Internet Engineering Task Force (IETF)
Request for Comments: 7432
Category: Standards Track
ISSN: 2070-1721

A. Sajassi, Ed.
Cisco
R. Aggarwal
Arktan
N. Bitar
Verizon
A. Isaac
Bloomberg
J. Uttaro
AT&T
J. Drake
Juniper Networks
W. Henderickx
Alcatel-Lucent
February 2015

BGP MPLS-Based Ethernet VPN

Ethernet VPN на основе BGP MPLS

Аннотация

Этот документ описывает процедуры для BGP MPLS Ethernet VPN (EVPN). Эти процедуры соответствуют требованиям RFC 7209 Requirements for Ethernet VPN (EVPN).

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/rfc7432.

Оглавление

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

1. Введение

Виртуальные частные ЛВС (VPLS), определённые в [RFC4664], [RFC4761] и [RFC4762], являются проверенной и широко развёрнутой технологией. Однако у имеющихся решений есть много ограничений в части многодомности и резервирования, оптимизации групповой передачи, простоты предоставления, распределения нагрузки по потокам и поддержки множества путей. Эти ограничения важны для центров обработки данных (ЦОД - Data Center или DC). В [RFC7209] описаны мотивы для новых решений, преодолевающих эти ограничения, а также приведены требования к новым решениям.

В этом документе описаны процедуры для решения на основе BGP MPLS, названного Ethernet VPN (EVPN) и соответствующего требованиям [RFC7209]. Следует прочесть [RFC7209], где требования и мотивы описаны подробно. Для EVPN нужны расширения имеющихся протоколов IP/MPLS, описанные в этом документе. В дополнение к этому EVPN использует некоторые «строительные блоки» имеющихся технологий MPLS.

2. Уровни требований

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].

3. Терминология

Broadcast Domain - широковещательный домен
: В сети с мостами широковещательный домен соответствует виртуальной ЛВС (Virtual LAN или VLAN), которая обычно представлена одним VLAN ID (VID), но может представляться несколькими VID при использовании обучения SVL (Shared VLAN Learning) в соответствии с [802.1Q].

Bridge Table - таблица моста
: Широковещательный домен в MAC-VRF.

CE
: Краевое устройство клиента, например, хост, маршрутизатор или коммутатор.

EVI
: Экземпляр EVPN, охватывающий краевые устройства провайдера (Provider Edge или PE), участвующие в EVPN.

MAC-VRF
: Таблица виртуальной маршрутизации и пересылки для MAC-адресов в PE.

Ethernet Segment (ES) - сегмент Ethernet
: При подключении клиентской стороны (устройство или сеть) к одному или нескольким PE через набор каналов Ethernet такой набор называют сегментов Ethernet.

Ethernet Segment Identifier (ESI) - идентификатор сегмента Ethernet
: Уникальный идентификатор, отличный от 0 и указывающий сегмент Ethernet.

Ethernet Tag - тег Ethernet
: Тег Ethernet указывает конкретный домен широковещания, например, VLAN. Экземпляр EVPN включает один или несколько доменой широковещания.

LACP
: Link Aggregation Control Protocol - протокол управления агрегированием каналов.

MP2MP
: Multipoint to Multipoint - множество со множеством.

MP2P
: Multipoint to Point - множество с одним.

P2MP
: Point to Multipoint - один со множеством.

P2P
: Point to Point - точка-точка (один с одним).

PE
: Крайвое устройство провайдера (Provider Edge).

Single-Active Redundancy Mode
: Когда лишь одному из устройств PE, подключённых к сегменту Ethernet, разрешено пересылать трафик в сегмент Ethernet и из него для данной VLAN, такой режим резервирования называют Single-Active (активен один).

All-Active Redundancy Mode
: Когда всем устройствам PE, подключённым к сегменту Ethernet, разрешено пересылать трафик в сегмент Ethernet и из него для данной VLAN, такой режим резервирования называют All-Active (активны все).

4. Обзор EVPN на основе BGP MPLS

В этом разделе представлен обзор EVPN. Экземпляр EVPN включает краевые устройства клиента (CE), соединённые с краевыми устройствами провайдера (PE), образующими границу инфраструктуры MPLS. CE может быть хостом, маршрутизатором или коммутатором. PE обеспечивают связность через виртуальные мосты L2 между CE. В сети провайдера может существовать множество экземпляров EVPN.

PE могут соединяться с инфраструктурой MPLS LSP, обеспечивающей преимущества технологии MPLS, такие как быстрая перемаршрутизация, отказоустойчивость и т. п. PE могут также соединяться с инфраструктурой IP и в этом случае между ними может применяться туннелировани IP/GRE (Generic Routing Encapsulation) или иные туннели IP. В документе подробно описаны процедуры лишь для туннелей MPLS LSP, однако их можно расширить на туннели IP в качестве туннелирования в сети с коммутацией пакетов (Packet Switched Network или PSN).

В EVPN изучение MAC между PE происходит не в плоскости данных (как с традиционными мостами в [RFC4761] [RFC4762]), а в плоскости управления. Это обучение обеспечивает лучший контроль процесса изучения MAC, включая указание, кто и что может изучать, а также возможность применения правил. Кроме того, плоскостью управления для анонсирования доступности MAC является многопротокольное (multi-protocol или MP) расширение BGP (как в IP VPN [RFC4364]). Это обеспечивает гибкость и возможность сохранять «виртуализацию» или изоляцию групп взаимодействующих агентов (хосты, серверы, виртуальные машины) друг от друга. В EVPN узлы PE анонсируют MAC-адреса, узнанные от подключённых к ним CE, вместе с меткой MPLS другим PE в плоскости управления с использованием Multiprotocol BGP (MP-BGP). Обучение в плоскости управления позволяет распределять трафик CE, подключённых к нескольким PE. Это дополняет распределение нагрузки через ядро MPLS по нескольким LSP между парой PE. Иными словами, это позволяет CE подключаться к нескольким активным точкам присоединения и улучшает время схождения при некоторых отказах в сети.

Однако обучение между PE и CE использует наиболее подходящий для CE метод - обучение в плоскости данных, IEEE 802.1x, протокол обнаружения канального уровня (Link Layer Discovery Protocol или LLDP), IEEE 802.1aq, ARP, плоскость управления сетью или иные протоколы.

Решение вопроса о заполнении таблицы пересылки L2 в PE всеми MAC-адресами получателей, известными плоскости управления, или реализации в PE решения на основе кэширования принимается локально. Например, таблица пересылки MAC может заполняться только MAC получателей из активных потоков, проходящих через PE.

Атрибуты правил EVPN очень похожи на применяемые в IP-VPN. Для экземпляра EVPN требуется отличитель маршрута (Route Distinguisher или RD), уникальный в масштабе MAC-VRF и 1 или несколько глобально уникальных целей маршрута (Route Target или RT). CE подключается к MAC-VRF на PE через интерфейс Ethernet, на котором может быть задан 1 или несколько тегов Ethernet, например, VLAN ID. Некоторые варианты развёртывания гарантируют уникальность VLAN ID среди экземпляров EVPN - все точки присоединения к данному экземпляру EVPN используют один VLAN ID, а другие экземпляры EVPN этот VLAN ID не применяют. В этом документе данный случай называется Unique VLAN EVPN и для него описаны упрощенные процедуры оптимизации.

5. Сегмент Ethernet

Как показано в [RFC7209], каждому сегменту Ethernet требуется уникальный идентификатор в EVPN. В этом разделе описано назначение идентификаторов и их кодирование для сигнализации EVPN. Далее описываются протокольные механизмы для работы с идентификаторами.

Когда сайт клиента подключён к одному или нескольким PE через набор каналов Ethernet, этот набор каналов называют сегментом Ethernet. Для многодомного сайта каждый сегмент Ethernet (ES) указывается уникальным ненулевым идентификатором (Ethernet Segment Identifier или ESI). ESI представляется 10-октетным целым числом в формате линии и первым передаётся старший октет. Указанные ниже значения ESI являются резервными.

- ESI 0 указывает однодомный сайт.
- ESI {0xFF} (повтор 10 раз) называется MAX-ESI.

В общем случае сегменту Ethernet **следует** иметь нерезервное значение ESI, уникальное в сети (т. е. во всех экземплярах EVPN на всех PE). Если CE сегмента Ethernet управляются оператором сети, следует гарантировать уникальность ESI, однако при неуправляемых CE оператор **должен** настроить уникальные ESI для этого сегмента Ethernet. Это нужно для автоматического обнаружения сегментов Ethernet и выбора назначенного узла пересылки (Designated Forwarder или DF).

В сети с управляемыми и неуправляемыми CE значение ESI имеет показанный ниже формат.

```
+---+---+---+---+---+---+---+---+---+---+
| T |          ESI Value                |
+---+---+---+---+---+---+---+---+---+---+
```

T (ESI Type)

1-октетное поле (старший октет), задающее формат остальных 9 октетов (ESI Value). Используемые 6 типов ESI описаны ниже.

Type 0 (T=0x00)

Указывает произвольное 9-октетное значение ESI, управляемое и настраиваемое оператором.

Type 1 (T=0x01)

При использовании IEEE 802.1AX LACP между PE и CE этот тип ESI указывает автоматически создаваемые ESI, определяемые из LACP путём конкатенации указанных ниже параметров.

- Адрес CE LACP System MAC (6 октетов) **должен** помещаться в 6 старших октетов поля ESI Value.
- Ключ CE LACP Port Key (2 октета) **должен** помещаться в 2 октета после адреса System MAC.
- Остающийся октет имеет значение 0x00.

Что касается CE, он будет рассматривать множество PE, к которым он подключён, как один коммутатор. Это позволяет CE агрегировать каналы, подключённые к разным PE, в одну группу (bundle).

Этот механизм можно применять лишь в случаях, когда ESI создаются в соответствии с приведёнными выше требованиями к уникальности.

Type 2 (T=0x02)

Этот тип применяется в случае подключения хостов через ЛВС с мостами между CE PE. ESI Value генерируется автоматически на основе протокола мостов L2 - при использовании протокола MSTP (Multiple Spanning Tree Protocol) в ЛВС на основе мостов значение ESI выводится путём прослушивания PDU мостов (Bridge PDU или BPDU) в сегменте Ethernet. Для этого на PE не требуется запускать MSTP, однако PE должен узнать MAC-адрес корневого моста (Root Bridge) и его приоритет (Bridge Priority) из внутреннего связующего дерева (Internal Spanning Tree или IST), прослушивая BPDU. Состав ESI Value описан ниже.

- MAC-адрес Root Bridge (6 октетов) **должен** помещаться в 6 старших октетов поля ESI Value.
- Root Bridge Priority (2 октета) **должен** помещаться в 2 октета после MAC-адреса Root Bridge.
- Остающийся октет имеет значение 0x00.

Этот механизм можно применять лишь в случаях, когда ESI соответствуют приведённым выше требованиям к уникальности.

Type 3 (T=0x03)

Этот тип указывает ESI Value на основе MAC, созданное автоматически или заданное оператором.

- Адрес System MAC (6 октетов). MAC-адрес PE **должен** помещаться в 6 старших октетов поля ESI Value.
- Значение Local Discriminator (3 октета) **должно** помещаться в 3 младших октета ESI Value.

Этот механизм можно применять лишь в случаях, когда ESI создаются в соответствии с приведёнными выше требованиями к уникальности.

Type 4 (T=0x04)

Значение ESI Value на основе router-ID, созданное автоматически или заданное оператором.

- Идентификатор Router ID (4 октета) **должен** помещаться в 4 старших октета поля ESI Value.
- Значение Local Discriminator (4 октета) **должно** помещаться в 4 октета после Router ID[1].
- Остающийся октет имеет значение 0x00.

Этот механизм можно применять лишь в случаях, когда ESI создаются в соответствии с приведёнными выше требованиями к уникальности.

Type 5 (T=0x05)

Значение ESI Value на основе номера автономной системы (Autonomous System или AS), созданное автоматически или заданное оператором.

- Номер AS (4 октета), принадлежащий системе, **должен** помещаться в 4 старших октета поля ESI Value. Если используется 2-октетный номер AS, 2 старших октета имеют значение 0x0000.

[1] В исходном документе вместо Router ID ошибочно сказано «адреса IP). См. https://www.rfc-editor.org/errata/eid5718. *Прим. перев.*

- Значение Local Discriminator (4 октета) **должно** помещаться в 4 октета после номера AS.
- Остающийся октет имеет значение 0x00.

Этот механизм можно применять лишь в случаях, когда ESI создаются в соответствии с приведёнными выше требованиями к уникальности.

6. Ethernet Tag ID

Ethernet Tag ID - 32-битовое поле, содержащее 12- или 24-битовый идентификатор, указывающий конкретный домен широковещания (например, VLAN) экземпляра EVPN. 12-битовый идентификатор называется VLAN ID (VID). Экземпляр EVPN содержит один или несколько доменов широковещания (одну или несколько VLAN). VLAN для данного экземпляра EVPN назначает поставщик услуг EVPN. Данная сеть VLAN может сама по себе представляться множеством VID. В таких случаях, узлы PE этой сети VLAN для данного экземпляра EVPN отвечают за трансляцию VLAN ID от локально подключённых устройств CE и к ним.

Если VLAN представляется одним VID на всех устройствах PE, участвующих в этой сети VLAN для экземпляра EVPN, трансляция VID на PE не требуется. Кроме того, в некоторых вариантах развёртывания гарантируется уникальность VID во всех экземплярах EVPN - все точки присоединения данного экземпляра EVPN используют одно значение VID, а в других экземплярах EVPN это значение не применяется. Это позволяет выводить RT для каждого экземпляра EVPN автоматически из соответствующего VID, как описано в параграфе 7.10.1. Автоматический вывод из Tag ID.

В последующих параграфах рассматривается связь между доменами широковещания (например, VLAN), Ethernet Tag ID (например, VID) и MAC-VRF, а также установка Ethernet Tag ID в разных маршрутах EVPN BGP (определены в разделе 8) для разных типов сервисных интерфейсов, описанных в [RFC7209].

Значение Ethernet Tag ID = 0xFFFFFFFF зарезервировано и называется MAX-ET.

6.1. Интерфейс сервиса на основе VLAN

С таким интерфейсом службы экземпляр EVPN лишь из одного домена широковещания (например, одной VLAN). Поэтому между VID и MAC-VRF имеется взаимно однозначное соответствие. Поскольку MAC-VRF соответствует одной сети VLAN, он состоит из одной таблицы моста, соответствующей VLAN. Если VLAN представляется множеством VID (например, свой VID на сегмент Ethernet и на PE), каждому PE требуется выполнять трансляцию VID для кадров, направленных в его сегменты. В таких вариантах кадры, передаваемые через сеть MPLS/IP, следует оставлять помеченными исходным VID, а трансляция VID **должна** поддерживаться в пути данных и **должна** выполняться на соответствующем PE. Ethernet Tag ID на всех маршрутах EVPN **должен** иметь значение 0.

6.2. Интерфейс сервиса VLAN Bundle

С таким интерфейсом экземпляр EVPN соответствует нескольким доменам широковещания (например, VLAN), однако поддерживается лишь одна таблица моста на MAC-VRF и ей пользуется несколько VLAN. Это предполагает, что MAC-адреса должны быть уникальными во всех VLAN для этого EVI, чтобы служба могла работать. Иными словами, существует сопоставление «множество в один» между VLAN и MAC-VRF, состоящей из одной таблицы моста. Кроме того, одна сеть VLAN должна представляться одним VID (например, не разрешается трансляция VID для этого типа интерфейса службы). Кадры с инкапсуляцией MPLS должны оставаться помеченными исходным VID. Трансляция не разрешается. Для Ethernet Tag ID на всех маршрутах EVPN должно быть установлено значение 0.

6.2.1. Интерфейс сервиса на базе порта

Этот интерфейс службы является особым случаем интерфейса VLAN bundle, где все VLAN на порту относятся к одной службе и отображаются на одну связку (bundle). Процедуры идентичны описанным в параграфе 6.2.

6.3. Интерфейс службы VLAN-Aware Bundle

С таким интерфейсом экземпляр EVPN содержит несколько доменов широковещания (например, VLAN) и свою таблицу моста для каждой сети VLAN, т. е. поддерживается множество (по одной на VLAN) таблиц на одну MAC-VRF для экземпляра EVPN.

Групповой, широковещательный и трафик неизвестных индивидуальных получателей (Broadcast, unknown unicast, or multicast или BUM) передаётся только узлам CE данного домена широковещания. Однако широковещательные домены в EVI **могут** иметь свой P-Tunnel или использовать общие P-туннели (например, один туннель на всех в EVI.

Когда VLAN представляется одним VID и трансляции VID не требуется, инкапсулированные в MPLS пакеты **должны** содержать это значение VID. Для Ethernet Tag ID на всех маршрутах EVPN **должно** быть установлено это значение VID. Анонсирующий узел PE **может** объявлять MPLS Label1 в анонсе маршрута MAC/IP **только** EVI или Ethernet Tag ID и EVI. Решение принимается локально анонсирующим PE (PE размещения) и не влияет на другие PE.

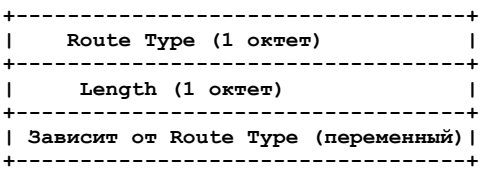
Когда VLAN представляется разными VID на разных CE и требуется трансляция VID, в каждом маршруте EVPN BGP **должен** передаваться нормализованный Ethernet Tag ID (VID). Кроме того, анонсирующий PE объявляет MPLS Label1 в анонсе маршрута MAC/IP, представляющем Ethernet Tag ID и EVI, так что при получении пакета с инкапсуляцией MPLS он может идентифицировать соответствующую таблицу моста по метке MPLS EVPN и выполнять трансляцию Ethernet Tag ID **только** на PE размещения, т. е. кадры Ethernet, передаваемые через сеть MPLS/IP **должны** сохранять исходный идентификатор VID, а трансляция VID выполняется на PE размещения. Для Ethernet Tag ID во всех маршрутах EVPN **должно** устанавливаться нормализованное значение Ethernet Tag ID, выданное провайдером EVPN.

6.3.1. Интерфейс службы на основе портов с поддержкой VLAN

Этот интерфейс службы является особым случаем интерфейса VLAN-aware bundle, где все VLAN на порту относятся к одному экземпляру сервиса и отображены на одну связку, но без трансляции VID. Процедуры идентичны описанным в параграфе 6.3.

7. Маршруты BGP EVPN

Этот документ определяет новое поле BGP NLRI (Network Layer Reachability Information - сведения о доступности на сетевом уровне) - EVPN NLRI.



Route Type определяет кодирование остальной части EVPN NLRI.

Поле Length указывает число октетов в зависящем от Route Type поле EVPN NLRI.

Этот документ определяет 4 типа маршрутов:

1 - Ethernet Auto-Discovery (A-D);
2 - MAC/IP Advertisement;
3 - Inclusive Multicast Ethernet Tag;
4 - Ethernet Segment.

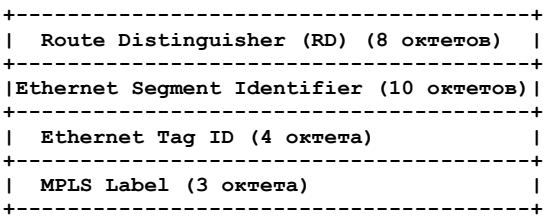
Подробное описание и процедуры для каждого типа описаны в последующих параграфах.

EVPN NLRI передаётся в BGP [RFC4271] с использованием расширений BGP Multiprotocol [RFC4760] с идентификатором семейства адресов (Address Family Identifier или AFI) 25 (L2VPN) и следующим идентификатором семейства адресов (Subsequent Address Family Identifier или SAFI) 70 (EVPN). Поле NLRI в атрибуте MP_REACH_NLRI/MP_UNREACH_NLRI содержит EVPN NLRI с указанным выше кодированием.

Чтобы два узла BGP обменивались помеченными EVPN NLRI, они должны использовать анонсы возможностей BGP (Capabilities Advertisement), подтверждающие способность обоих обрабатывать такие NLRI. Это делается в соответствии с [RFC4760] путём указания кода возможности 1 (multiprotocol BGP) с AFI 25 (L2VPN) и SAFI 70 (EVPN).

7.1. Ethernet Auto-discovery

EVPN NLRI для маршрутов типа Ethernet A-D имеет вид



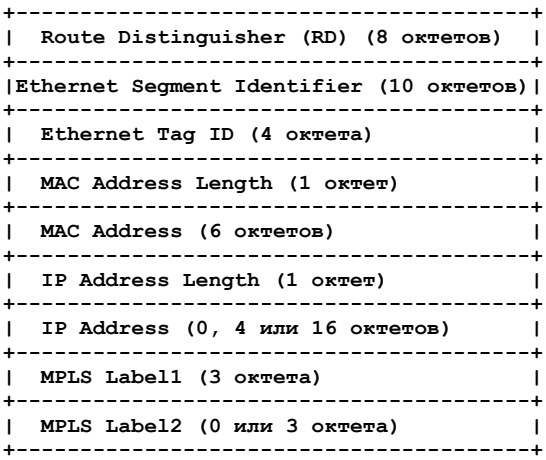
При обработке ключей маршрута BGP только поля Ethernet Segment Identifier и Ethernet Tag ID считаются частью префикса в NLRI. Поле MPLS Label трактуется как атрибут маршрута, а не его часть.

Процедуры и использование таких маршрутов описаны в параграфах 8.2. Быстрое схождение и 8.4. Псевдонимы и резервный путь.

Значение 20-битовой метки MPLS помещается в старшие биты 3-октетного поля MPLS Label.[1]

7.2. MAC/IP Advertisement

EVPN NLRI для маршрутов типа MAC/IP Advertisement имеет вид



При обработке ключей маршрута BGP только поля Ethernet Tag ID, MAC Address Length, MAC Address, IP Address Length и IP Address считаются частью префикса в NLRI. Поля Ethernet Segment Identifier, MPLS Label1, MPLS Label2 трактуется как атрибут маршрута, а не его часть. Размеры адресов IP и MAC указываются в битах.

Процедуры и использование таких маршрутов описаны в разделах 9. Определение доступности индивидуального MAC-адреса и 14. Распределение нагрузки для индивидуальных пакетов.

Значение 20-битовой метки MPLS помещается в старшие биты 3-октетных полей MPLS Label1 и MPLS Label2.[1]

[1] В исходном документе этого предложения не было. См. https://www.rfc-editor.org/errata/eid5554. *Прим. перев.*

7.3. Inclusive Multicast Ethernet Tag

EVPN NLRI для маршрутов типа Inclusive Multicast Ethernet Tag имеет вид

```
+---------------------------------------+
|  RD (8 октетов)                       |
+---------------------------------------+
|  Ethernet Tag ID (4 октета)           |
+---------------------------------------+
|  IP Address Length (1 октет)          |
+---------------------------------------+
|  Originating Router's IP Address      |
|          (4 или 16 октетов)           |
+---------------------------------------+
```

Процедуры и использование таких маршрутов описаны в разделах 11. Обработка трафика с множеством получателей, 12. Обработка индивидуальных пакетов неизвестных получателей и 16. Групповая передача и широковещание. Размер адреса IP указывается в битах. При обработке ключей маршрута BGP только поля Ethernet Tag ID, IP Address Length и Originating Router's IP Address считаются частью префикса в NLRI.

7.4. Ethernet Segment

EVPN NLRI для маршрутов типа Ethernet Segment имеет вид

```
+---------------------------------------+
|  RD (8 октетов)                       |
+---------------------------------------+
|Ethernet Segment Identifier (10 октетов)|
+---------------------------------------+
|  IP Address Length (1 октет)          |
+---------------------------------------+
|  Originating Router's IP Address      |
|          (4 или 16 октетов)           |
+---------------------------------------+
```

Процедуры и использование таких маршрутов описаны в параграфе 8.5. Выбор DF. Размер адреса IP указывается в битах. При обработке ключей маршрута BGP только поля Ethernet Segment ID, IP Address Length и Originating Router's IP Address считаются частью префикса в NLRI.

7.5. ESI Label Extended Community

Эта расширенная группа является новой переходной группой с полями Type = 0x06 и Sub-Type = 0x01. Она может анонсироваться с маршрутами Ethernet Auto-discovery и разрешает процедуры расщепления горизонта (split-horizon) для многодомных сайтов, как указано в параграфе 8.3. Расщепление горизонта. Поле ESI Label представляет ES анонсирующим PE и применяется в фильтрации split-horizon другими PE, подключёнными к тому же многодомному сегменту Ethernet.

Каждая расширенная группа ESI Label кодируется 8 октетами, как показано ниже.

```
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=0x06     | Sub-Type=0x01 | Flags(1 октет)|  Reserved=0   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Reserved=0   |           ESI Label                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Младший бит октета Flags определён как Single-Active. Значение 0 указывает, что многодомный сайт работает в режиме избыточности All-Active, а 1 указывает режим Single-Active.

Значение 20-битовой метки MPLS помещается в старшие биты 3-октетного поля ESI Label.[1]

7.6. ES-Import Route Target

Это новая расширенная группа Route Target, передаваемая с маршрутом Ethernet Segment. Она позволяет всем PE одного многодомного сайта импортировать маршруты Ethernet Segment. Значение выводится автоматически для ESI Type 1, 2, 3 путём указания MAC-адреса в 6 старших октетах 9-октетного поля ESI Value в ES-Import Route Target.

```
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    | Type=0x06     | Sub-Type=0x02 |          ES-Import            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     Продолжение ES-Import                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Этот документ добавляет для расширенной группы Route Target значение 0x06 в старшем октета (поле Type) в дополнение к значениям, заданным в [RFC4360]. Значение второго октета (поле Sub-Type) 0x02 указывает, что это расширенная группа (Extended Community) типа Route Target. Новое значение Type = 0x06 указывает, что структура RT содержит 6-октетное значение (например, адрес MAC). Узлы BGP, реализующие RT Constraint [RFC4684], **должны** применять процедуры RT Constraint и для ES-Import RT.

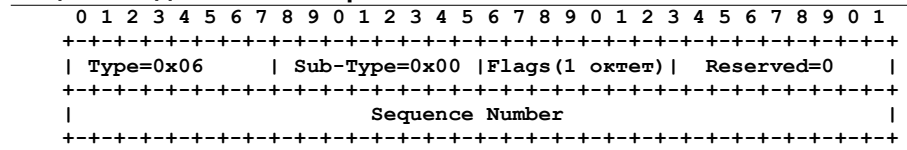
Процедуры и использование этого атрибута описаны в параграфе 8.1. Автоматическое обнаружение многодомных ES.

7.7. Расширенная группа MAC Mobility

Эта новая переходная расширенная группа имеет поля Type = 0x06 и Sub-Type = 0x00 и может анонсироваться с маршрутами MAC/IP Advertisement. Процедуры использования группы описаны в разделе 15. Мобильность MAC.

MAC Mobility кодируется 8-октетным значением, как показано ниже.

[1] В исходном документе этого предложения не было. См. https://www.rfc-editor.org/errata/eid5554. *Прим. перев.*



Младший бит октета Flags определён как флаг Sticky/static и может иметь значение 1, указывающее статический MAC-адрес, который нельзя поменять. Порядковый номер служит для того, чтобы устройства PE сохраняли корректный маршрут MAC/IP Advertisement при наличии нескольких обновлений для одного MAC-адреса.

7.8. Расширенная группа Default Gateway

Расширенная группа Default Gateway относится к типу Opaque (см. параграф 3.3 в [RFC4360]) и является переходной группой (первый октет имеет значение 0x03). Второй октет (Sub-Type) имеет значение 0x0d (Default Gateway), выделенное IANA. Поле Value для этой группы является резервным (отправитель устанавливает 0, получатель игнорирует). Процедуры и применение атрибута описаны в параграфе 10.1. Принятый по умолчанию шлюз.

7.9. Назначение RD для MAC-VRF

Для отличителя маршрута (RD) должно быть установлено значение RD из MAC-VRF, анонсирующей NLRI. Значение RD **должно** назначаться для данной MAC-VRF на устройстве PE и **должно** быть уникальным среди всех MAC-VRF на PE. **Рекомендуется** использовать Type 1 RD [RFC4364]. Поле Value содержит IP-адрес PE (обычно loopback), за которым следует уникальное в масштабе PE число, которое может генерироваться самим PE. В случае уникальной VLAN EVPN младшие 12 битов могут содержать VLAN ID, а старшие 4 - 0.

7.10. Цели маршрутов

Маршрут EVPN **может** включать один или несколько атрибутов цели (Route Target или RT. RT могут настраиваться (как в IP VPN) или выводиться автоматически.

Если PE применяет RT Constraint, узел PE анонсирует все такие RT, используя RT Constraint в соответствии с [RFC4684]. Применение RT Constraint позволяет каждому маршруту EVPN достигать лишь тех PE, которым указан импорт хотя бы одного RT из набора, передаваемого в маршруте EVPN.

7.10.1. Автоматический вывод из Tag ID

Для сценария Unique VLAN EVPN крайне желательно автоматически выводить RT из Ethernet Tag ID (VLAN ID) для данного экземпляра EVPN, как показано ниже.

- В поле Global Administrator для RT **должен** указываться номер автономной системы (AS), с которой связано устройство PE.
- 12-битовое значение VLAN ID **должно** помещаться в 12 младших битов поля Local Administrator, а остальные должны иметь значение 0.

8. Многодомные функции

В этом разделе рассматриваются функции, процедуры и связанные маршруты BGP, используемые для поддержки многодомности в EVPN. Описаны как многодомные устройства (multihomed device или MHD), так и многодомные сети (multihomed network или MHN).

8.1. Автоматическое обнаружение многодомных ES

PE, подключённые к одному сегменту Ethernet, могут автоматически находить друг друга путём обмена маршрутами Ethernet Segment без настройки или с минимальной настройкой.

8.1.1. Создание маршрута Ethernet Segment

Отличитель маршрута (RD) **должен** быть Type 1 RD [RFC4364]. Значение поля включает IP-адрес PE (обычно петлевой), за которым следует уникальный номер PE.

Идентификатор сегмента (ESI) **должен** иметь 10-октетное значение, описанное в разделе 5. Сегмент Ethernet.

Анонсы BGP для маршрутов Ethernet Segment **должны** включать ES-Import Route Target, как указано в параграфе 7.6.

Фильтрация маршрутов Ethernet Segment **должна** выполняться так, чтобы маршруты Ethernet Segment импортировались только устройствами PE, которые многодомны в одном сегменте Ethernet. С этой целью каждый узел PE, подключенный к определённому сегменту Ethernet, создаёт правило фильтрации для импорта маршрута, содержащего цель ES-Import Route Target, созданную из ESI.

8.2. Быстрое схождение

В EVPN доступность MAC-адресов определяется черз плоскость управления BGP в сети MPLS. Поэтому при отсутствии какого-либо механизма быстрой защиты время схождения для сети зависит от числа маршрутов MAC/IP Advertisement, которые должен отозвать столкнувшийся с отказом узел PE. Для больших сред такая схема замедляет схождение.

Для решения проблемы в EVPN определён механизм, позволяющий быстро и эффективно сообщить удаленным узлам PE о необходимости обновления их таблиц пересылки при возникновении отказа связности с сегментом Ethernet. Для этого каждый узел PE анонсирует набор из одного или нескольких Ethernet A-D на маршрут ES для каждого локально подключённого сегмента Ethernet (создание этих маршрутов описано в параграфе 8.2.1. Создание Ethernet A-D для маршрута Ethernet Segment). PE может потребоваться анонсировать более одного Ethernet A-D на маршрут ES для данного сегмента ES, поскольку ES может входить в несколько EVI и RT для всех этих EVI могут не поместиться в один маршрут. Анонсирование нескольких Ethernet A-D на маршрут ES для ES позволяет включать в каждый маршрут часть полного набора RT. Ethernet A-D на маршрут ES в наборе различаются по RD.

При отказе связности с подключённым сегментом PE отзывает соответствующие наборы Ethernet A-D для маршрутов ES. В результате все PE, получившие отзыв, обновляют свои смежности next-hop для всех MAC-адресов, связанных с соответствующим сегментом Ethernet. Если нет PE, анонсирующих Ethernet A-D для того же сегмента, получивший отзыв узел PE просто аннулирует записи MAC для сегмента. В иных случаях PE обновляет смежности next-hop.

8.2.1. Создание Ethernet A-D для маршрута Ethernet Segment

В этом параграфе описаны процедуры, применяемые при создании Ethernet A-D для маршрута ES, используемых для быстрого схождения (см. выше) и анонсирования меток ESI, используемых для фильтрации split-horizon (параграф 8.3. Расщепление горизонта). **Требуется** поддержка этих маршрутов.

Для отличителя маршрута (RD) **должен** указываться Type 1 RD [RFC4364]. Поле включает IP-адрес PE (обычно, loopback), за которым следует число, уникальное в масштабе PE.

В поле Ethernet Segment Identifier **должно** помещаться 10-октетное значение, как указано в разделе 5. Сегмент Ethernet. Маршрут Ethernet A-D не требуется при нулевом значении идентификатора сегмента (однодомный вариант).

В поле Ethernet Tag ID **должно** быть указано MAX-ET.

Метка MPLS в NLRI **должна** иметь значение 0.

В маршрут **должна** включаться расширенная группа ESI Label. Если желателен режим All-Active, в поле флагов Single-Active расширенной группы ESI Label **должно** быть установлено значение 0, а в поле MPLS label этой расширенной группы **должна** помещаться действительная метка MPLS. Метка MPLS в Extended Community называется меткой ESI и она **должна** иметь одинаковое значение во всех Ethernet A-D для маршрута ES, анонсируемых для ES. Это **должна** быть назначенная в нисходящем направлении метка MPLS, если анонсирующий узел PE использует входную репликацию для принимаемого от других PE трафика BUM. Если анонсирующий узел PE использует P2MP MPLS LSP для передачи трафика BUM, это **должна** быть назначенная в восходящем направлении метка MPLS. Использование метки описано в параграфе 8.3. Расщепление горизонта.

Если желателен режим Single-Active, бит Single-Active в поле флагов расширенной группы ESI Label **должен** иметь значение 1, а для метки ESI **следует** указывать действительное значение метки MPLS.

8.2.1.1. Ethernet A-D RT

Каждый маршрут Ethernet A-D для ES **должен** включать хотя бы один атрибут RT. Набор маршрутов Ethernet A-D для ES **должен** передавать весь набор RT для всех экземпляров EVPN, к которым относится сегмент Ethernet.

8.3. Расщепление горизонта

Рассмотрим узел CE, подключенный к 2 или более PE в сегменте Ethernet ES1, работающем в режиме All-Active. Если CE передаёт пакет BUM одному из PE, не являющемуся назначенным узлом пересылки (Designated Forwarder или DF), скажем, PE1, то PE1 будет пересылать этот пакет всем или части других PE в этом экземпляре EVPN, включая DF PE для этого сегмента Ethernet. В этом случае DF PE, с которым CE имеет несколько соединений, **должен** отбросить пакет и не пересылать его обратно CE. Эта фильтрация в документе называется расщеплением горизонта (split-horizon).

При работе множества PE в режиме Single-Active использование фильтрации split-horizon настоятельно рекомендуется, поскольку оно предотвращает временные петли в моменты отказа или восстановления, которые могут влиять на сегмент Ethernet (например, когда два PE считают себя DF для сегмента, пока не завершена процедура выбора DF).

Для реализации функции split-horizon каждый пакет fBUM, исходящий не от DF PE, инкапсулируется с меткой MPLS, указывающей сегмент происхождения (Ethernet), т. е. сегмент, из которого кадр попал в сеть EVPN. Эта метка называется меткой ESI и **должна** распространяться всеми PE, работающими в режиме All-Active, с использованием набора Ethernet A-D для маршрутов ES, как указано в параграфе 8.2.1. Метку ESI **следует** распространять всем PE при работе в режиме Single-Active, с использованием набора Ethernet A-D для маршрутов ES. Эти маршруты импортируются узлами PE, подключёнными к сегменту Ethernet, а также PE, имеющими хотя бы один общий экземпляр EVPN с сегментом Ethernet в маршруте. Как описано в параграфе 8.2.1[1], маршрут **должен** содержать расширенную группу ESI Label с действительной меткой ESI. PE полагается на значение метки ESI для решения вопроса о выходе кадра BUM из конкретного сегмента Ethernet.

8.3.1. Назначение метки ESI

В последующих параграфах описываются процедуры назначения меток ESI, которые зависят от типа туннелей, применяемых для доставки в сети EVPN пакетов с множеством адресатов.

8.3.1.1. Входная репликация

Каждый узел PE в режиме All-Active или Single-Active, применяющий репликацию на входе для трафика BUM, анонсирует в нисходящем направлении назначенную метку ESI в наборе маршрутов Ethernet A-D на ES для подключённых ES. Эта метка **должна** программироваться в пространстве меток платформы анонсирующим PE, а запись пересылки для этой метки должна указывать, что пакеты с такой меткой **не** пересылаются в сегмент Ethernet, куда была распространена эта метка.

Правила включения метки ESI в пакет BUM входным PE в режиме All-Active приведены ниже.

- Входной узел PE, на являющийся DF, **должен** включать метку ESI, распространяемую выходным PE DF, в копию передаваемого им пакета BUM.
- Входному узлу PE (DF или не DF) **следует** включать метку ESI, распространяемую каждым выходным PE, не являющимся DF, в копию передаваемого им пакета BUM.

Ниже приведено правило включения метки ESI в пакет BUM входным PE в режиме Single-Active.

[1] В оригинале ошибочно указан параграф 8.1.1. См. https://www.rfc-editor.org/errata/eid6286. *Прим. перев.*

- Входному узлу PE DF **следует** включать метку ESI, распространяемую выходным PE, в копию передаваемого им пакета BUM.

В обоих режимах All-Active и Single-Active входному PE **недопустимо** включать метку ESI в копию пакета BUM, передаваемого выходному узлу PE, не подключённому к сегменту ES, через который пакет BUM попал в EVI.

В качестве примера рассмотрим PE1 и PE2 с многодомными подключениями к CE1 в сегменте ES1, работающими в режиме All-Active. Предположим, что PE1 использует LSP P2P или MP2P для передачи пакетов PE2. Пусть PE1 не является DF для VLAN1, а PE2 служит DF для VLAN1 и PE1 получает пакет BUM от CE1 в VLAN1 сегмента ES1. В этом случае PE2 распространяет маршрут Inclusive Multicast Ethernet Tag для VLAN1, соответствующей экземпляру EVPN. Когда PE1 передаёт полученный от CE1 пакет BUM, он **должен** сначала втолкнуть в стек MPLS метку ESI, которую PE2 распространил для ES1. Затем он **должен** втолкнуть в стек MPLS метку MPLS, распространённую PE2 в маршруте Inclusive Multicast Ethernet Tag для VLAN1. Полученный пакет инкапсулируется в стек меток LSP P2P или MP2P для передачи пакета PE2. Когда PE2 получает этот пакет, он определяет по верхней метке MPLS набор ESI, в которые он будет реплицировать пакет после удаления меток LSP P2P или MP2P. Если следующей является метка ESI, назначенная PE2 для ES1, то PE2 **недопустимо** пересылать пакет в ES1. Если следующей является метка ESI, не назначенная PE2, то PE2 **должен** отбросить пакет. Следует отметить, что в этом случае при получении PE2 пакета BUM для VLAN1 от CE1 ему **следует** инкапсулировать пакет с меткой ESI, полученной от PE1, при передаче его PE1, чтобы избежать временных петель в случае отказа, который влияет на ES1 (например, отказ порта или канала).

8.3.1.2. P2MP MPLS LSP

PE в режиме All-Active, не являющиеся DF и применяющие P2MP LSP для передачи трафика BUM анонсируют назначенную восходящим направлением метку ESI в наборе маршрутов Ethernet A-D на ES для своего общего подключённого сегмента ES. Эта метка назначается восходящим PE, анонсирующим маршрут. Метка **должна** программироваться другими PE, которые подключены к ESI, анонсируемому в маршруте, в контексте пространства меток для анонсирующего PE. Кроме того, запись пересылки для этой метки должна приводить к тому, что пакеты с этой меткой **не** пересылаются в сегмент Ethernet, для которого распространена метка. Эта метка **должна** также программироваться другими PE, которые импортируют маршрут, но не подключены к ESI, анонсируемому в маршруте, в контексте пространства меток анонсирующего PE. Кроме того, запись пересылки для этой метки должна приводить к выталкиванию (pop) метки без других действий.

DF PE в режиме Single-Active, использующим P2MP LSP для передачи трафика BUM, следует анонсировать назначенную восходящим направлением метку ESI в наборе маршрутов Ethernet A-D на ES для своего подключённого сегмента ES, как описано выше.

В качестве примера рассмотрим PE1 и PE2 с многодомным подключением к CE1 в ES1, работающие в режиме All-Active. Рассмотрим также PE3 одного из экземпляров EVPN в ES1. Предположим, что узел PE1, не являющийся DF, применяет P2MP MPLS LSP для передачи пакетов BUM. Когда PE1 передаёт пакет BUM, полученный от CE1, он **должен** сначала втолкнуть (push) в стек MPLS метку ESI, назначенную для ESI, где был получен пакет. Результирующий пакет далее инкапсулируется в стек меток P2MP MPLS, требуемый для передачи пакета другим PE. Выталкивание на предпоследнем этапе **должно** быть отключено в P2MP LSP, используемых в транспортной инфраструктуре MPLS для EVPN. Когда PE2 получит этот пакет, он декапсулирует верхнюю метку MPLS и перешлёт пакет с использованием контекста пространства меток, определяемого верхней меткой. Если следующей является метка ESI, назначенная PE1 для ES1, тогда PE2 **недопустимо** пересылать пакет в ES1.Когда PE3 получит этот пакет, он декапсулирует верхнюю метку MPLS и перешлёт пакет, используя контекст пространства меток, определяемый верхней меткой. Если следующей является метка ESI, назначенная PE1 для ES1 и PE3 не подключён к ES1, тогда PE3 **должен** вытолкнуть метку и лавинно разослать пакет через все локальные ESI в этом экземпляре EVPN. Нужно отметить, что при передаче узлом PE2 кадра BUM через P2MP LSP, следует инкапсулировать кадр с меткой ESI, даже если этот узел служит DF для данной VLAN, чтобы избежать временных петель в случае отказа, который может влиять на ES1 (например, отказ порта или канала).

8.4. Псевдонимы и резервный путь

В случае, когда CE подключён к нескольким PE, с использованием группы агрегирования (Link Aggregation Group или LAG) с избыточностью All-Active, возможно, что лишь один PE узнает набор MAC-адресов, связанных с трафиком от CE. Это ведёт к ситуации, когда удалённые PE получают маршруты MAC/IP Advertisement для этих адресов от одного PE даже при подключении к многодомному сегменту нескольких PE. В результате удалённые PEs не смогут эффективно распределить нагрузку между узлами PE, подключёнными к многодомному сегменту Ethernet. Это может возникать, например, когда PE выполняют обучение в плоскости данных при доступе, а функция распределения нагрузки в CE хэширует трафик с заданного MAC-адреса отправителя в один узел PE. Другой случай возникает, когда PE полагаются на обучение в плоскости управления на доступе (например, через ARP), поскольку трафик ARP будет хэшироваться одним каналом в LAG.

Для решения этой проблемы в EVPN введена концепция псевдонимов (aliasing), обеспечивающая PE возможность сигнализировать свою доступность для экземпляра EVPN в данном ES даже без изучения MAC-адресов из EVI/ES. Для этого служит маршрут Ethernet A-D для EVI. Удалённому PE, получающему маршрут MAC/IP Advertisement с незарезервированным ESI, **следует** считать анонсированный MAC-адрес доступным через все PE, анонсировавшие доступность EVI/ES для этого MAC-адреса через комбинацию маршрута Ethernet A-D на EVI для EVI/ES (и тега Ethernet, если это применимо) и Ethernet A-D на ES для ES со сброшенным (0) битом Single-Active в флагах расширенной группы ESI Label.

Отметим, что маршрут Ethernet A-D для EVI может быть получен удаленным PE до получения набора маршрутов Ethernet A-D для ES. Поэтому маршрут Ethernet A-D для EVI **недопустимо** применять для пересылки трафика удаленным PE, пока он не получит маршруты Ethernet A-D для ES.

Резервный путь является тесно связанным решением, но применяется в режиме Single-Active. В этом случае PE также анонсирует свою доступность для данного EVI/ES, применяя ту же комбинацию маршрутов Ethernet A-D для EVI и Ethernet A-D для ES, что указана выше, но с установленным (1) битом Single-Active в флагах расширенной группы ESI Label. Удалённому PE, получающему маршрут MAC/IP Advertisement с незарезервированным ESI, **следует** считать

анонсированный MAC-адрес доступным через любой PE, анонсировавший эту комбинацию маршрутов, а также ему **следует** установить резервный путь для этого MAC-адреса.

8.4.1. Создание маршрута Ethernet A-D на экземпляр EVPN

В этом параграфе описаны процедуры создания маршрута Ethernet A-D на экземпляр EVPN (EVI), который применяется для псевдонима (см. выше). Поддержка этого маршрута **необязательна**.

Отличитель маршрута (RD) **должен** устанавливаться в соответствии с 7.9. Назначение RD для MAC-VRF.

Идентификатор сегмента Ethernet (ESI) **должен** быть 10-октетным значением, как описано в разделе 5. Сегмент Ethernet. Маршрут Ethernet A-D не требуется при Segment Identifier = 0.

Ethernet Tag ID - это идентификатор тега Ethernet в сегменте Ethernet. Его значением может быть 12-битовый идентификатор VLAN ID, занимающий 12 младших битов (старшие 20 битов сбрасываются в 0), или другой тег Ethernet, используемый EVPN. Для него **можно** установить принятый по умолчанию тег Ethernet для сегмента Ethernet или 0.

Отметим, что сказанное выше позволяет анонсировать маршрут Ethernet A-D с одним из указанных ниже уровней детализации:

- Один маршрут Ethernet A-D на пару <ESI, Ethernet Tag ID> в MAC-VRF. Это применимо при использовании узлом PE размещения на основе MPLS с трансляцией VID и может быть применимо при использовании PE размещения по MAC-адресам с трансляцией VID.
- Один маршрут Ethernet A-D для каждого <ESI> в MAC-VRF (где Ethernet Tag ID = 0). Это применимо при использовании узлом PE размещения по MAC или MPLS без трансляции VID.

Использование метки MPLS описано в разделе 14. Распределение нагрузки для индивидуальных пакетов.

Поле Next Hop атрибута MP_REACH_NLRI из маршрута **должно** содержать адрес IPv4 или IPv6 анонсирующего PE.

Маршрут Ethernet A-D **должен** содержать один или несколько атрибутов RT, как указано в 7.10. Цели маршрутов.

8.5. Выбор DF

Рассмотрим CE, который является хостом или маршрутизатором с прямыми подключениями к нескольким PE в экземпляре EVPN данного сегмента Ethernet. В сегменте Ethernet может быть задан один или несколько тегов Ethernet. В этом примере лишь один из узлов PE, назначенный пересылающим (Designated Forwarder или DF), отвечает за указанные ниже действия.

- Передача группового и широковещательного трафика с данным тегом Ethernet из конкретного сегмента Ethernet в CE.
- Лавинная рассылка неизвестного индивидуального трафика (трафика, для которого PE не знает MAC-адрес получателя) с данным тегом Ethernet из конкретного сегмента Ethernet узлу CE, если среда требует этого.

Отметим, что выбор DF с детализацией <ES, VLAN> или <ES, VLAN bundle> для трафика BUM принят по умолчанию в данной спецификации.

CE всегда передаёт пакеты конкретного потока по одному каналу к PE. Например, если CE является хостом, он, как отмечено выше, считает набор каналов, используемых для доступа к узлам PE группой агрегирования (LAG). CE применяет локальную функцию хэширования для отображения потоков трафика в каналы LAG.

Если сеть мостов подключена к множеству PE в сети EVPN через коммутаторы, поддержка избыточности All-Active требует соединения сети мостов с двумя или более PE с применением LAG.

Если сеть мостов не соединена с узлами PE через LAG, лишь один из каналов между сетью мостов и PE должен быть активным для данной пары <ES, VLAN> или <ES, VLAN bundle>. В этом случае набор маршрутов Ethernet A-D для ES, анонсируемый каждым PE, **должен** иметь установленный (1) бит Single-Active в флагах расширенной группы ESI Label.

Принятая по умолчанию процедура выбора DF с детализацией <ES, VLAN> для службы на основе VLAN или <ES, VLAN bundle> для службы VLAN-Aware Bundle называется «нарезкой сервиса» (service carving). При таком выделении можно выбрать несколько DF на сегмент Ethernet (по одному на VLAN или VLAN bundle), чтобы распределять нагрузку многоадресного трафика в данный сегмент. Процедуры распределения нагрузки равномерно делят пространство VLAN в ES между PE и каждый PE служит DF для отдельных (не пересекающихся) VLAN или VLAN bundle в данном ES. Процедура «нарезки» описана ниже.

1. Когда узел PE обнаруживает ESI подключённого сегмента Ethernet, он анонсирует маршрут Ethernet Segment со связанным атрибутом расширенной группы ES-Import.
2. Затем PE запускает таймер (по умолчанию на 3 секунды), чтобы разрешить получение маршрутов Ethernet Segment от других узлов PE, подключённых к тому же сегменту Ethernet. Значение таймера следует задавать одинаковым для всех PE, подключённых к одному сегменту Ethernet.
3. По завершении отсчёта таймера каждый PE создаёт список адресов IP всех узлов PE, подключённых к сегменту Ethernet (включая себя) в порядке роста числовых значений. Адреса IP для этого списка извлекаются из поля Originating Router's IP address в анонсированном маршруте Ethernet Segment. Затем каждому PE назначается номер, указывающий его позицию в списке (начиная с 0 для PE с наименьшим численным значением адреса IP. Номера служат для выбора PE на роль DF для данного экземпляра EVPN в сегменте Ethernet по приведённым ниже правилам.

   В предположении группы резервирования из N узлов PE для службы на основе VLAN узел PE с номером i будет DF для <ES, VLAN V>, когда (V mod N) = i. Для службы VLAN-Aware Bundle при делении по модулю **должно** применяться наименьшее значение VLAN в группе на данном ES.

Следует отметить, что поле Originating Router's IP address в маршруте Ethernet Segment для получения IP-адреса PE должно содержать упорядоченный список, позволяющий узлу CE быть многодомным в разных AS, если такая потребность когда-нибудь возникнет.

4. Узел PE, выбранный в качестве DF для данной пары <ES, VLAN> или <ES, VLAN bundle>, разблокирует многоадресный трафик для данной VLAN или VLAN bundle в соответствующем сегменте ES. Отметим, что DF PE разблокирует многоадресный трафик, выходящий в сегмент. Остальные (не DF) PE продолжат отбрасывать многоадресный трафик в выходном направлении к <ES, VLAN> или <ES, VLAN bundle>.

   При отказе порта или канала затронутый узел PE отзывает свой маршрут Ethernet Segment. Это заново запустит процедуру нарезки сервиса на всех PE в группе резервирования. При отказе узла PE, вводе или выводе PE из эксплуатации узлы PE снова запустят процедуру нарезки сервиса. В случае многодомного режима Single-Active при переносе службы с одного PE в группе резервирования на другой, который в конечном итоге будет выбран DF для службы, **следует** инициировать уведомление об очистке MAC-адреса в соответствующий сегмент Ethernet. Это можно сделать, например, с помощью обновления new в протоколе IEEE 802.1ak Multiple VLAN Registration Protocol (MVRP).

8.6. Взаимодействие с однодомными PE

Будем называть PE, поддерживающие лишь однодомные устройства CE, однодомными PE. Для таких PE упомянутые выше многодомные процедуры можно опустить, однако для полной совместимости однодомных PE с многодомными следует поддерживать на однодомных PE некоторые из описанных выше многодомных процедур, указанные ниже.

- Процедуры обработки маршрутов Ethernet A-D для быстрого схождения (8.2. Быстрое схождение), позволяющие однодомным PE воспользоваться преимуществами быстрого схождения.
- Процедуры обработки маршрутов Ethernet A-D для псевдонимов (8.4. Псевдонимы и резервный путь), позволяющие однодомным PE воспользоваться преимуществами распределения нагрузки.
- Процедуры обработки маршрутов Ethernet A-D для резервного пути (8.4. Псевдонимы и резервный путь), позволяющие однодомным PE воспользоваться преимуществами улучшенного схождения.

9. Определение доступности индивидуального MAC-адреса

PE пересылают полученные пакеты по MAC-адресам получателей. Это предполагает способность PE изучать пути достижения индивидуальных MAC-адресов получателей. Имеется два варианта изучения MAC-адресов - локальное и удалённое.

9.1. Локальное обучение

Конкретный узел PE должен быть способен изучать MAC-адреса от подключённых к нему CE. Это называется локальным обучением.

Узлы PE в конкретном экземпляре EVPN **должны** поддерживать обучение в локальной плоскости данных по стандартным процедурам IEEE Ethernet. Узел PE должен быть способен изучать MAC-адреса в плоскости данных при получении из сети CE:

- запросов DHCP;
- ARP Request для своего MAC;
- ARP Request для партнёра.

Дополнительно PE **могут** изучать MAC-адреса от CE в плоскости управления или через интеграцию плоскости поддержки (management) между PE и CE.

Имеются приложения, в которых MAC-адрес, доступный через данный PE в локально подключённом сегменте (например, с ESI X), может перемещаться, становясь доступным через иной PE в другом сегменте (например, с ESI Y). Это называется мобильностью MAC (MAC Mobility), процедуры для этого описаны в 15. Мобильность MAC.

9.2. Удалённое обучение

Конкретный узел PE должен быть способен определить, как передать трафик по MAC-адресу, принадлежащему CE, подключённым к другим PE, или находящемуся за ними, т. е. удаленным CE или находящимся за ними хостам. Такие MAC-адреса называются здесь удалёнными.

Этот документ требует от PE изучать удалённые MAC-адреса в плоскости управления. Для этого каждый узел PE анонсирует MAC-адреса, узнанные от локально подключённых CE в плоскости управления, всем другим PE этого экземпляра EVPN, используя MP-BGP и, в частности, маршрут MAC/IP Advertisement.

9.2.1. Создание анонса адреса MAC/IP

Протокол BGP расширен для анонсирования MAC-адресов с использованием типа маршрута MAC/IP Advertisement в EVPN NLRI.

Значение RD **должно** быть установлено в соответствии с 7.9. Назначение RD для MAC-VRF. Для Ethernet Segment Identifier устанавливается 10-октетное значение ESI, описанное в 5. Сегмент Ethernet. Ethernet Tag ID может иметь значение 0 или представлять действительный тег Ethernet Tag ID, когда имеется несколько таблиц мостов в MAC-VRF (т. е. PE нужно поддерживать сервис VLAN-Aware Bundle для этого EVI).

Когда Ethernet Tag ID в NLRI имеет ненулевое значение для конкретного домена широковещания, это может быть значение Ethernet для CE (например, VLAN ID у CE) или провайдера EVPN (например VLAN ID провайдера). Последний вариант применяется, когда теги Ethernet у CE (например, VLAN ID у CE) для конкретного домена широковещания различаются у разных CE.

Поле MAC Address Length указывается в битах и имеет значение 48. Иные значения размера MAC-адреса выходят за рамки этого документа. Кодирование MAC-адреса **должно** использовать 6-октетный формат, заданный в [802.1Q] и [802.1D-REV].

Поле IP Address не обязательно. По умолчанию в поле IP Address Length установлено значение 0 и IP Address не включается в маршрут. Если нужно анонсировать действительный адрес IP, он кодируется в маршруте. При наличии IP-адреса поле IP Address Length указывает его размер в битах (32 или 128). Другие значения IP Address Length выходят за рамки этого документа. IP-адрес **должен** быть задан 4 октетами для IPv4 или 16 октетами для IPv6. Поля Length в EVPN NLRI (размер в октетах, как указано в 7. Маршруты BGP EVPN) достаточно для определения наличия и формата адреса IP в маршруте как для IPv4, так и для IPv6.

Поле MPLS Label1 кодируется как 3 октета, где старшие 20 битов содержат значение метки. Метка MPLS Label1 **должна** назначаться в нисходящем направлении и связывается с MAC-адресом анонсируемым PE. Анонсирующий PE использует эту метку при получении инкапсулированного в MPLS пакета для пересылки на по MAC-адресу получателя в сторону CE. Процедуры пересылки заданы в разделах 13 и 14.

PE может анонсировать одну и ту же метку EVPN для всех MAC-адресов в данном экземпляре MAC-VRF. Назначение такой метки называется назначением метки для MAC-VRF. Как вариант, PE может анонсировать уникальную метку EVPN для пары <MAC-VRF, тег Ethernet>. Это называется назначением метки <MAC-VRF, тег Ethernet>. В качестве третьего варианта PE может анонсировать уникальную метку EVPN для пары <ESI, тег Ethernet>. Это называется назначением метки <ESI, тег Ethernet>. Четвёртым вариантом является анонсирование узлом PE уникальной метки EVPN для MAC-адреса. Это называют назначением метки для MAC. У всех этих вариантов есть свои недостатки и выбор конкретного метода определяется локально узлом PE, от которого исходит маршрут.

Назначение по метке MAC-VRF требует наименьшего числа меток EVPN, но нужен роиск MAC в дополнение к поиску MPLS для пересылки на выходном PE. С другой стороны, уникальная метка на <ESI, тег Ethernet> или MAC позволяет выходному PE пересылать пакеты, полученные от другого PE, подключённому CE после поиска лишь по метке MPLS (без поиска MAC). Это включает возможность выполнять подходящую трансляцию VLAN ID на выходе к CE.

Поле MPLS Label2 не обязательно и при наличии кодируется 3 октетами, 20 старших битов которых содержат метку. В поле Next Hop атрибута MP_REACH_NLRI для маршрута **должен** быть помещён адрес анонсирующего PE (IPv4 или IPv6).

Анонс BGP для маршрута MAC/IP Advertisement **должен** содержать хотя бы один атрибут RT. Значения RT могут настраиваться (как в IP VPN) или выводиться автоматически из Ethernet Tag ID в случае Unique VLAN, как описано в параграфе 7.10.1. Автоматический вывод из Tag ID.

Следует отметить, что этот документ не требует от узлов PE создания состояний пересылки для удалённых MAC, когда те получены из плоскости управления. Вопрос создания таких состояний решается локально.

9.2.2. Распознавание маршрута

Если поле Ethernet Segment Identifier в полученном маршруте MAC/IP Advertisement имеет зарезервированное значение ESI (0 или MAX-ESI), то при решении PE установить состояние пересылки для соответствующего MAC-адреса оно **должно** основываться только на маршруте MAC/IP Advertisement.

Если поле Ethernet Segment Identifier в полученном маршруте MAC/IP Advertisement имеет незарезервированное значение ESI и принимающий узел PE локально подключён к тому же ESI, тогда PE не меняет своё состояние пересылки на основе полученного маршрута. Это гарантирует предпочтение локальных маршрутов перед удалёнными.

Если поле Ethernet Segment Identifier в полученном маршруте MAC/IP Advertisement имеет незарезервированное значение ESI и принимающий узел PE решит установить состояние пересылки для соответствующего MAC-адреса, это **должно** происходить при получении как маршрута MAC/IP Advertisement, так и связанного набора маршрутов Ethernet A-D для ES. Зависимость установки маршрута MAC от маршрутов Ethernet A-D для ES нужна для гарантированного предотвращения случайной установки маршрутов MAC во время массового отзыва.

Для иллюстрации этого рассмотрим два PE (PE1 и PE2) соединённые с многодомным сегментом Ethernet (ES1). Предполагается режим избыточности All-Active. Данный MAC-адрес M1 известен PE1, но не PE2. На PE3 могут возникать показанные ниже состояния.

T1

Когда получен маршрут MAC/IP Advertisement от PE1 и наборы маршрутов Ethernet A-D для ES и Ethernet A-D для EVI от PE1 и PE2, узел PE3 может пересылать трафик для M1 как PE1, так и PE2.

T2

Если после T1 узел PE1 отзовёт свой набор маршрутов withdraws Ethernet A-D для ES, узел PE3 будет пересылать трафик для M1 только PE2.

T2'

Если после T1 узел PE2 отзовёт свой набор маршрутов withdraws Ethernet A-D для ES, узел PE3 будет пересылать трафик для M1 только PE1.

T2"

Если после T1 узел PE1 отзовёт свой маршрут MAC/IP Advertisement, узел PE3 будет считать трафик для M1 направленным по неизвестному индивидуальному адресу.

T3

PE2 анонсирует маршрут MAC для M1, а затем PE1 отзывает свой маршрут MAC для M1. Узел PE3 продолжает пересылать трафик для M1 обоим узлам PE1 и PE2. Иными словами, несмотря на отзыв M1 узлом PE1, PE3 пересылает трафик для M1 обоим узлам PE1 PE2. Это связано с тем, что поток от CE, ведущий к хешированию трафика для M1 узлом PE1, может быть прерван, что приведёт к устареванию M1 на PE1, однако M1 может быть доступен обоим узлам PE1 и PE2.

10. ARP и ND

Поле IP Address в маршруте MAC/IP Advertisement может содержать один из адресов IP, связанных с MAC-адресом. Это можно использовать для минимизации лавинной рассылки сообщений ARP или ND (Neighbor Discovery) через сеть

MPLS и удалённые CE. Минимизируется также обработка сообщение ARP (или ND) на конечных станциях (хостах), подключённых к сети EVPN. Узел PE может узнать адрес IP, связанный с MAC-адресом, из плоскости управления или данных между CE и PE или путём отслеживания некоторых сообщений от CE или к нему. Когда PE знает адрес IP, связанный с MAC-адресом локально подключённого CE, он может анонсировать этот адрес другим PE путём включения в маршрут MAC/IP Advertisement. Это может быть адрес IPv4, представленный 4 октетами, или IPv6 из 16 октетов. Для ARP и ND в поле IP Address Length **должно** устанавливаться значение 32 при адресе IPv4 и 128 для IPv6.

Если с MAC-адресом связано несколько адресов IP, маршрут MAC/IP Advertisement **должен** создаваться для каждого адреса IP. Например, это могут быть адреса IPv4 и IPv6 для одного MAC-адреса при использовании двойного стека IP. Когда привязка адреса IP к MAC-адреса удаляется, маршрут MAC/IP Advertisement для этого IP **должен** отзываться.

Отметим, что маршрут только для MAC может анонсироваться вместе с маршрутом MAC/IP, но независимо от него, в случаях, когда изучение MAC через сеть или узел выполняется в плоскости данных и не зависит от отслеживания ARP для создания маршрута MAC/IP. В таких случаях при завершении срока действия записи ARP и отзыве MAC/IP информация MAC не будет теряться. Когда MAC/IP изучается в плоскости поддержки (management) или управления, передающий узел PE может создавать и анонсировать лишь маршрут MAC/IP. Если принимающий PE получает оба маршрута MAC и MAC/IP, то при получении отзыва маршрута MAC/IP он **должен** удалить соответствующую запись из таблицы ARP, но не запись для MAC из таблицы MAC-VRF, если не получен отзыв и для маршрута MAC.

Когда PE получает ARP Request для адреса IP от CE и у этого PE есть привязка MAC-адреса к IP, узлу PE **следует** выступить посредником ARP, отвечая на ARP Request.

10.1. Принятый по умолчанию шлюз

Когда PE нужно выполнить пересылку между подсетями, каждая из которых представляет свой домен широковещания (например, другую сеть VLAN), пересылка происходит на уровне L3 и выполняющий такую функцию узел PE называется принятым по умолчанию шлюзом (default gateway) для экземпляра EVPN. В этом случае при получении PE запроса ARP для IP-адреса, настроенного как адрес принятого по умолчанию шлюза, узел PE возвращает ARP Reply.

Каждый узел PE, служащий принятым по умолчанию шлюзом для данного экземпляра EVPN, **может** анонсировать в плоскости управления EVPN свой MAC-адрес этого шлюза, используя маршрут MAC/IP Advertisement, и каждый такой PE указывает, что маршрут связан с заданным по умолчанию шлюзом. Для этого нужно включать в маршрут расширенную группу Default Gateway, определённую в параграфе 7.8. Расширенная группа Default Gateway. В анонсах маршрутов MAC с расширенной группой Default Gateway устанавливается ESI = 0.

В поле IP Address маршрута MAC/IP Advertisement устанавливается IP-адрес принятого по умолчанию шлюза этой подсети (например, экземпляра EVPN). Для данной подсети (например, VLAN или экземпляр EVPN) IP-адрес принятого по умолчанию шлюза совпадает на всех участвующих PE. Включение этого адреса IP позволяет принимающему PE сравнить заданный у него IP-адрес шлюза по умолчанию с полученным в маршруте MAC/IP Advertisement для этой подсети (или экземпляра EVPN), при наличии расхождений PE **следует** уведомить оператора и записать ошибку в системный журнал (log).

Если заранее не известно (из не описанных в документе источников), что все PE в данном экземпляре EVPN выступают принятыми по умолчанию шлюзами для этого экземпляра EVPN, **должна** быть установлена действительная метка MPLS для нисходящего направления (downstream).

Кроме того, даже если все PE данного экземпляра EVPN выступают принятыми по умолчанию шлюзами для этого экземпляра EVPN, но не у всех PE имеется достаточно (маршрутной) информации для маршрутизации между подсетями для всего трафика между подсетями из подсети, связанной с экземпляром EVPN, тогда при анонсировании таким PE в плоскости управления EVPN MAC-адреса своего шлюза по умолчанию в маршруте MAC/IP Advertisement с указанием связанного с маршрутом шлюза по умолчанию, этот маршрут **должен** включать действительную метку для нисходящего направления.

Если все PE данного экземпляра EVPN выступают принятыми по умолчанию шлюзами для этого экземпляра EVPN и используется один и тот же MAC-адрес шлюза по умолчанию на всех, такие анонсы не требуются. Однако при наличии у каждого шлюза своего MAC-адерса, все шлюза должны знать MAC-адреса других шлюзов и анонсы нужны. Это называется псевдонимами MAC-адреса (MAC address aliasing), поскольку один принятый по умолчанию шлюз может быть представлен несколькими адресами MAC.

Каждый PE, получающий такой маршрут и импортирующий его по процедурам этого документа, отвечает на полученные сообщения ARP Requests, как описано в этом параграфе.

Каждый PE в роли заданного по умолчанию шлюза для данного экземпляра EVPN при получении и импорте такого маршрута по процедурам этого документа **должен** создать состояние пересылки MAC, позволяющее ему выполнять пересылку IP для пакетов, направленных по MAC-адресу из маршрута.

11. Обработка трафика с множеством получателей

Данному PE нужны процедуры для передачи широковещательного и группового трафика, полученного от CE с данным тегом Ethernet (VLAN) в данном экземпляре EVPN, всем другим PE, охватывающим этот тег Ethernet (VLAN) в данном экземпляре EVPN. В некоторых ситуациях, как описано в разделе 12. Обработка индивидуальных пакетов неизвестных получателей, PE может также потребоваться лавинная рассылка индивидуального трафика для неизвестные адресатов другим PE.

Узлы PE в конкретном экземпляре EVPN могут применять репликацию на входе, P2MP LSP или MP2MP LSP для передачи трафика BUM другим PE.

Каждый узел PE **должен** для указанного выше анонсировать маршрут Inclusive Multicast Ethernet Tag. В последующих параграфах рассмотрены процедуры создания маршрута Inclusive Multicast Ethernet Tag.

11.1. Создание маршрута Inclusive Multicast Ethernet Tag

Отличитель маршрута RD **должен** быть установлен в соответствии с параграфом 7.9. Назначение RD для MAC-VRF.

Ethernet Tag ID - это идентификатор тега Ethernet, который может иметь значение 0 или действительного тега Ethernet.

В поле Originating Router's IP Address **должен** быть указан IP-адрес PE, которому следует быть общим для всех EVI на PE (например, этом может быть адрес loopback-интерфейса PE). Поле IP Address Length указывает размер в битах.

Поле Next Hop атрибута MP_REACH_NLRI в маршруте **должно** содержать адрес анонсирующего PE (IPv4 или IPv6).

Анонс BGP для маршрута Inclusive Multicast Ethernet Tag **должен** включать хотя бы 1 атрибут Route Target (RT). Назначение RT **должно** выполняться по процедуре из параграфа 7.10. Цели маршрутов.

11.2. Идентификация P-Tunnel

Для идентификации туннеля P-tunnel, служащего для пересылки трафика BUM, маршрут Inclusive Multicast Ethernet Tag **должен** включать атрибут Provider Multicast Service Interface (PMSI), как укащано в [RFC6514]. В зависимости от технологии, применяемой в P-tunnel для экземпляра EVPN на PE, атрибут PMSI Tunnel маршрута Inclusive Multicast Ethernet Tag создаётся, как указано ниже.

- Если создавший анонс узел PE использует дерево P-multicast для P-tunnel в EVPN, атрибут PMSI Tunnel **должен** содержать отождествление дерево (Отметим, что PE может создать отождествление до фактического создания экземпляра дерева).
- Узел PE, использующий дерево P-multicast для P-tunnel, **может** агрегировать несколько экземпляров EVPN (EVI), присутствующих на PE, в одно дерево. В этом случае дополнительно к отождествлению дерева атрибут PMSI Tunnel **должен** включать назначенную в восходящем направлении метку MPLS, которую PE однозначно привязал к EVI, связанному с обновлением (как указано его атрибутами RT).

  Если PE уже анонсировал маршруты Inclusive Multicast Ethernet Tag для двух или более EVI, которые теперь хочет объединить, он **должен** анонсировать эти маршруты заново. Анонсируемые повторно маршруты **должны** совпадать с исходными, за исключением атрибута PMSI Tunnel и содержащейся в нем метки.
- Если создавший анонс узел PE использует использует репликацию на входе для P-tunnel в EVPN, маршрут **должен** включать атрибут PMSI Tunnel с Tunnel Type = Ingress Replication и Tunnel Identifier с маршрутизируемым адресом PE. Атрибут PMSI Tunnel **должен** содержать назначенную в нисходящем направлении метку MPLS. Эта метка служит для демультиплексирования трафика EVPN BUM, полученного PE через туннель MP2P.
- Флаг Leaf Information Required в атрибуте PMSI Tunnel **должен** быть сброшен (0), а при получении **должен** игнорироваться.

12. Обработка индивидуальных пакетов неизвестных получателей

Процедуры этого документа не требуют от узлов PE лавинной рассылки индивидуального трафика с неизвестными адресатами другим PE. Если PE узнают MAC-адреса CE из протокола плоскости управления, они могут распространять MAC-адреса через BGP и все индивидуальные адреса MAC будут известны до того, как появится трафик для них.

Если узел PE не знает MAC-адрес получателя, он может разослать пакет лавинно. При этом необходимо учитывать «расщепление горизонта», как описано ниже. Принципы, лежащие в основе описываемых процедур, заимствованы из правил пересылки с расщеплением горизонта в решениях VPLS [RFC4761] [RFC4762]. Когда PE, способный к лавинной рассылке (скажем, PEx) получает кадр с неизвестным MAC-адресом получателя, он применяет лавинную рассылку. Если кадр получен от присоединённого CE, узел PEx должен передать копию кадра в каждый сегмент Ethernet (с тем же EVI), для которого он служит DF, за исключением сегмента Ethernet, откуда получен кадр. Кроме того, узел PE должен разослать кадр всем другим PE, участвующим в экземпляре EVPN. Если же кадр получен от другого PE (скажем, PEy), PEx должен передать копию пакета в каждый сегмент Ethernet (с тем же EVI), для которого он служит DF. PEx **недопустимо** передавать кадр другим PE, поскольку узел PEy уже сделал это. Правила пересылки с расщеплением горизонта применяются к пакетам с неизвестными MAC-адресами.

Решение о применении лавинной рассылки пакетов для неизвестных MAC-адресов зависит от способа изучения адресов между узлами CE и узлами PE.

Узлы PE в конкретном экземпляре EVPN могут применять репликацию на входе, используя RSVP-TE P2P LSP или LDP MP2P LSP для передачи индивидуального трафика неизвестных получателей другим PE. Для этого можно также использовать RSVP-TE P2MP или LDP P2MP.

12.1. Входная репликация

При использовании репликации на входе атрибут P-tunnel в маршрутах Inclusive Multicast Ethernet Tag для экземпляра EVPN указывает нисходящую метку, которую другие PE могут применять при передаче трафика BUM для данного экземпляра EVPN этому PE.

Узел PE, получивший пакет с такой меткой MPLS, **должен** считать его пакетом BUM. Если MAC-адрес получателя является индивидуальным, PE **должен** считать пакет индивидуальным с неизвестным адресатом.

12.2. P2MP MPLS LSP

Процедуры использования P2MP LSP очень похожи на процедуры VPLS, описанные в [RFC7117]. Атрибут P-tunnel, используемый PE для передачи трафика BUM для конкретного экземпляра EVPN, анонсируется в маршруте Inclusive Multicast Ethernet Tag, как описано в разделе 11. Обработка трафика с множеством получателей.

Атрибут P-tunnel задаёт идентификатор P2MP LSP. Это эквивалент дерева Inclusive, описанного в [RFC7117]. Отметим, что множество тегов Ethernet, которые могут различаться в разных экземплярах EVPN, может применяться в одном P2MP LSP с использованием меток восходящего направления [RFC7117]. Это эквивалент дерева Aggregate Inclusive [RFC7117]. При использовании P2MP LSP для лавинной рассылки трафика неизвестных индивидуальных получателей порядок пакетов может меняться.

Узел PE, получающий пакет по пути P2MP LSP, заданному в атрибуте PMSI Tunnel, **должен** считать его пакетом BUM. Если MAC-адрес получателя является индивидуальным, PE **должен** считать пакет индивидуальным с неизвестным адресатом.

13. Пересылка индивидуальных пакетов

В этом разделе описаны процедуры для пересылки индивидуальных пакетов узлами PE, когда такие пакеты получены от напрямую подключённых CE или других PE.

13.1. Пересылка пакетов, полученных от CE

Когда узел PE получает пакет от CE с данным тегом Ethernet Tag ID, он должен сначала найти в пакете MAC-адрес отправителя. В некоторых средах со включённой защитой MAC адрес отправителя **может** служить для проверки отождествления хоста и и принятия решения о допуске трафика от этого хоста в сеть. Поиск MAC источника **может** также применяться для локального изучения MAC-адресов.

Если PE решает переслать пакет, он должен найти MAC-адрес получателя. Если PE получил анонс MAC для этого адреса получателя от одного или нескольких других PE или узнал его от присоединённых локально CE, этот MAC-адрес считается известным. Остальные MAC считаются неизвестными.

Для известного MAC-адреса узел PE пересылает пакет одному или нескольким удаленным PE или локально подключённому CE. При пересылке удалённому PE пакет инкапсулируется с меткой EVPN MPLS, анонсированной удаленным PE для этого MAC-адреса, в стек меток MPLS LSP для достижения удалённого PE.

Если MAC-адрес неизвестен и административная политика PE требует лавинной рассылки для неизвестных индивидуальных получателей, выполняются указанные ниже процедуры.

- Узел PE **должен** лавинно разослать пакет другим PE. Сначала PE **должен** инкапсулировать пакет с меткой ESI MPLS, как указано в параграфе 8.3. Расщепление горизонта. При использовании репликации на входе пакет **должен** реплицироваться каждому удалённому PE с меткой VPN, являющейся меткой MPLS. Это метка MPLS, анонсируемая удаленным PE в атрибуте PMSI Tunnel маршрута Inclusive Multicast Ethernet Tag для <MAC-VRF> или комбинации <MAC-VRF, тег Ethernet>.

  Тег Ethernet в маршруте может совпадать с тегом Ethernet, связанным с интерфейсом, через который входной узел PE получил пакет. При использовании P2MP LSP пакет **должен** передаваться через P2MP LSP, в котором PE является корнем для тега Ethernet в экземпляре EVPN. При использовании одного P2MP LSP для всех тегов Ethernet все PE в экземпляре EVPN **должны** быть листьями P2MP LSP. Если применяются разные P2MP LSP для данного тега Ethernet в экземпляре EVPN, листьями P2MP LSP должны быть лишь PE с тем же тегом Ethernet. Пакеты **должны** инкапсулироваться в стек меток P2MP LSP.

Если MAC-адрес неизвестен и административная политика PE запрещает лавинную рассылку для неизвестных индивидуальных получателей, узел PE **должен** отбросить пакет.

13.2. Пересылка пакетов, полученных от удалённого PE

В этом параграфе описаны процедуры пересылки индивидуальных пакетов от удалённого PE известным и неизвестным адресатам.

13.2.1. Пересылка неизвестным индивидуальным адресатам

При получении узлом PE пакета MPLS от удалённого PE сначала обрабатывается стек меток MPLS, затем, если верхняя метка в стеке MPLS оказывается меткой P2MP LSP, связанного с экземпляром EVPN или (при репликации на входе) нисходящей меткой, анонсированной в атрибуте P-tunnel, и после выполнения процедуры расщепления горизонта применяются указанные ниже действия. 8.3:

- Если PE служит DF для трафика BUM в конкретном наборе ESI для тега Ethernet, по умолчанию PE лавинно рассылает пакет этим ESI. Иными словами, по умолчанию PE предполагает, что для трафика BUM не требуется поиск MAC-адреса получателя. Как вариант, PE может выполнить такой поиск для лавинной рассылки пакета только части интерфейсов CE с таким тегом Ethernet. Например, PE может отказаться от пересылки пакетов BUM в некоторые сегменты Ethernet на основании административной политики, даже будучи DF для сегмента Ethernet.
- Если PE не является DF ни в одном ESI для тега Ethernet, по умолчанию пакет отбрасывается.

13.2.2. Пересылка известным индивидуальным адресатам

Если верхняя метка MPLS оказывается меткой EVPN, заданной в анонсе индивидуального MAC, узел PE пересылает пакет на основе сведений CE next-hop, связанных с меткой, или выполняет поиск MAC-адреса получателя для пересылки пакета узлу CE.

14. Распределение нагрузки для индивидуальных пакетов

В этом параграфе рассмотрены процедуры распределения нагрузки при пересылке индивидуальных пакетов для известных адресатов многодомному CE.

14.1. Распределение трафика от PE к удаленным CE

При импорте удаленным PE маршрута MAC/IP Advertisement для данной пары <ESI, тег Ethernet> в MAC-VRF, узле должен проверить все импортируемые маршруты Ethernet A-D для данного ESI с целью определения параметром распределения нагрузки в сегменте Ethernet.

14.1.1. Режим избыточности Single-Active

Для данного ES при импорте удаленным PE набора маршрутов Ethernet A-D для ES хотя бы от одного PE с установленным флагом Single-Active в расширенной группе ESI Label удалённый узел PE **должен** сделать вывод, что ES работает в режиме избыточности Single-Active. Поэтому MAC-адрес будет доступен лишь через PE, анонсировавший соответствующий маршрут MAC/IP Advertisement, его называют основным (primary) PE. Другие PE, анонсирующие наборы маршрутов Ethernet A-D для ES в том же ES, обеспечивают резервные пути для этого ES на случая отказа основного PE и называются резервными (backup) PE. Следует отметить, что основным PE для данной пары <ES, VLAN> (или <ES, VLAN bundle>) является DF для этой пары <ES, VLAN> (или <ES, VLAN bundle>).

При отказе основного PE он **может** отозвать свой набор маршрутов Ethernet A-D на ES для затронутого сегмента ES до отзыва своего набора маршрутов MAC/IP Advertisement.

Если имеется лишь один резервный PE для данного ES, удалённый PE **может** использовать отзыв набора маршрутов Ethernet A-D для ES от основного PE с целью обновления записей пересылки для связанных адресов MAC, чтобы указать резервный PE. Когда резервный PE начинает изучать MAC-адреса своих подключённых ES, он начнёт передачу своих маршрутов MAC/IP Advertisement, пока отказавший узел PE отзывает свои. Эти минимизирует лавинные рассылки трафика в случаях отказа.

Если имеется несколько резервных PE для данного ES, удалённый PE **должен** использовать отзыв набора маршрутов Ethernet A-D для ES от основного PE с целью начать лавинную рассылку трафика для связанных MAC-адресов (коль скоро лавинная рассылка для неизвестных индивидуальных адресатов разрешена административно), поскольку нет возможности выбрать один резервный узел PE.

14.1.2. Режим избыточности All-Active

Если удалённый узел PE импортировал набор маршрутов Ethernet A-D для ES от одного или нескольких PE и ни один из маршрутов не имеет установленного (1) флага Single-Active в расширенной группе ESI Label, удалённый PE **должен** сделать вывод, что данные сегмент ES работает в режиме избыточности All-Active. Удалённому узлу PE, получившему маршрут MAC/IP Advertisement с нерезервным ESI, **следует** считать все анонсированные MAC-адреса доступными через все PE, которые анонсировали для этого MAC доступность EVI/ES через комбинацию маршрута Ethernet A-D на EVI для данного EVI/ES (и тега Ethernet, если это применимо) и маршрута Ethernet A-D на ES для данного сегмента ES. Удалённый PE **должен** использовать полученные маршруты MAC/IP Advertisement и Ethernet A-D на EVI/ES для создания набор следующих узлов (next hop) на пути к анонсированным MAC-адресам.

Каждый следующий узел (next hop) включает стек меток MPLS, который применяется выходным PE для пересылки пакета. Этот стек описан ниже.

- Если next hop создаётся из маршрута MAC, **должен** применяться этот стек меток. Однако, если маршрута MAC для данного PE нет, next hop и стек меток MPLS создаются из маршрутов Ethernet A-D. Отметим, что последующее описание относится к определению стека меток для конкретного next hop на пути к данному PE, откуда удалённый PE получил и импортировал маршруты Ethernet A-D с теми же ESI и тегом Ethernet, какие указаны в анонсе MAC. Упомянутые ниже маршруты Ethernet A-D относятся к импортированным от данного PE.
- Если набор маршрутов Ethernet A-D на ES для этого сегмента ES и маршрут Ethernet A-D для EVI существуют, лишь тогда используется метка из последнего.

Для объяснения рассмотрим узел CE (CE1) с подключением к двум PE (PE1 и PE2) на интерфейсе LAG (ES1), передающий пакеты с MAC-адресом источника MAC1 в VLAN1 (отображена на EVI1). Удалённый узел PE (PE3) способен узнать о доступности MAC1 от PE1 и PE2. Оба PE1 и PE2 могут анонсировать MAC1 в BGP, если они получали пакеты с MAC1 от CE1. Если это не так и MAC1 анонсирует лишь узел PE1, PE3 все равно будет считать MAC1 доступным через PE1 и PE2, поскольку оба анонсировали набор маршрутов Ethernet A-D на ES для сегмента ES1, а также маршрут Ethernet A-D на EVI для <EVI1, ES1>.

Стек меток MPLS для передачи пакетов узлу PE1 является стеком MPLS LSP для доступа к PE1 (на вершине стека), за которым следует метка EVPN анонсируемая PE1 для MAC от CE1. Стеком меток MPLS для передачи пакетов узлу PE2 является стек MPLS LSP для доступа к PE2 (на вершине стека), за которым следует метка MPLS из маршрута Ethernet A-D, анонсированного PE2 для <ES1, VLAN1>, если PE2 не анонсировал MAC1 через BGP. Назовём такой стек MPLS next hop.

Удалённый PE (PE3) может сейчас распределять трафик от своих CE, адресованный в CE1, между PE1 и PE2. PE3 может использовать кортеж сведений о потоке для хеширования трафика в один из MPLS next hop с целью распределения нагрузки для трафика IP. Кроме того, PE3 может распределять нагрузку по MAC-адресу отправителя.

Когда PE3 решит передать конкретный пакет узлу PE1 или PE2, он может выбрать один из возможных путей к конкретному удалённому PE, используя обычные процедуры MPLS. Например, если применяются туннели на основе RSVP-TE LSP и PE3 решает передать пакет PE1, тогда PE3 может выбрать из набора RSVP-TE LSP путь, где PE1 служит получателем.

Когда PE1 или PE2 получает пакет для CE1 от PE3 и его индивидуальный адресат известен, пакет пересылается CE1. Если это пакет BUM, пересылать его CE должен лишь один из узлов PE1 и PE2 - тот, который служит DF.

14.2. Распределение трафика между PE и локальным CE

На CE можно настроить более одного интерфейса для подключения к разным или одному PE с целью распределения нагрузки за счёт применения таких технологий, как LAG. PE и CE могут распределять трафик между интерфейсами с использованием одного из описанных ниже механизмов.

14.2.1. Обучение в плоскости данных

PE используют обучение в плоскости данных для локальных MAC-адресов, узнанных от локальных CE. Это позволяет узлам PE узнать конкретный MAC-адрес и связать его с один или несколькими интерфейсами, если технология между PE и CE разрешает несколько путей. PE могут распределять трафик для этого MAC-адреса между интерфейсами.

Способность CE распределять нагрузку, создаваемую на разных интерфейсах, зависит от реализации CE.

14.2.2. Обучение в плоскости управления

CE может быть хостом, анонсирующим с использованием протокола управления один MAC на всех интерфейсах. Это позволяет узлам PE узнать MAC-адрес хоста и связать его со всеми интерфейсами. Тогда PE могут распределять трафик для хоста между интерфейсами. Хост также может распределять трафик и получающий его узел PE использует для его пересылки процедуры EVPN.

15. Мобильность MAC

Данный хост или конечная станция (определяемые MAC-адресом) может перемещаться из одного сегмента Ethernet в другой. Это называется мобильностью или перемещением MAC (MAC Mobility или MAC move) и отличается от многодомной ситуации, где данный адрес MAC доступен через несколько PE для одного сегмента Ethernet. При перемещении MAC имеется два набора маршрутов MAC/IP Advertisement - по одному для каждого сегмента Ethernet - и MAC-адрес будет казаться доступным в каждом из этих сегментов.

Чтобы все PE экземпляра EVPN могли корректно определить текущее местоположение MAC-адреса, все анонсы о его доступности в прежнем сегменте Ethernet **должны** быть отозваны узлами PE, которые их анонсировали.

При использовании локального обучения в плоскости данных эти PE не могут определить перенос MAC-адреса в другой сегмент и получение маршрутов MAC/IP Advertisement с атрибутом расширенной группы MAC Mobility от других PE инициирует отзыв прежних анонсов. Если локальное обучение происходит в плоскости управления, оно само вызовет отзыв прежних анонсов узлами PE.

Когда данный MAC-адрес перемещался неоднократно, возможно, между одной парой сегментов Ethernet, может возникать множество отзывов и повторных анонсов. Чтобы все PE в экземпляре EVPN корректно получили сведения через инфраструктуру BGP, требуется добавить порядковый номер для атрибутов расширенной группы MAC Mobility. Для корректной обработки перемещений реализация **должна** отслеживать изменения порядковых номеров. Каждый случай перемещения данного MAC-адреса имеет порядковый номер, задаваемый в соответствии с приведёнными ниже правилами.

- PE при первом анонсировании MAC-адреса не использует атрибут расширенной группы MAC Mobility.
- PE, обнаруживший локально присоединённый MAC-адрес, который ранее был указан в маршруте MAC/IP Advertisement с другим идентификатором сегмента Ethernet, анонсирует MAC-адрес в маршруте MAC/IP Advertisement с атрибутом расширенной группы MAC Mobility и порядковым номером на 1 больше номера в атрибуте расширенной группы MAC Mobility из полученного маршрута MAC/IP Advertisement. При первом переносе данного MAC-адреса, когда в принятом маршруте MAC/IP Advertisement нет атрибута расширенной группы MAC Mobility предполагается номер 0.
- PE, обнаруживший локально присоединённый MAC-адрес, для которого уже был получен маршрут MAC/IP Advertisement с тем же ненулевым идентификатором сегмента Ethernet, анонсирует его:
  1. без атрибута расширенной группы MAC Mobility, если в принятом маршруте не было такого атрибута;
  2. с атрибутом расширенной группы MAC Mobility и порядковым номером, равным большему номеров в полученных маршрутах MAC/IP Advertisement с атрибутом расширенной группы MAC Mobility.
- PE, обнаруживший локально присоединённый MAC-адрес, для которого уже был получен маршрут MAC/IP Advertisement с тем же нулевым идентификатором сегмента Ethernet (однодомный вариант), анонсирует его с атрибутом расширенной группы MAC Mobility и порядковым номером, установленным должным образом. В однодомном варианте ESI не сравниваются, а в многодомном сравнение ESI служит для предотвращения ложного обнаружения переноса MAC между PE, подключёнными к одному многодомному сайту.

PE, получивший маршрут MAC/IP Advertisement для MAC-адреса с другим идентификатором сегмента Ethernet и большим порядковым номером, чем он ранее анонсировал, отзывает свой маршрут MAC/IP Advertisement. Если два или более PE анонсируют один MAC-адрес с одинаковым порядковым номером, но разными идентификаторами сегментов Ethernet, получивший маршруты PE выбирает маршрут, анонсированный PE с меньшим адресом IP. Если PE является инициатором маршрута MAC и получает тот же MAC-адрес с таким же номером, как создал он сам, PE сравнивает свой адрес IP с IP-адресом удалённого PE и выбирает меньший IP. Если его маршрут не является лучшим, PE отзывает его.

15.1. Проблема дублирования MAC

Возможны случаи, когда один MAC-адрес узнают разные PE в одной VLAN поскольку два (или более) хоста ошибочно настроены с одним адресом MAC (дубликат). В таких ситуациях трафик от этих хостов будет вызывать постоянные «перемещения» MAC между PE, подключёнными к этим хостам. Важно распознавать такие ситуации и избегать увеличения порядкового номера (в атрибуте расширенной группы MAC Mobility) до бесконечности. Для исправления ситуации узел PE, обнаруживший перемещение MAC путём локального обучения, запускает таймер на M секунд (по умолчанию of M = 180) и при обнаружении N (по умолчанию 5) перемещений MAC до завершения отсчёта таймера принимает решение о наличии дубликата MAC. Узел PE **должен** уведомить оператора и прекратить передачу и обработку маршрутов BGP MAC/IP Advertisement для этого MAC-адреса, пока оператор не устранить проблему. Значения M и N **должны** быть настраиваемыми для обеспечения гибкости оперативного управления. Отметим, что другие PE экземпляра EVPN будут пересылать трафик для дублированного MAC-адреса одному из анонсирующих его узлов PE.

15.2. Закреплённые MAC-адреса

В некоторых случаях желательно задать некоторые MAC адреса как статические, чтобы они не могли перемещаться. Такие MAC-адреса анонсируются с расширенной группой MAC Mobility, где установлен (1) флаг static и порядковый номер имеет значение 0. если PE получил такой анонс, а потом узнал этот же MAC-адрес от локального обучения, он **должен** уведомить оператора.

16. Групповая передача и широковещание

Узлы PE конкретного экземпляра EVPN могут применять репликацию на входе или P2MP LSP для пересылки группового трафика другим PE.

16.1. Входная репликация

PE могут применять репликацию на входе для лавинной рассылки трафика BUM, как описано в разделе 11. Обработка трафика с множеством получателей. Данный широковещательный пакет должен передаваться всем удаленным PE. Однако данный групповой пакет для группового потока может пересылаться лишь подмножеству PE. В частности, данный групповой поток может передаваться лишь тем PE, у которых получатели заинтересованы в нём. Определение PE, имеющих получателей для данного группового потока выполняется с помощью явного отслеживания в соответствии с [RFC7117].

16.2. P2MP LSP

PE может использовать дерево Inclusive для отправки пакетов BUM. Этот термин заимствован из [RFC7117].

Можно применять разные транспортные технологии в сети сервис-провайдера (SP). Для деревьев Inclusive P-multicast такие технологии включают LSP «один со многими», создаваемые RSVP-TE или Multipoint LDP (mLDP).

16.2.1. Деревья Inclusive

Дерево Inclusive позволяет использовать одно дерево группового распределения, называемое деревом Inclusive P-multicast, в сети SP для передачи всего группового трафика от заданного набора экземпляров EVPN в данный узел PE. Можно нстроить конкретное дерево P-multicast для передачи трафика от сайтов одного или нескольких экземпляров EVPN. Способность передавать через одно дерево трафик нескольких экземпляров EVPN называют агрегированием (Aggregation), а дерево - Aggregate Inclusive P-multicast или, короче, Aggregate Inclusive. Дерево Aggregate Inclusive должно включать каждый узел PE, относящийся к любому из экземпляров EVPN, использующих дерево. Это означает, что PE может получать трафик BUM даже при отсутствии у него заинтересованных в этом трафике получателей.

Деревья Inclusive или Aggregate Inclusive, определённые в этом документе, являются деревьями P2MP (один со многими). Дерево P2MP служит для передачи трафика лишь EVPN CE, подключённых к PE, являющемуся корнем дерева.

Процедуры сигнализации дерева Inclusive совпадают с описанными в [RFC7117] с заменой маршрута VPLS A-D на маршрут Inclusive Multicast Ethernet Tag. Атрибут P-tunnel attribute [RFC7117] для дерева Inclusive анонсируется в маршруте Inclusive Multicast Ethernet Tag, как описано в разделе 11. Обработка трафика с множеством получателей. Отметим, что для дерева Aggregate Inclusive узел PE может «агрегировать» несколько экземпляров EVPN на одном пути P2MP LSP, используя метки восходящего направления. Процедуры агрегирования совпадают с описанными в [RFC7117] с заменой маршрутов VPLS A-D на маршруты Inclusive Multicast Ethernet Tag.

17. Схождение

В этом разделе описано восстановление при различных типах отказов в сети.

17.1. Отказы транзитных каналов и узлов между PE

Использование имеющихся механизмов быстрой перемаршрутизации MPLS может обеспечить восстановление за время порядка 50 мсек при отказе транзитного канала или узла в инфраструктуре, соединяющей узлы PE.

17.2. Отказы PE

Рассмотрим хост CE1, подключенный к PE1 и PE2. При отказе PE1 удалённый узел PE3 может обнаружить это по отказу сессии BGP. Такое обнаружение может занимать доли секунды при использовании обнаружения двухсторонней пересылки (Bidirectional Forwarding Detection или BFD) для сессий BGP. PE3 может обновить своё состояние пересылки для начале передачи трафика CE1 только через PE2.

17.3. Отказы сети на пути от PE к CE

Если прерывается связь между многодомным CE и одним из PE, к которым он подключён, этот PE **должен** отозвать набор маршрутов Ethernet A-D для ES, анонсированный ранее для ES. Это позволяет удаленным PE исключить MPLS next hop для этого PE из набора MPLS next hop, которые могут служить для пересылки трафика CE. При устаревании записи MAC на PE, этот PE **должен** отозвать MAC-адрес через BGP.

Когда тег Ethernet перестаёт применяться в сегменте Ethernet, узел PE **должен** отозвать маршруты Ethernet A-D для EVI, анонсированные для пар <ESI, тег Ethernet tags>, затронутых исключением тега. Кроме того, PE **должен** отозвать маршруты MAC/IP Advertisement, затронутые таким исключением.

Реализации следует применять маршруты Ethernet A-D на ES для оптимизации отзыва маршрутов MAC/IP Advertisement. Когда PE получает отзыв конкретного маршрута Ethernet A-D от анонсировавшего его PE, ему **следует** считать отозванными все маршруты MAC/IP Advertisement, полученные из того же ESI как маршруты Ethernet A-D от анонсирующего PE. Это оптимизирует время схождения сети при отказах между PE и CE.

18. Порядок кадров

Если в MAC первый полубайт (биты 5 - 8) старшего октета MAC-адреса получателя (следует за последней меткой MPLS) имеет значение 0x4 или 0x6, кадр Ethernet может быть ошибочно истолкован пакет IPv4 или IPv6 промежуточными узлами, выполняющими ECMP на основе глубокой инспекции пакетов, что ведёт к распределению пакетов одного потока по разным путям ECMP, способному влиять на задержку. В результате может нарушаться порядок следования пакетов одного потока. Такие нарушения могут возникать в устойчивом состоянии сети без каких-либо отказов и существенно влиять на работу сети. Для предотвращения таких нарушений служит ряд правил.

- При использовании в сети глубокой инспекции пакетов для ECMP **следует** применять Preferred PW MPLS Control Word [RFC4385] со значением 0 (например, 4-октетное поле со значением 0) при передаче пакетов с инкапсуляцией EVPN через MP2P LSP.
- Если в сети применяются энтропийные метки [RFC6790], **не следует** использовать слово управления при передаче пакетов с инкапсуляцией EVPN через MP2P LSP.
- При передаче пакетов с инкапсуляцией EVPN через P2MP LSP или P2P LSP **не следует** применять слово управления.

19. Вопросы безопасности

Соображения безопасности, рассмотренные в [RFC4761] и [RFC4762], применимы к этому документу в части изучения MAC в плоскости данных через устройство присоединения (Attachment Circuit или AC) и лавинной рассылки трафика неизвестных индивидуальных адресатов и сообщений ARP через ядро MPLS/IP. Соображения безопасности из [RFC4364] применимы к этому документу в части изучения MAC в плоскости управления через ядро MPLS/IP. Далее рассмтриваются дополнительные вопросы безопасности.

Как отмечено в [RFC4761], есть два аспекта обеспечения приватности данных и защиты от DoS[1]-атак в VPN - защита плоскости управления и защита путей пересылки. Компрометация плоскости управления может привести к передаче узлом PE данных клиента из той или иной сети EVPN в другую EVPN, отправке данных клиентов EVPN в «черную дыру» или перехватчику, что неприемлемо с точки зрения приватности данных. Кроме того, компрометация плоскости управления может открывать возможность несанкционированного использования данных EVPN (например, репликация трафика в дереве групповой рассылки для усиления DoS-атак с передачей больших объёмов трафика).

Механизмы этого документа используют для плоскости управления протокол BGP. Описанные в [RFC5925] методы помогут проверить подлинность сообщений BGP, осложняя подделку обновлений (может служить для перенаправления трафика EVPN в другой экземпляр EVPN) и отзывов (DoS-атаки). В вариантах магистралей с несколькими AS (b) и (c) в [RFC4364] рассмотрены меры защиты сессий BGP через несколько AS между граничными маршрутизаторами AS (Autonomous System Border Router или ASBR), PE или рефлекторами маршрутов (Route Reflector).

Дополнительное обсуждение вопросов безопасности для BGP приведено в спецификации BGP [RFC4271] и анализе безопасности BGP [RFC4272]. Исходное обсуждение опции подписи TCP MD5 для защиты сессий BGP приведено в [RFC5925], а [RFC6952] включает вопросы применения ключей и аутентификации в BGP.

Отметим, что [RFC5925] может помочь в сохранении приватности меток MPLS - знание меток позволяет перехватить трафик EVPN. Для такого перехвата нужен ещё доступ к пути данных в сети SP. Предполагается, что пользователи VPN принимают меры предосторожности (такие как шифрование) для защиты данных, передаваемых через VPN.

Одним из требований защиты плоскости данных является восприятие меток MPLS лишь от корректных интерфейсов. Для PE такие интерфейсы включают каналы от других маршрутизаторов в AS узла PE. Для ASBR такими интерфейсами служат каналы от других маршрутизаторов в автономной системе ASBR и каналы от других ASBR в AS, где имеются экземпляры данной сети EVPN. Это особенно важно для экземпляров EVPN в нескольких AS.

Важно также ограничивать вредоносный трафик в сети от самозванных адресов MAC. Механизм, описанный в параграфе 15.1. Проблема дублирования MAC, показывает, как можно обнаружить дубликаты MAC-адресов и предотвратить фиктивные перемещения MAC. Механизм, описанный в параграфе 15.2. Закреплённые MAC-адреса, показывает, как привязать MAC-адреса к данному сегменту Ethernet, чтобы при их появлении в других сегментах Ethernet трафик этих MAC-адресов не мог мог попасть из тех сегментов в сеть EVPN.

20. Взаимодействие с IANA

Этот документ определяет EVPN NLRI для передачи в многопротокольных расширениях BGP. NLRI импользует имеющийся идентификатор AFI = 25 (L2VPN). Агентство IANA выделило для BGP EVPN значение SAFI = 70.

Агентство IANA выделило приведённые ниже субтипы EVPN Extended Community из [RFC7153] и данный документ становится единственной ссылкой для них.

0x00	MAC Mobility	[RFC7432]
0x01	ESI Label	[RFC7432]
0x02	ES-Import Route Target	[RFC7432]

Этот документ создаёт реестр EVPN Route Types, новые значения в который вносятся по процедуре RFC Required, заданной в [RFC5226]. Максимальным значением для этого реестра является 255. Исходно выделенные значения указаны ниже.

0	Резерв	[RFC7432]
1	Ethernet Auto-discovery	[RFC7432]
2	MAC/IP Advertisement	[RFC7432]
3	Inclusive Multicast Ethernet Tag	[RFC7432]
4	Ethernet Segment	[RFC7432]

21. Литература

21.1. Нормативные документы

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC4271] Rekhter, Y., Ed., Li, T., Ed., and S. Hares, Ed., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006, <http://www.rfc-editor.org/info/rfc4271>.

[1]Denial-of-service - отказ в обслуживании.

[RFC4360] Sangli, S., Tappan, D., and Y. Rekhter, "BGP Extended Communities Attribute", RFC 4360, February 2006, <http://www.rfc-editor.org/info/rfc4360>.

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006, <http://www.rfc-editor.org/info/rfc4364>.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007, <http://www.rfc-editor.org/info/rfc4760>.

[RFC4761] Kompella, K., Ed., and Y. Rekhter, Ed., "Virtual Private LAN Service (VPLS) Using BGP for Auto-Discovery and Signaling", RFC 4761, January 2007, <http://www.rfc-editor.org/info/rfc4761>.

[RFC4762] Lasserre, M., Ed., and V. Kompella, Ed., "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007, <http://www.rfc-editor.org/info/rfc4762>.

[RFC7153] Rosen, E. and Y. Rekhter, "IANA Registries for BGP Extended Communities", RFC 7153, March 2014, <http://www.rfc-editor.org/info/rfc7153>.

21.2. Дополнительная литература

[802.1D-REV] "IEEE Standard for Local and metropolitan area networks - Media Access Control (MAC) Bridges", IEEE Std. 802.1D, June 2004.

[802.1Q] "IEEE Standard for Local and metropolitan area networks - Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q(tm), 2014 Edition, November 2014.

[RFC4272] Murphy, S., "BGP Security Vulnerabilities Analysis", RFC 4272, January 2006, <http://www.rfc-editor.org/info/rfc4272>.

[RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", RFC 4385, February 2006, <http://www.rfc-editor.org/info/rfc4385>.

[RFC4664] Andersson, L., Ed., and E. Rosen, Ed., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, September 2006, <http://www.rfc-editor.org/info/rfc4664>.

[RFC4684] Marques, P., Bonica, R., Fang, L., Martini, L., Raszuk, R., Patel, K., and J. Guichard, "Constrained Route Distribution for Border Gateway Protocol/MultiProtocol Label Switching (BGP/MPLS) Internet Protocol (IP) Virtual Private Networks (VPNs)", RFC 4684, November 2006, <http://www.rfc-editor.org/info/rfc4684>.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008, <http://www.rfc-editor.org/info/rfc5226>.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010, <http://www.rfc-editor.org/info/rfc5925>.

[RFC6514] Aggarwal, R., Rosen, E., Morin, T., and Y. Rekhter, "BGP Encodings and Procedures for Multicast in MPLS/BGP IP VPNs", RFC 6514, February 2012, <http://www.rfc-editor.org/info/rfc6514>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, November 2012, <http://www.rfc-editor.org/info/rfc6790>.

[RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, May 2013, <http://www.rfc-editor.org/info/rfc6952>.

[RFC7117] Aggarwal, R., Ed., Kamite, Y., Fang, L., Rekhter, Y., and C. Kodeboniya, "Multicast in Virtual Private LAN Service (VPLS)", RFC 7117, February 2014, <http://www.rfc-editor.org/info/rfc7117>.

[RFC7209] Sajassi, A., Aggarwal, R., Uttaro, J., Bitar, N., Henderickx, W., and A. Isaac, "Requirements for Ethernet VPN (EVPN)", RFC 7209, May 2014, <http://www.rfc-editor.org/info/rfc7209>.

Благодарности

Большое спасибо Yakov Rekhter за неоднократное рецензирование документа и важные комментарии, а также за увлекательные обсуждения нескольких тем, которое помогло сформировать документ. Спасибо также Pedro Marques, Kaushik Ghosh, Nischal Sheth, Robert Raszuk, Amit Shukla, Nadeem Mohammed за обсуждения, которые помогли сформировать документ. Спасибо Han Nguyen за комментарии и поддержку работы. Спасибо Steve Kensil и Reshad Rahman за их рецензии. Спасибо Jorge Rabadan за вклад в раздел 5. Спасибо Thomas Morin за обзор документа и вклад в параграф 8.6. Большое спасибо Jakob Heitz за помощь в улучшении некоторых разделов документа.

Спасибо Clarence Filsfils, Dennis Cai, Quaizar Vohra, Kireeti Kompella, Apurva Mehta за их вклад в документ.

Последняя, но не менее важная благодарность Giles Heron (руководитель WG) за подробную рецензию при подготоке WG Last Call и многочисленные ценные предложения.

Участники работы

Кроме авторов, указанных на титульной странице, в создание документа внесли вклад указанные ниже люди.

Keyur Patel
Samer Salam
Sami Boutros
Cisco

Yakov Rekhter
Ravi Shekhar

Juniper Networks

Florin Balus
Nuage Networks

Адреса авторов

Ali Sajassi (editor)
Cisco
EMail: sajassi@cisco.com

Rahul Aggarwal
Arktan
EMail: raggarwa_1@yahoo.com

Nabil Bitar
Verizon Communications
EMail : nabil.n.bitar@verizon.com

Aldrin Isaac
Bloomberg
EMail: aisaac71@bloomberg.net

James Uttaro
AT&T
EMail: uttaro@att.com

John Drake
Juniper Networks
EMail: jdrake@juniper.net

Wim Henderickx
Alcatel-Lucent
EMail: wim.henderickx@alcatel-lucent.com

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru