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)

Предоставление, автоматическое обнаружение и сигнализация L2VPN

Аннотация

Предоставляемые провайдером виртуальные частные сети канального уровня (Provider Provisioned Layer 2 Virtual Private Network или L2VPN) могут использовать разные модели предоставления, т. е. модели, указывающие настраиваемые элементы и их конфигурацию. После настройки сведения о предоставлении распространяются процессом обнаружения (discovery process). Когда этот процесс завершается, автоматически вызывается сигнальный протокол для организации mesh-сети псевдопроводов (PW), формирующей (виртуальную) опорную сеть L2VPN. Этот документ задаёт множество моделей предоставления L2VPN, а также семантическую структуру идентификаторов конечных точек, требуемых для каждой модели. Рассматривается распространение этих идентификаторов процессом обнаружения, особенно для случая использования протокола граничного шлюза Border Gateway Protocol или BGP). Указывается, что идентификаторы конечных точек передаются двумя сигнальными протоколами для организации PW - протоколом распространения меток (Label Distribution Protocol или LDP) и протоколом туннелирования на канальном уровне (Layer 2 Tunneling Protocol version 3 или L2TPv3).

Статус документа

Документ относится к категории Internet Standards Track.

Документ является результатом работы IETF[1] и представляет согласованный взгляд сообщества IETF. Документ прошёл открытое обсуждение и был одобрен для публикации IESG[2]. Дополнительную информацию о стандартах Internet можно найти в разделе 2 в RFC 5741.

Информацию о текущем статусе документа, ошибках и способах обратной связи можно найти по ссылке http://www.rfc-editor.org/info/rfc6074.

Оглавление

[1]Internet Engineering Task Force - комиссия по решению инженерных задач Internet.
[2]Internet Engineering Steering Group - комиссия по инженерным разработкам Internet.

1. Введение

В [RFC4664] описано несколько вариантов, где наборы псевдопроводов могут объединяться в предоставляемые провайдером L2 VPN (Provider Provisioned Layer 2 VPN, L2 PPVPN или L2VPN), дающих разные виды L2VPN. Эти виды могут иметь разные модели предоставления (provisioning), т. е. разные варианты информации, которая должна быть настроена и разные наборы элементов. После настройки сведения о предоставлении распространяются процессом обнаружения (discovery process), а после обнаружения информации автоматически вызывается сигнальный протокол для организации требуемых псевдопроводов. Семантика идентификаторов конечных точек, используемых протоколом сигнализации для конкретного типа L2VPN, определяется моделью предоставления. Т. е. для разных типов L2VPN с разными моделями предоставления требуются различные типы идентификаторов конечных точек. В этом документе задано несколько моделей предоставления L2VPN и семантические структуры идентификаторов конечных точек для этих моделей.

Протокол LDP ([RFC5036] с расширением [RFC4447]) или L2TP версии 3 ([RFC3931] с расширением [RFC4667]) может служить для сигнализации при установке и поддержке PW [RFC3985]. Любой протокол, организующий соединения, должен обеспечивать каждой конечной точке способ представить себя другим, каждый сигнальный протокол PW, таким образом, обеспечивает способ идентификации конечных точек PW. Поскольку каждый сигнальный протокол должен поддерживать все виды L2VPN и модели предоставления, такой протокол должен иметь очень общий способ представления идентификаторов конечных точек в соответствующих полях каждого протокола сигнализации. Этот документ задаёт способ кодирования идентификаторов конечных точек для каждой модели предоставления в протоколах сигнализации LDP и L2TPv3.

Здесь используются термины из [RFC3985], [RFC4026], [RFC4664], [RFC5659], в частности, устройство присоединения (Attachment Circuit), псевдопровод (pseudowire), PE (provider edge - граница провайдера), CE (customer edge - граница клиента), многосегментный псевдопровод (multi-segment pseudowire).

В разделе 2 представлен обзор относящихся к документу аспектов [RFC4447] и [RFC4667]. Раздел 3 подробно описывает модели предоставления и связывает их с сигнальным процессом и процессом обнаружения. Достаточно подробно рассмотрены способы интеграции сигнальных механизмов с процессом автоматического обнаружения на основе BGP. В разделе 4 описано, как процедуры обнаружения и сигнализации можно применять в среде с множеством AS и рассмотрено несколько вариантов организации multi-AS L2VPN.

Ключевые слова **необходимо** (MUST), **недопустимо** (MUST NOT), **требуется** (REQUIRED), **нужно** (SHALL), **не нужно** (SHALL NOT), **следует** (SHOULD), **не следует** (SHOULD NOT), **рекомендуется** (RECOMMENDED), **возможно** (MAY), **необязательно** (OPTIONAL) в данном документе должны интерпретироваться в соответствии с [RFC2119]

2. Модель сигнального протокола

2.1. Идентификация конечных точек

В соответствии с [RFC4664] псевдопровод (pseudowire или PW) можно считать связью между двумя узлами пересылки (Forwarder). В простых экземплярах службы виртуальных частных проводов (Virtual Private Wire Service или VPWS) узел пересылки связывает псевдопровод в одним устройством подключения (Attachment Circuit или AC) так, что кадры, полученные одной стороной пересылаются на другую и наоборот. В услугах виртуальных частных ЛВС (Virtual Private LAN Service или VPLS) узел пересылки связывает набор псевдопроводов с набором AC и при получении кадра любым из элементов набора выполняется обращение к таблице MAC[1]-адресов (и выполнение различных процедур 802.1d) для определения элемента или элементов набора, которым пересылается кадр. В более сложных вариантах узлы пересылки могут связывать PW с PW, тем самым соединяя их. Это нужно для поддержки распределенных VPLS и некоторых вариантов обмена между автономными системами (AS).

[1]Media Access Control - управление доступом к среде.

В простой VPWS, где Forwarder связывает один PW в одни AC, узел пересылки можно указать через его AC. В простых VPLS узел пересылки можно идентифицировать по его устройству PE и VPN.

Для организации PW между парой Forwarder сигнальный протокол должен позволять узлу пересылки одной конечной точки идентифицировать Forwarder в другой точке. В [RFC4447] применяется термин «идентификатор присоединения» (Attachment Identifier или AI) для указания элемента, идентифицирующего Forwarder. В [RFC4667] для этого же служит термин «идентификатор узла пересылки» (Forwarder Identifier). В этом документе применяются оба термина.

[RFC4447] задаёт два элемента классов эквивалентности пересылки (Forwarding Equivalence Class или FEC), которые можно использовать при организации псевдопроводов - это элементы PWid FEC и Generalized ID FEC. Элемент PWid FEC передаёт лишь один Forwarder identifier и может применяться лишь при использовании обоими узлами пересылки одного идентификатора, который можно представить 32-битовым значением. Элемент Generalized ID FEC передаёт два Forwarder identifier, по одному для каждого из соединённых узлов пересылки. Каждый идентификатор называют AI и сигнальное сообщение содержит идентификаторы источника (Source Attachment Identifier или SAI) и цели (Target Attachment Identifier или TAI).

Элемент Generalized ID FEC обеспечивает дополнительное структурирование идентификаторов. Предполагается, что SAI и TAI иногда будут иметь общую часть, называемую идентификатором группы присоединения (Attachment Group Identifier или AGI) так, что SAI и TAI могут быть конкатенацией AGI с соответствующим индивидуальным идентификатором присоединения (Attachment Individual Identifier или AII). Таким образом, пара идентификаторов представляется 3 полями: AGI, Source AII (SAII), Target AII (TAII). SAI будет конкатенацией AGI и SAII, TAI - AGI и TAII.

[RFC4667] позволяет использовать один или два Forwarder Identifier для организации псевдопроводов. При использовании одного Forwarder Identifier в сигнальных сообщениях L2TP предполагается одно значение у исходного и целевого узла пересылки. Если два Forwarder Identifier передаются в сигнальных сообщениях L2TP, каждый Forwarder использует локально значимый идентификатор.

Forwarder Identifier в [RFC4667] является эквивалентом Attachment Identifier из [RFC4447]. Forwarder Identifier состоит из идентификатора группы присоединения (AGI) и индивидуального идентификатора подключения (AII). В отличие от элемента Generalized ID FEC, AGI и AII передаются в разных парах «атрибут-значение» L2TP (Attribute-Value Pair или AVP). AGI кодируется в AGI AVP, а SAII и TAII в Local End ID AVP и Remote End ID AVP, соответственно. Forwarder Identifier источника является конкатенацией AGI и SAII, Forwarder Identifier цели - конкатенацией AGI и TAII.

В приложениях, группирующих PW в Layer 2 Virtual Private Network, AGI можно считать VPN Identifier.

Следует отметить, что хотя разные узлы пересылки поддерживают различные приложения, тип приложения (например, VPLS или VPWS) не всегда можно вывести из идентификаторов узлов пересылки. Маршрутизатор, получающий сигнальное сообщение с конкретным TAI, должен иметь возможность определить, какой из локальных узлов пересылки указывается данным TAI, и определить приложение, представляемое этим узлом пересылки. Но другие узлы могут быть не в состоянии определить приложение, просто просматривая сигнальные сообщения.

В этом документе предлагается некая дополнительная структура AGI и AII для ряда приложений L2VPN. Отметим, что [RFC4447] задаёт структуру TLV для полей AGI и AII. Таким образом, оператор, выбравший определённую здесь структуру AII, может также применять другие типы AGI или AII, если он хочет использовать иную структуру для этих идентификаторов для некоторых других приложений. Например, можно использовать тип длинного префикса [RFC5003] для обеспечения возможности передачи административных данных, возможно, с сочетании со сведениями, полученными при автоматическом обучении.

2.2. Создание одностороннего псевдопровода

При любой форме сигнализации на основе LDP каждая конечная точка PW должна инициировать создание одностороннего LSP. Пара таких LSP образует PW. В большинстве моделей предоставления L2VPN две конечных точки данного PW могут одновременно инициировать сигнализацию для псевдопровода. Поэтому им нужен способ определить, когда данная пара LSP предназначена для объединения в PW. Такой способ отличается в разных службах и моделях предоставления L2VPN, а детали представлены в последующих параграфах.

Сигнализация L2TP по своей природе организует двухстороннюю сессию, создающую PW между двумя конечными точками. Эти точки также могут одновременно инициировать сигнализацию для данного PW. Возможна организация между парой узлов пересылки двух PW. Чтобы избежать организации дублированных псевдопроводов между узлами Forwarder, каждый узел PE должен быть способен независимо обнаруживать такую связь между псевдопроводами. Процедуры обнаружения описаны в [RFC4667].

2.3. Идентификаторы присоединения и узлы пересылки

С каждым узлом Forwarder в PE должен быть связан идентификатор присоединения (AI) путём настройки или использования некого алгоритма. AI должны быть уникальными в контексте маршрутизатора PE, на котором размещается Forwarder. Комбинация <маршрутизатор PE, AI> должна быть уникальна в глобальном масштабе.

Как указано в [RFC4447], AI может состоять из группового (AGI) и индивидуального (AII) идентификаторов подключения. В контексте этого документа AGI можно рассматривать как VPN-ID или некий иной атрибут, общий для всех устройств присоединения, которым разрешено подключение.

Иногда полезно рассматривать устройства присоединения на одном PE как относящиеся к общему пулу. Например, набор устройств присоединения, подключающих отдельное устройство CE к данному PE можно считать пулом. Применение пулов рассмотрено в параграфе 3.3. Пулы с окрашиванием - полная связность псевдопроводов P2P.

Детали создания полей AGI и AII, указывающих конечные точки псевдопроводов в различных моделях предоставления, рассматриваются ниже.

Можно считать, что LSP для одного направления псевдопровода идентифицируется набором

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

а LSP двух противоположных направлений набором

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

Псевдопровод является парой таких LSP. При использовании сигнализации L2TP это относится к двум направлениям сессии L2TP.

Когда сигнальное сообщение передаётся от PE1 к PE2 и PE1 нужно указать, что AI для настройки является одно из его устройств присоединения (или пулов), AI называют идентификатором присоединения источника (SAI). Если PE1 нужно указать, что AI для настройки является одно из устройств присоединения (или пулов) PE2, AI называют идентификатором присоединения цели (TAI). SAI относится к одной конечной точке, а TAI - к другой (удалённой).

В сигнальном протоколе определяется кодирование трёх указанных ниже полей.

- Групповой идентификатор присоединения (Attachment Group Identifier или AGI).
- Индивидуальный идентификатор присоединения источника (Source Attachment Individual Identifier или SAII).
- Индивидуальный идентификатор присоединения цели (Target Attachment Individual Identifier или TAII).

Если поле AGI не пусто (non-null), SAI состоит из AGI и SAII, а TAI - из AGI и TAII. При пустом AGI в качестве SAI и TAI выступают SAII и TAII, соответственно.

Цель состоит в том, чтобы узел PE, получающий сообщение LDP Label Mapping или L2TP Incoming Call Request (ICRQ) с TAI, был способен однозначно сопоставить TAI с одним из своих устройств присоединения (или пулов). Способ сопоставления следует задавать локально (включая использование для этого части или всех байтов TAI). В части процедур сигнализации TAI на деле представляет собой просто строку байтов (cookie).

3. Применение

В этом разделе указано, как применяется сигнализация псевдопроводов с использованием обозначения исходного и целевого узла пересылки для различных приложений. Для некоторых из них указываются способы применения различных моделей предоставления. Однако описания не являются исчерпывающими.

3.1. Индивидуальные псевдопровода «точка-точка»

Заданная в этом документе сигнализация может служить для организации индивидуально предоставляемых псевдопроводов «точка-точка». В этом случае каждый узел пересылки связывает один PW с одним устройством присоединения. Каждый узел PE должен предоставляться с необходимым набором устройств присоединения, а затем этим устройствам должны предоставляться некоторые параметры.

3.1.1. Модели предоставления

3.1.1.1. Двухстороннее предоставление

В этой модели устройство присоединения должно обеспечиваться локальным именем, адресом удалённого PE и удаленным именем. При сигнализации локальное имя передаётся как SAII, удалённое - как TAII, а AGI остаётся пустым. Если два устройства присоединения подключены к PW, локальное имя одного должно быть удаленным именем для другого.

Отметим, что при совпадении локального и удалённого имени можно применять элемент PWid FEC вместо элемента Generalized ID FEC в сигнализации на основе LDP.

При сигнализации L2TP локальное имя передаётся в Local End ID AVP, удалённое - в Remote End ID AVP. AGI AVP использовать необязательно и при наличии оно содержит значение AGI нулевого размера. Если локальное имя совпадает с удаленным, Local End ID AVP можно не включать в сигнальные сообщения L2TP.

3.1.1.2. Одностороннее предоставление с обнаружением

В этой модели каждое устройство присоединения должно иметь локальное имя. Имя состоит из VPN-ID (передаётся как AGI) и идентификатора AII, который уникален в рамках AGI. Если два устройства присоединения связаны PW, лишь одному из них требуется удалённое имя (локальное имя другого устройства). Ни одному из них не требуется адрес удалённого PE, но оба должны иметь одно значение VPN-ID.

В рамках процедуры автоматического обнаружения каждый узел PE анонсирует свою пару <VPN-id, local AII> и каждый PE сравнивает свою пару <VPN-id, remote AII> с парами <VPN-id, local AII>, анонсируемыми другими PE. Если у PE1 имеется пара <VPN-id, remote AII> со значением <V, fred>, а у PE2 - <VPN-id, local AII> с <V, fred>, PE1 сможет понять, что ему нужно соединиться с PE2. При сигнализации узел будет использовать значение fred для TAII и V для AGI. Локальное имя PE1 для устройства присоединения передаётся как SAII.

Основное преимущество этой модели по сравнению с двухсторонним предоставлением заключается в том, что она позволяет переносить устройство присоединения от одного PE к другому без перенастройки удалённой конечной точки. Однако по сравнению с подходом, описанным в параграфе 3.3. Пулы с окрашиванием - полная связность псевдопроводов P2P, эта модель сильнее загружает механизм обнаружения, поскольку каждое имя устройства присоединения должно анонсироваться отдельно (имена устройств присоединения в этой схеме не объединяются).

3.1.2. Сигнализация

Сигнализация на основе LDP следует процедурам, заданным в [RFC4447], т. е. один узел PE (PE1) передаёт сообщение Label Mapping другому PE (PE2) для организации LSP в одном направлении. Если сообщение успешно обработано и ещё нет LSP для псевдопровода в обратном направлении (PE1->PE2), PE2 передаёт сообщение Label Mapping узлу PE1.

В дополнение к процедурам [RFC4447] при получении узлом PE сообщения Label Mapping с указанием в TAI конкретного устройства присоединения, настроенного на привязку к PW «точка-точка» должны быть выполнены указанные ниже проверки.

Если устройство присоединения уже связано с псевдопроводом (включая случай наличия лишь одного из двух LSP) и удалённой конечной точкой является не PE1, PE2 передаёт PE1 сообщение Label Release со Status Code, указывающим, что устройство присоединения связано с другим PE, и обработка сообщения Mapping завершается.

Если устройство присоединения уже связано с псевдопроводом (включая случай наличия лишь одного из двух LSP) и AI на PE1 отличается от заданного в полях AGI/SAII сообщения Mapping, PE2 передаёт PE1 сообщение Label Release со Status Code, указывающим, что устройство присоединения связано с другим удаленным устройством присоединения, и обработка сообщения Mapping завершается.

Когда PE при использовании сигнализации на основе L2TP получает сообщение ICRQ и TAI указывает конкретное устройство присоединения, настроенное на привязку к PW «точка-точка», выполняются указанные ниже проверки.

Если устройство присоединения уже связано с псевдопроводом и удалённой конечной точкой является не PE1, PE2 передаёт PE1 сообщение Call Disconnect Notify (CDN) со Status Code, указывающим, что устройство присоединения связано с другим PE, и обработка сообщения ICRQ завершается.

Если устройство присоединения уже связано с псевдопроводом, но тот привязан к Forwarder на PE1 с AI, отличающимся от указанного в полях SAI сообщения ICRQ, PE2 передаёт PE1 сообщение CDN со Status Code, указывающим, что устройство присоединения связано с другим удаленным устройством присоединения, и обработка сообщения ICRQ завершается.

Указанные ошибки могут быть результатом некорректной конфигурации.

3.2. Виртуальные частные ЛВС

В VPLS [RFC4762] устройства присоединения можно считать интерфейсами ЛВС, подключёнными к «виртуальным коммутаторам ЛВС» или, в терминологии [RFC4664], к «экземплярам виртуальной коммутации» (Virtual Switching Instance или VSI). Каждый узел Forwarder является VSI, подключённым к множеству PW и множеству устройств присоединения. Служба VPLS требует создания одного псевдопровода между каждой парой VSI в одной сети VPLS. Каждое устройство PE может иметь несколько VSI, каждый из которых относится к своей VPLS.

3.2.1. Предоставление

У каждой сети VPLS должен быть глобально уникальный идентификатор, который в [RFC4762] назван VPLS identifier (или VPLS-id). На каждом VSI должен быть указан идентификатор VPLS-id сети VPLS, к которой он относится.

У каждого VSI должен быть уникальный идентификатор, который называется VSI-ID. Идентификатор может создаваться автоматически конкатенацией VPLS-id и IP-адреса его маршрутизатора PE. Отметим, что адрес PE здесь служит лишь формой уникального идентификатора и провайдер может выбрать иную схему нумерации, коль скоро она обеспечивает уникальность каждого идентификатора VSI в рамках экземпляра VPLS. В параграфе 4.4 рассматривается назначение идентификаторов при наличии нескольких провайдеров.

3.2.2. Автоматическое обнаружение

3.2.2.1. Автоматическое обнаружение на основе BGP

В этом параграфе описано применение BGP для обнаружения информации, требуемой при создании экземпляров VPLS. При автоматическом обнаружении на основе BGP для VPLS идентификаторами AFI/SAFI [RFC4760] являются:

- AFI (25) для L2VPN (как для всех схем L2VPN);
- SAFI (65) конкретно для службы L2VPN, псевдопровода которой организованы по процедурам, описанным в этом документе.

Назначение AFI/SAFI рассматривается в разделе 6. Взаимодействие с IANA.

Для автоматического обнаружения на основе BGP нужен хотя бы один глобально уникальный идентификатор, связанный с VPLS, и такие идентификаторы должны быть представимыми в форме 8-байтовых различителей маршрутов (Route Distinguisher или RD). Подойдёт любой метод назначения одного или нескольких уникальных идентификаторов VPLS и представления каждого в форме RD (с использованием кодирования из [RFC4364]).

Для каждого экземпляра VSI требуется уникальный идентификатор, кодируемый в форме BGP NLRI[1]. Он формируется путём добавления RD (см. выше) перед IP-адресом PE, содержащего VSI. Отметим, что этот адрес служит лишь легко доступным уникальным идентификатором VSI в сети VPN и не обязан быть маршрутизируемым в глобальном масштабе, но должен быть уникальным внутри экземпляра VPLS. При желании можно применять иную схему назначения уникальных идентификаторов для каждого VSI внутри экземпляра VPLS (например, нумерация VSI в одной сети VPN от 1 до n).

При использовании описанных в этом документе процедур требуется назначить один глобально уникальный идентификатор VPLS-id для каждого экземпляра VPLS [RFC4762]. Этот идентификатор должен кодироваться в форме расширенной группы BGP (Extended Community) [RFC4360]. Как указано в разделе 6. Взаимодействие с IANA, этот документ определяет два субтипа Extended Community, которые **должны** быть переходными.

Первым субтипом расширенной группы является 2-октетное значение AS Specific Extended Community, вторым - IPv4 Address Specific Extended Community. Их кодирование задано в [RFC4360] и обеспечивает сервис-провайдерам возможность выделять VPLS-id без риска возникновения конфликтов с другими провайдерами. Однако следует отметить, что для межпровайдерских L2VPN требуется координация VPLS-id, как описано в параграфе 4.4. Назначение RT и RD.

Каждый экземпляр VSI необходимо связать с одной или несколькими расширенными группами Route Target (RT). Это управляет распространением NLRI и поэтому будет контролировать формирование топологии псевдопроводов, образующих конкретную сеть VPLS.

[1] Network Layer Reachability Information - информация о доступности на сетевом уровне.

Автоматическое обнаружение выполняется путём распространения каждым PE по протоколу BGP данных NLRI для каждого из своих VSI с указанием себя в качестве следующего интервала BGP (next hop) и подходящего RT для каждого NLRI. Обычно каждый узел PE будет клиентом небольшого числа рефлекторов BGP, которые будут распространять эти сведения другим клиентам.

Если PE получает обновление BGP, в котором отсутствуют какие-либо из указанных выше элементов, такое обновление следует игнорировать.

Если у PE имеется VSI с определенным RT, он может импортировать все NLRI с тем же RT и узнать из атрибута BGP next hop в этих NLRI адреса IP других маршрутизаторов PE, имеющих VSI с тем же RT. Для этого применимо использование маршрутных рефлекторов, описанное в параграфе 4.3.3 [RFC4364].

Если конкретная сеть VPLS предназначена для использования в качестве одной полностью подключённой ЛВС, все её VSI будут иметь общую цель RT и в этом случае RT может (но не обязательно) быть кодированием VPN-id. Экземпляр VSI можно разместить в нескольких VPLS, назначая ему несколько RT.

Отметим, что можно создать иерархическую VPLS, назначая несколько RT некоторым из VSI, механизм RT обеспечивает полный контроль над перекрытием псевдопроводов, входящих в топологию VPLS.

При реализации распределенной VPLS (параграф 3.5. Распределенные VPLS) в автоматическом обнаружении на основе BGP участвуют лишь обращенные в сеть PE (Network-facing PE или N-PE). Это значит, что N-PE должны анонсировать доступность каждому поддерживаемому VSI, включая на обращенные к пользователю PE (User-facing PE или U-PE), с которыми они соединены. Для создания виртуального идентификатора каждому VSI можно использовать IP-адрес каждого U-PE в комбинации с RD экземпляра VPLS.

Таким образом, анонс BGP для отдельного VSI на данном PE будет включать:

- NLRI с AFI = L2VPN, SAFI = VPLS в форме RD:PE_addr;
- loopback-адрес PE в качестве BGP next hop;
- Extended Community Attribute с VPLS-id;
- Extended Community Attribute с одним или несколькими RT.

Значения AFI и SAFI рассмотрены в разделе 6. Взаимодействие с IANA. Формат NLRI представлен ниже.

```
+-----------------------------------+
|  Length (2 октета)                |
+-----------------------------------+
|  Route Distinguisher (8 октетов)  |
+-----------------------------------+
|  PE_addr (4 октета)               |
+-----------------------------------+
```

Отметим, что этот анонс похож на формат NLRI, определённый в [RFC4761], а основное отличие состоит в том, что [RFC4761] включает в NLRI блок меток. Взаимодействие между определённой здесь схемой VPLS и схемой [RFC4761] выходит за рамки этого документа.

3.2.3. Сигнализация

Необходимо создавать идентификаторы присоединения, указывающие VSI. В предыдущем параграфе указано кодирование VSI-ID как RD:PE_addr, и передача VPLS-id в BGP Extended Community. Здесь описано кодирование этой информации для сигнальных процессов. VPLS-id помещается в поле AGI, а PE_addr (точнее, VSI-ID из NLRI в BGP без RD) - в поле TAII. Комбинации AGI и TAII достаточно для полного указания VSI, к которому подключён данный псевдопровод в средах с одной или несколькими автономными системами (AS). Для SAII передающее устройство PE **должно** устанавливать значение PE_addr (точнее, VSI-ID без RD экземпляра VSI, связанного с данной сетью VPLS в передающем PE), чтобы включить при необходимости сигнализацию обратной половины (направления) PW.

Структура полей AGI и AII для Generalized ID FEC в LDP задана в [RFC4447]. Поле AGI в этом случае содержит Type = 1, размер 8 и поле VPLS-id размером 8 байтов. Поля AII содержат Type = 1, размер 4 и адрес или иной идентификатор PE размером 4 байта. Назначение типа AGI и AII описано в разделе 6. Взаимодействие с IANA.

Кодирование AGI и AII в L2TP описано в [RFC4667].

Отметим, что этот метод не позволяет организовать более одного PW на пару VSI.

3.2.4. Псевдопровода как устройства присоединения VPLS

Можно использовать этот метод для организации PW, подключённого одним концом к VSI, а другим - к конечной точке на устройстве присоединения. На данном VSI может быть множество PW, которые нужно различать, поэтому каждый псевдопровод (PW) должен иметь значение SAII, уникальной для VSI-ID.

3.3. Пулы с окрашиванием - полная связность псевдопроводов P2P

Модель пулов с окрашиванием (Colored Pools) обеспечивает автоматизированный способ предоставления услуг VPWS. В этой модели каждое устройство PE может включать несколько пулов устройств присоединения, каждый из которых связан со своей VPN. PE может включать несколько пулов VPN, поскольку каждый пул может соответствовать своему устройству CE. Может быть желательно создание одного псевдопровода между каждой парой пулов в одной VPN и результатом станет создание полносвязной (full mesh) сети виртуальных устройств CE-CE для каждой VPN.

3.3.1. Предоставление

Каждый пул настраивается и связывается с:

- набором устройств присоединения;
- «цветом», который можно считать неким идентификатором VPN-id;

- относительным идентификатором пула, который уникален в рамках данного цвета.

Примечание. В зависимости от технологии, применяемой для устройств присоединения (AC), может потребоваться предоставление таких устройств. Например, если AC являются устройствами трансляции кадров (frame relay), может существовать отдельная система предоставления таких устройств. «Предоставление» AC может быть столь же простым, как выделение неиспользуемого VLAN ID на интерфейсе и указание его клиенту. Это не зависит от описанных в документе процедур.

Идентификатор пула и цвет совместно задают глобально уникальный идентификатор для пула. При наличии n пулов данного цвета их идентификаторами могут быть числа от 1 до n, но это не обязательно.

Семантика заключается в том, что псевдопровод будет создаваться между каждой парой пулов одного цвета и каждый псевдопровод будет связан с одним AC каждого из этих двух пулов.

Если каждый пул является набором AC, ведущих к одному устройству CE, связность L2 между CE контролируется тем же способом, каким выделяются цвета для пулов. Для создания полной связности «цвет» будет просто VPN-id.

Можно настроить отдельное устройство присоединения с относительным идентификатором удалённого пула. Затем AC привязывается к определённому псевдопроводу, только когда удалённый конец PW является пулом с относительным идентификатором. В этом случае пары устройств AC всегда будут связаны псевдопроводами.

3.3.2. Автоматическое обнаружение

3.3.2.1. Автоматическое обнаружение на основе BGP

В этом параграфе описано, как можно использовать протокол BGP для обнаружения сведений, требуемых при создании экземпляров VPWS. При использовании автоматического обнаружения на основе BGP для VPWS поля AFI/SAFI имеют указанные ниже значения.

- Значение AFI, заданное IANA для L2VPN (как и для всех схем L2VPN).
- Значение SAFI, выделенное IANA специально для служб L2VPN, где псевдопровода создаются с использованием описанных в этом документе процедур.

Назначение AFI/SAFI рассматривается в разделе 6. Взаимодействие с IANA.

Для автоматического обнаружения на основе BGP с экземпляром VPWS должен быть связан хотя бы один уникальный идентификатор. Каждый из таких идентификаторов должен быть представим как RD (отличитель маршрута). Глобально уникальный идентификатор пула должен быть представим как NLRI, идентификатор пула, определённый как 4-байтовое значение, добавляется в конец RD для создания NLRI.

При использовании описанных в этом документе процедур необходимо назначить один уникальный в глобальном масштабе идентификатор каждому экземпляру VPWS. Этот идентификатор должен быть представим в виде BGP Extended Community [RFC4360]. Как указано в разделе 6, этот документ определяет для этого два субтипа Extended Community. Расширенные группы (Extended Community) **должны** быть переходными.

Первый субтип Extended Community является Two-octet AS Specific Extended Community, второй - IPv4 Address Specific Extended Community. Их кодирование определено в [RFC4360] и обеспечивает сервис-провайдерам возможность выделять идентификаторы VPWS без риска конфликтов с другими провайдерами. Однако следует отметить, что согласование идентификаторов VPWS требуется при создании межпровайдерских L2VPN, описанных в параграфе 4.4.

Каждый пул также должен быть связан со значением RT (цель маршрута), которое может также представлять цвет. Если желаемая топология является полносвязной сетью псевдопроводов, все пулы будут иметь одно значение RT. Другие варианты топологии рассмотрены в параграфе 3.4. Пулы с окрашиванием - частичная связность.

Автоматическое обнаружение основано на распространении каждым PE (через BGP) сведений NLRI для каждого из своих пулов с указанием себя как BGP и кодированием цвета пула в RT. Если у данного PE имеется пул с конкретным цветом (RT), он должен получать через BGP все NLRI с тем же цветом (RT). Обычно каждый PE является клиентом небольшого числа маршрутных рефлекторов BGP, которые будут распространять эти сведения другим клиентам.

Если PE получает обновление BGP, в котором отсутствует какой-либо из указанных выше элементов, обновление следует игнорировать.

Если у PE есть пул определённого цвета, он может получать все NLRI с тем же цветом а из атрибута BGP next hop в NLRI узнает адреса IP других маршрутизаторов PE, у который имеются пулы того же цвета. Он узнает также уникальный идентификатор каждого из этих удалённых пулов, поскольку он представлен в NLRI. Относительный идентификатор удалённого пула можно извлечь из NLRI и применять для сигнализации, как указано ниже.

Таким образом, анонс BGP для определённого пула устройств присоединения на данном PE будет содержать:

- NLRI с AFI = L2VPN, SAFI = VPLS в форме RD:pool_num;
- BGP next hop с loopback-адресом PE;
- Extended Community Attribute с идентификатором VPWS;
- Extended Community Attribute с одним или несколькими RT.

Значения AFI и SAFI приведены в разделе 6. Взаимодействие с IANA.

3.3.3. Сигнализация

Сигнализация на основе LDP следует процедурам, заданным в [RFC4447], т. е. узел PE (PE1) передаёт сообщение Label Mapping другому PE (PE2) для организации LSP в одном направлении. Адресом PE2 является адрес next-hop, полученный от BGP, как описано выше. Если сообщение успешно обработано и ещё нет LSP для псевдопровода в обратном направлении (PE1->PE2), PE2 передаёт сообщение Label Mapping узлу PE1. Сигнализация на основе L2TPv3 следует процедурам [RFC4667]. Детали использования протоколов сигнализации приведены ниже.

Когда PE отправляет сообщение Label Mapping или ICRQ для организации PW между двумя пулами, идентификатор VPWS (распространяемый BGP в Extended Community Attribute) кодируется как AGI, относительный идентификатор локального пула - как SAII, а относительный идентификатор удалённого пула - как TAII.

Структура полей AGI и AII для Generalized ID FEC в LDP задана в [RFC4447]. Поле AGI в этом случае включает Type = 1, размер 8 и идентификатор VPWS (8 байтов). TAII включает Type = 1, размер 4 и номер удалённого пула (4 байта). SAII включает Type = 1, размер 4 и номер локального пула (4 байта. Назначение типов AGI и AII описано в разделе 6. Взаимодействие с IANA. Отметим, что заданные в этом документе процедуры VPLS и VPWS могут применять те же типы AGI (1) и AII (1).

Кодирование AGI и AII в L2TP задано в [RFC4667].

Когда PE2 получает сообщение Label Mapping или ICRQ от PE1, а TAI указывает пул и уже есть псевдопровод, связывающий устройство присоединения этого пуля с устройством присоединения на PE1, а AI на PE1 этого псевдопровода совпадает с SAI из сообщения Label Mapping или ICRQ, узел PE2 передаёт сообщение Label Release или CDN узлу PE1 со Status Code, указывающим, что устройство присоединения уже связано с удаленным устройством присоединения. Это предотвращает создание множества псевдопроводов между данной парой пулов.

Отметим, что сама сигнализация идентифицирует лишь удалённый пул, к которому ведёт псевдопровод, а не удалённое устройство присоединения, связанное с псевдопроводом. Однако удалённый узел PE может проверить поле SAII для определения устройства присоединения, связанного с псевдопроводом.

3.4. Пулы с окрашиванием - частичная связность

Процедуры создания сети псевдопроводов без полной связности (partial mesh) между окрашенныеми пулами, по существу такие же, как при организации полной связности, и отличаются лишь в нескольких аспектах.

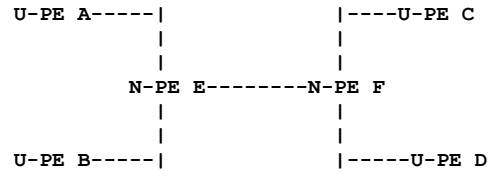
- Для каждого пула можно задать наборы «импортируемых RT» и экспортируемых RT».
- В процессе автоматического обнаружения на основе BGP цвет пула по-прежнему кодируется в RD, но если в пуле задан набор «экспортируемых RT», они представляются в RT сообщений BGP Update **вместо** цвета.
- Если пул имеет экспортируемый RT со значением X, будет создаваться PW с каждым другим пулом, имеющим X в одном из своих экспортируемых RT. Сигнальные сообщения и процедуры соответствуют параграфу 3.3.3. Сигнализация.

В качестве простого примера рассмотрим задачу создания топологии «звезда» (hub-and-spoke) с одним хабом. Один пул (хаб) настраивается с экспортируемым RT_hub и импортируемым RT_spoke. Все остальные пулы (лучи) настраиваются с экспортирумым RT_spoke и импортируемым RT_hub. Таким образом, пул хаба соединяется с лучами и наоборот, но пулы лучей не соединяются между собой.

3.5. Распределенные VPLS

В Distributed VPLS ([RFC4664]) функциональность VPLS узла PE делится между двумя системами - U-PE и N-PE, U-PE размещается между пользователем и N-PE. Функциональность VSI (например, изучение MAC и функции моста) реализуется в U-PE. К N-PE может быть подключено несколько U-PE. Для каждой VPLS, поддерживаемой U-PE, обеспечивается псевдопровод к каждому из других U-PE той же VPLS. Однако U-PE не поддерживает управляющие сигнальные сообщения с другими, а вместо этого имеет лишь одно сигнальное соединение со своим N-PE. По сути, каждый псевдопровод между U-PE состоит из 3 псевдопроводов, соединённых вместе - от U-PE к N-PE, от N-PE к N-PE и от N-PE к U-PE. В терминологии [RFC5659] N-PE выполняют функцию коммутации псевдопроводов для организации многосегментных PW между U-PE.

Рассмотрим в качестве примера показанную на рисунке топологию, где 4 U-PE имеют общую сеть VPLS.



Покажем, как PW соединяются между собой в приведённой выше топологии для организации требуемых псевдопроводов от U-PE A к другим U-PE.

Имеется 3 PW от A к E. Обозначим их A-E/1, A-E/2, A-E/3. Для правильного подключения A к другим U-PE, нужны 2 PW от E к F (E-F/1 и E-F/2), 1 PW от E к B (E-B/1), 1 от F к C (F-C/1) и 1 от F к D (F-D/1).

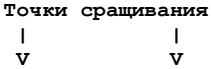
N-PE должны затем «срастить» эти псевдопровода, чтобы получить эквивалент механизма нераспределенной сигнализации VPLS:

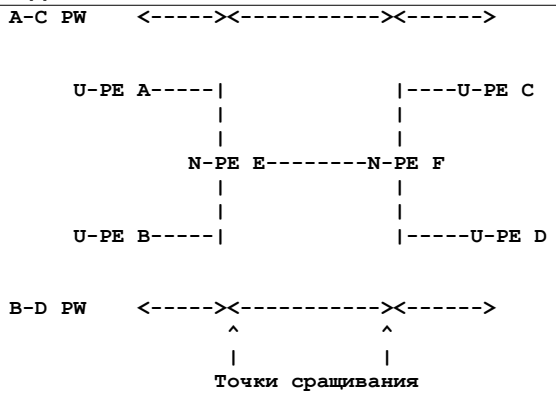
- PW от A к B: A-E/1 сращивается с E-B/1.
- PW от A к C: A-E/2 сращивается с E-F/1, а тот - с F-C/1.
- PW от A к D: A-E/3 сращивается с E-F/2, а тот - с F-D/1.

Неважно, какие PW сращиваются, пока результат даёт псевдопровода от A в B, C и D.

Аналогично требуется срастить дополнительные PW для подобающего соединения U-PE B с U-PE C и D, а также U-PE C с U-PE D.

На рисунке ниже показаны PW от A к C и от B к D. Для простоты остальные 4 PW опущены.





Можно видеть, что распределенная VPLS не сокращает число PW на U-PE, но снижает число управляющих соединений на U-PE. Следует ли это делать, зависит от того, где находится узкое место (bottleneck).

3.5.1. Сигнализация

Сигнализацию для поддержки Distributed VPLS можно реализовать с помощью описанных здесь механизмов. Однако процедуры для VPLS (3.2.3. Сигнализация) требуют дополнительных механизмов, обеспечивающих приемлемое число PW между разными N-PE и U-PE, а также между N-PE.

На данном N-PE подключённые напрямую U-PE данной VPLS можно пронумеровать от 1 до n. Эти номера указывают U-PE для конкретных VPN-id и N-PE (т. е. для однозначного указания U-PE нужно знать N-PE, VPN-id и номер U-PE).

В результате настройки/обнаружения каждому U-PE должен быть предоставлен список пар <j, IP-адрес>. Каждый элемент этого списка указывает U-PE настроить j псевдопроводов к указанному адресу IP. Когда U-PE передаёт сигнал N-PE, для AGI устанавливается подобающий идентификатор VPN-id, для SAII - номер PW, а для TAII - null.

В приведённом выше примере U-PE A будет передано <3, E> для организации 3 PW к E. При сигнализации A установит для AGI подходящее значение VPN-id, а для SAII - значение 1, 2 или 3 (номер PW, для которого передаётся сигнал).

В результате настройки/обнаружения каждому N-PE должны быть предоставлены указанные ниже сведения для каждой VPLS.

- «Локальный» список {<j, IP-адрес>}, где каждый элемент задаёт организацию j псевдопроводов к локально подключённому U-PE по указанному адресу. Число элементов в этом списке (n) будет число локально подключённых U-PE в этой сети VPLS. В приведённом выше примере E будет передан локальный список {<3, A>, <3, B>}, задающий организовать по 3 PW к A и B.
- Локальная нумерация U-PE для конкретной сети VPLS и определённого N-PE. В приведённом выше примере E можно было бы указать, что U-PE A имеет номер 1, а U-PE B - 2.
- «Удалённый» список {<IP-адрес, k>}, задающий организацию k псевдопроводов для каждого U-PE к указанному адресу IP. Каждый из IP-адресов указывает N-PE, а k - число U-PE на N-PE, входящих в VPLS. В приведённом выше примере E получит удалённый список {<2, F>}. Поскольку N-PE E имеет 2 U-PE, это задаёт создание 4 PW к N-PE F, по 2 для каждого из U-PE узла N-PE E.

Сигнализация PW от N-PE к U-PE основана на локальном списке и локальной нумерации U-PE. При передаче сигнализации определённого PW от N-PE к U-PE для AGI устанавливается подходящее значение VPN-id, SAII пусто (null), а для TAII устанавливается номер PW (для конкретных VPLS и U-PE). В приведённом выше примере при передаче сигнала от E к A для TAII будет устанавливаться значение 1, 2 или 3 в соответствии с тремя PW, которые нужно организовать к A. Аналогичным способом передаются 3 PW для B.

LSP, передаваемые от U-PE к N-PE, связаны с LSP от N-PE к U-PE обычным способом. PW между U-PE и N-PE называют U-PW.

Сигнализация соответствующего набора PW от N-PE к N-PE основана на удалённом списке. Все PW между N-PE можно считать эквивалентными. Пока установлено корректное число PW, N-PE могут сращивать эти PW с подходящими U-PW. Сигнализация корректного числа PW от N-PE к N-PE основана на удалённом списке, который задаёт число организуемых PW к конкретному удалённому N-PE на локальный U-PE.

При сигнализации конкретного PW от N-PE к N-PE для AGI устанавливается подходящее значение VPN-id. TAII указывает удалённый узел N-PE, как в нераспределенном случае, т. е. содержит IP-адрес удалённого N-PE. При наличии n таких PW, они различаются значениями SAII. Для поддержки множества значений SAII в одной сети VPLS передающему N-PE требуется столько VSI-ID, сколько у него U-PE. Как отмечено в параграфе 3.2.2, это можно обеспечить, например, использованием IP-адреса каждого подключённого U-PE. PW между N-PE называют N-PW.

Каждый U-PW должен быть «срощен» с N-PW. Это делается на основе удалённого списка. Если этот список содержит элемент <i, F>, то i псевдопроводов U-PW от каждого локального U-PE должны быть сращены с i псевдопроводов N-PW от удалённого N-PE F. Не имеет значения, какие U-PW срощены с какими N-PW, пока это ограничение соблюдается.

Если N-PE имеет более одного локальной U-PE для данной VPLS, он должен также обеспечивать сращивание U-PW от каждого такого U-PE с U-PW от каждого из других U-PE.

3.5.2. Предоставление и обнаружение

Каждый N-PE должен предоставляться с набором поддерживаемых экземпляров VPLS, VPN-id для каждого из них и списком локальных U-PE для каждой из этих VPLS. В рамках процедуры обнаружения N-PE анонсирует число U-PE для каждой VPLS (см. 3.2.2. Автоматическое обнаружение).

Может применяться автоматическое обнаружение (например, через BGP) для всех других N-PE в VPLS и для каждого из них - числа U-PE, локальных для этого N-PE. Исходя из этого можно найти общее число U-PE в VPLS. Этих сведений достаточно для расчёта локального и удалённого списка каждого N-PE.

3.5.3. Нераспределенная VPLS как частный случай

Узел PE, обеспечивающий «нераспределенную VPLS» (PE в роли U-PE и N-PE сразу) может взаимодействовать с парами N-PE-U-PE, обеспечивающими распределенную VPLS. Такой PE просто анонсирует в процедуре обнаружения, что он имеет один локальный узел U-PE на VPLS. Такой PE, естественно, не поддерживает коммутации PW.

Если каждый узел PE в VPLS поддерживает нераспределенную VPLS, т. е. анонсирует себя как N-PE с одним локальным U-PE, результирующая сигнализация будет совпадать с описанной в параграфе 3.2.3. Сигнализация.

3.5.4. Сращивание и плоскость данных

Сращивание двух PW вместе достаточно просто в плоскости данных MPLS, поскольку перемещение пакета из одного PW напрямую в другой является простой операцией замены для метки PW. Когда PW состоит из двух или более срощенных PW, предполагается, что данные пойдут к узлу, где организовано сращивание, т. е. путь данных будет проходить через узлы, участвующие в сигнализации PW.

Дополнительные сведения о сращивании приведены в [RFC6073].

4. Работа в разных AS

Механизмы обеспечения, автоматического обнаружения и сигнализации, описанные выше, можно применить в среде с разными автономными системами (AS). Как и в [RFC4364], имеется много вариантов работы в такой среде.

4.1. Распределение L2VPN NLRI через несколько этапов EBGP

Этот вариант более всего похож на (c) из [RFC4364], т. е. применяется распространение EBGP (External BGP) через несколько интервалов (hop) для L2VPN NLRI между исходной и целевой AS с передачей по EBGP помеченных маршрутов IPv4 или IPv6 из данной AS в соседнюю.

Граничный маршрутизатор AS (Autonomous System Border Router или ASBR) должен поддерживать помеченные маршруты IPv4 /32 (или IPv6 /128) к маршрутизаторам PE в его AS. Он применяет EBGP для распространения этих маршрутов в другие AS, указывая в них себя как BGP next hop. Маршрутизаторы ASBR в любой транзитной AS также используют EBGP для передачи помеченных маршрутов /32 (или /128). Это ведёт к созданию набора путей с коммутацией по меткам от всех входных маршрутизаторов PE ко всем выходным маршрутизаторам PE. В результате маршрутизаторы PE в разных AS могут создавать многоэтапные (multi-hop) соединения EBGP между собой и обмениваться L2VPN NLRI через эти соединения. После такого обмена пары PE в разных AS могут организовать между собой сессии LDP для сигнализации PW.

Для VPLS анонсы BGP и сигнализация PW соответствуют параграфу 3.2. В результате создания многоэтапной сессии EBGP между исходной и целевой AS узлы PE в одной AS, имеющие VSI той или иной VPLS, будут обнаруживать PE в другой AS с теми же VSI из той же VPLS. Эти PEs смогут организовать подходящий сеанс протокола сигнализации PW и создать полную связность (full mesh) псевдопроводов VSI-VSI для построения сети VPLS, как описано в параграфе 3.2.3. Сигнализация.

Для VPWS анонсы BGP и сигнализация PW соответствуют параграфу 3.3. В результате создания многоэтапной сессии EBGP между исходной и целевой AS узлы PE в одной AS, имеющие пулы того или иного цвета (VPN) , будут обнаруживать PE в другой AS с пулами того же цвета. Эти PEs смогут организовать подходящий сеанс протокола сигнализации PW и создать полную связность (full mesh) псевдопроводов, как описано в параграфе 3.2.3. Сигнализация. Аналогичным путём можно создать частичную связность по процедурам параграфа 3.4.

Как и в L3 VPN, создание L2VPN, охватывающей сети нескольких провайдеров, требует некоторой координации использования RT и RD. Этот вопрос рассматривается в параграфе 4.4. Назначение RT и RD.

4.2. Распространение L2VPN NLRI для EBGP с многосегментными PW

Возможным недостатком подхода из предыдущего параграфа является организация сигнальных сессий PW между всеми PE в данной сети L2VPN (VPLS или VPWS). Это ведёт к потенциально большому числу сессий LDP или L2TPv3 через границу AS и участие в этих сеансах большого числа устройств внутри AS. В случае AS, относящихся к разным провайдерам, можно предположить, что провайдеры захотят сократить число сигнальных сессий через границы AS и ограничить число устройств, участвующих в этих сеансах. Кроме того, принудительное завершение сигнальных сессий LDP или L2TPv3 на меньшем количестве ASBR, позволит провайдеру применять стандартные процедуры аутентификации для меньшего числа межпровайдерских сессий. Эти опасения стали мотивом разработки предложенного ниже решения.

В [RFC6073] описан подход для «коммутации» пакетов из одного псевдопровода в другой на конкретном узле. Это позволяет создать сквозной, многосегментный псевдопровод из нескольких сегментов псевдопроводов без поддержки сквозного управляющего соединения. Здесь этот подход служит для работы в нескольких AS, подобно варианту (b) из [RFC4364].

В этой модели применяется распространение EBGP для L2VPN NLRI из AS в соседнюю AS. Сначала маршрутизаторы PE используют IBGP (Internal BGP) для распространения L2VPN NLRI маршрутизатору ASBR или рефлектору маршрутов, клиентом которого является ASBR. Затем ASBR использует EBGP для распространения этих L2VPN NLRI маршрутизатору ASBR в другой AS, который распространяет их маршрутизаторам PE в своей AS или другому ASBR, который, в свою очередь, распространяет их и т. д.

В этом случае PE может узнать адрес ASBR, через который доступен другой PE, с которым нужно организовать PW. Локальный local PE будет получать анонсы BGP с записью L2VPN NLRI, соответствующей экземпляру L2VPNЭ в котором локальный PE имеет подключённые элементы. BGP next-hop в L2VPN NLRI будет ASBR локальной AS. Затем вместо организации управляющего соединения с удаленным PE локальный PE создаст соединение с ASBR. После

этого можно организовать сегмент псевдопровода от PE к ASBR. Маршрутизатор ASBR может создать PW к ASBR в следующей AS и срастить его с PW от PE, как описано в параграфе 3.5.4 и [RFC6073]. Повторение процесса на каждом ASBR создаёт цепочку сегментов PW, которые посре сращивания соединяют два PE.

Отметим, что в этом случае локальный PE может не узнать IP-адрес удаленного PE. Он знает L2VPN NLRI от удалённого PE, где может не быть адреса удалённого PE, и IP-адрес ASBR, который является BGP next-hop для NLRI.

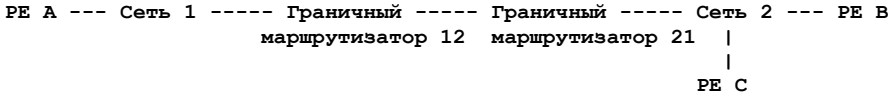
Использование этого подхода для VPLS или полносвязной VPWS ведёт к полносвязной сети псевдопроводов между PE как и в предыдущем параграфе, но не требует полной связности для управляющих соединений (сессии LDP или L2TPv3). Вместо этого управляющие соединения внутри AS организуются между всеми PE этой AS и маршрутизаторами ASBR в ней. Одно управляющее соединение между ASBR смежных AS может служить для поддержки множества псевдопроводных сегментов между AS.

Отметим, что описанные здесь процедуры ведут к совмещению точек сращивания (PW Switching PE или S-PE) в терминологии [RFC5659]) с маршрутизаторами ASBR. Между данной парой AS возможно несколько соединений ASBR-ASBR и в этом случае PE может выбирать доступные ASBR на основе ряда критериев, таких как метрика IGP, локальная конфигурация и т. п., аналогично выбору точек выхода при обычно маршрутизации IP. Использование нескольких ASBR повышает отказоустойчивость (в масштабе времени схождения маршрутов BGP), поскольку PE может выбрать другой маршрутизатор ASBR при отказе используемого.

Как и в L3 VPN, создание L2VPN через сети нескольких провайдеров требует некоторой координации использования RT и RD. Этот вопрос рассмотрен в параграфе 4.4. Назначение RT и RD.

4.3. Межпровайдерское применение распределенной сигнализации VPLS

Дополнительный вариант межпровайдерского решения VPLS можно вывести из описанного выше подхода Distributed VPLS. Рассмотрим показанную на рисунке топологию.



Здесь A, B и C являются PE в одной VPLS, а сети 1 и 2 принадлежат разным сервис-провайдерам. Граничный маршрутизатор 12 (BR12) соединяет сеть 1 с сетью 2, а граничный маршрутизатор 21 (BR21) - сеть 2 с сетью 1. Предположим, что PE не являются «распределенными», т. е. каждый выполняет функции U-PE и N-PE. В этой топологии нужны лишь 2 псевдопровода между провайдерами - A-B и A-C.

Предположим, что сервис-провайдер по какой-то причине решил, что он не хочет, чтобы каждый из его PE имел управляющее соединение с любыми PE в другой сети. Взамен он хочет организовать межпровайдерское управляющее соединение между парой граничных маршрутизаторов. Это можно реализовать, используя методы из параграфа 3.5, где PE ведут себя подобно U-PE, а BR - подобно N-PE. В этом примере PE A будет вести себя как U-PE, локально подключённый к BR12, PE B и C - как U-PE, подключённые локально к BR21, а оба BR - подобно N-PE.

В результате PW от A к B будет состоять из 3 сегментов - A-BR12, BR12-BR21, BR21-B. Граничные маршрутизаторы будут сращивать соответствующие сегменты псевдопроводов.

Это требует нумерации PE внутри VPLS от 1 до n.

4.4. Назначение RT и RD

Отметим, что для корректной работы в разных AS по любой из описанных выше процедур в AS должны применяться согласованные значения RT и RD, как в L3 VPN [RFC4364]. Структура RT и RD делает риск случайных конфликтов небольшим. Основная проблема заключается в том, что оператору одной AS нужно знать RT в другой AS, выбранные для любой VPN, имеющей сайты в обеих AS. Как и в L3 VPN, имеется много способов сделать это, но все они требуют координации между провайдерами. Например, провайдер A может пометить все NLRI для данной сети VPN одним RT, скажем, RT_A, а затем провайдер B может настроить PE, подключённые к сайтам этой VPN для импорта NLRI, содержащих это значение RT. Провайдер B может выбрать своё значение RT (RT_B), помечая все NLRI для этой VPN данным RT и тогда провайдер A может импортировать эти NLRI с RT в соответствующих PE. Однако это требует от провайдеров обменяться своими значениями RT для каждой VPN. Провайдеры могут также договориться об использовании общего RT для данной VPN. В любом случае важно обменяться значениями RT между провайдерами. Как и в L3 VPN, провайдеры могут настроить фильтрацию RT, чтобы через границу AS проходили только согласованные значения RT.

Отметим, что требуется один идентификатор VPN (передаётся в BGP Extended Community) для каждого экземпляра VPLS или VPWS. Правила кодирования этих идентификаторов [RFC4360] обеспчивают отсутствие конфликтов с другими провайдерами. Однако для одного экземпляра VPLS или VPWS, охватывающего сети двух или более провайдеров, один провайдер должен назначить идентификатор и сообщить его другим, а те должны использовать это значение для сайтов того же экземпляра VPLS или VPWS.

5. Вопросы безопасности

В этом документе рассмотрено множество моделей предоставления услуг L2VPN и заданы идентификаторы конечных точек, требуемые для поддержки каждой из предложенных моделей. Указано также, как эти идентификаторы конечных точек сопоставляются с полями протоколов автоматического обнаружения и сигнализации.

Вопросы безопасности, связанные с сигнальными протоколами, рассмотрены в спецификациях соответствующих протоколов ([RFC5036], [RFC4447], [RFC3931], [RFC4667]).

Вопросы безопасности автоматического обнаружения на основе BGP, в том числе для работы в разных AS, рассмотрены в [RFC4364]. L2VPN использующие автоматическое обнаружение на основе BGP, могут автоматизировать и установку механизмов защиты. Задание автоматизированных механизмов защиты выходит за рамки этого документа и рекомендуется для будущих работ.

Вопросы безопасности, связанные с конкретными услугами L2VPN. Рассмотрены в [RFC4664], [RFC4665] и [RFC4762].

Способ сопоставления идентификаторов конечных точек с полями протоколов не создаёт дополнительных проблем безопасности.

6. Взаимодействие с IANA

Агентство IANA выделило AFI и SAFI для L2VPN NLRI, имеющие те же значения, которые заданы в [RFC4761], т. е. AFI = 25 (L2VPN) и SAFI = 65 (уже выделено для VPLS). Одни значения AFI и SAFI применяются для автоматического обнаружения VPLS и VPWS, описанного в этом документе.

В [RFC4446] заданы реестры Attachment Group Identifier (AGI) Type и Attachment Individual Identifier (AII) Type. Тип 1 в каждом из реестров назначен для форматов AGI и AII, определённых в этом документе.

Агентство IANA выделило два новых кода статуса LDP. IANA уже поддерживает реестр STATUS CODE NAME SPACE, заданный в [RFC5036]. В нем выделены указанные ниже значения.

```
0x00000030 Устройство присоединения связано с другим PE
0x0000002D Устройство присоединения связано с другим удаленным AC
```

Зарегистрированы два новых L2TP Result Code для сообщений CDN. IANA поддерживает реестр L2TP Result Code Values for the CDN message, заданный в [RFC3438]. В нем выделены указанные ниже значения.

```
   27: Устройство присоединения связано с другим PE
   28: Устройство присоединения связано с другим удаленным AC
```

В [RFC4360] задан реестр Two-octet AS Specific Extended Community, в котором агентство IANA выделило значение из «переходного» диапазона (0x0000-0x00FF), указанное ниже.

```
0x000A Связанный с 2-октетной AS идентификатор L2 VPN
```

В [RFC4360] задан реестр IPv4 Address Specific Extended Community, в котором агентство IANA выделило значение из «переходного» диапазона (0x0100-0x01FF), указанное ниже.

```
0x010A Идентификатор L2 VPN
```

7. Совместимость BGP-AD и VPLS-BGP

BGP-AD и VPLS-BGP [RFC4761] используют одни AFI и SAFI. Чтобы BGP-AD и VPLS-BGP могди работать вместе, размер NLRI должен применяться как демультиплексор. BGP-AD NLRI имеет размер 12 байтов и содержит 8-байтовое значение RD и 4-байтовое значение VSI-ID. В VPLS-BGP [RFC4761] применяются 17-батовый NLRI. Поэтому реализации BGP-AD должны игнорировать NLRI размером больше 12 байтов.

8. Благодарности

Спасибо Dan Tappan, Ted Qian, Ali Sajassi, Skip Booth, Luca Martini, Dave McDysan, Francois Le Faucheur, Russ Gardo, Keyur Patel, Sam Henderson, Matthew Bocci за комментарии, критику и полезные предложения.

Спасибо Tissa Senevirathne, Hamid Ould-Brahim, Yakov Rekhter за обсуждение вопросов автоматического обнаружения.

Спасибо Vach Kompella за обсуждение подходящей для обобщённых идентификаторов семантики.

9. Литература

9.1. Нормативные документы

[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. Дополнительная литература

[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.

Адреса авторов

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

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru