Network Working Group
Request for Comments: 4364
Obsoletes: 2547
Category: Standards Track

E. Rosen
Cisco Systems, Inc.
Y. Rekhter
Juniper Networks, Inc.

February 2006

BGP/MPLS IP Virtual Private Networks (VPNs)

Виртуальные частные сети IP (VPN) BGP/MPLS

Статус документа

В этом документе содержится спецификация протокола, предложенного сообществу Internet. Документ служит приглашением к дискуссии в целях развития и совершенствования протокола. Текущее состояние стандартизации протокола вы можете узнать из документа Internet Official Protocol Standards (STD 1). Документ может распространяться без ограничений.

Аннотация

Этот документ описывает метод, с помощью которого сервис-провайдер (SP[1]) может использовать опорную сеть IP для предоставления своим абонентам услуг виртуальных частых сетей IP VPN[2]. Метод использует «партнерскую модель», в которой краевые маршрутизаторы абонента (CE[3]) передают свои маршруты краевым маршрутизаторам SP (PE[4]). Здесь нет «наложения», видимого для абонентских алгоритмов маршрутизации и маршрутизаторы CE на разных сайтах не имеют партнерских отношений. Пакеты данных туннелируются через опорную сеть так, что маршрутизаторам ядра не нужно знать маршруты VPN.

Этот документ отменяет RFC 2547.

Оглавление

1. Введение....2
1.1. Виртуальные частные сети....2
1.2. Периметр абонента и периметр провайдера....3
1.3. VPN с перекрывающимися адресами....3
1.4. VPN с разными маршрутами в одну систему....3
1.5. Магистральные маршрутизаторы SP....3
1.6. Безопасность....4
2. Сайты и CE....4
3. VRF - множество таблиц пересылки в PE....4
3.1. VRF и устройства подключения....4
3.2. Связывание пакетов IP с VRF....5
3.3. Заполнение VRF....5
4. Распространение маршрутов VPN по протоколу BGP....5
4.1. Семейство адресов VPN-IPv4....6
4.2. Кодирование RD....6
4.3. Контроль распространения маршрутов....7
4.3.1. Атрибут RT....7
4.3.2. Распространение маршрутов между PE по протоколу BGP....8
4.3.3. Использование рефлекторов....9
4.3.4. Передача VPN-IPv4 NLRI в BGP....10
4.3.5. Построение VPN с использованием RT....10
4.3.6. Распространение маршрутов между VRF в одном PE....10
5. Пересылка....10
6. Поддержка подобающей изоляции VPN....11
7. Как PE узнают маршруты от CE....11
8. Как CE узнают маршруты от PE....13
9. Операторы для операторов....13
10. Магистрали с множеством AS....13
11. Доступ в Internet из VPN....14
12. VPN для управления....15
13. Вопросы безопасности....15
13.1. Уровень данных....15
13.2. Уровень управления....16
13.3. Защита устройств P и PE....16
14. Качество обслуживания....16
15. Расширяемость....16

[1]Service Provider.
[2]Virtual Private Network.
[3]Customer edge.
[4]Provider edge.

16. Взаимодействие с IANA.....17
17. Благодарности.....17
18. Участники работы.....17
19. Нормативные документы.....18
20. Дополнительная литература.....18

1. Введение

Этот документ описывает метод, с помощью которого сервис-провайдер (SP) может использовать опорную сеть IP для предоставления своим абонентам услуг виртуальных частых сетей IP VPN. Метод использует «партнерскую модель», в которой краевые маршрутизаторы абонента (CE) передают свои маршруты краевым маршрутизаторам SP (PE). Сервис-провайдер использует протокол BGP[1] [BGP, BGP-MP] для обмена маршрутами конкретной сети VPN между маршрутизаторами PE, подключёнными к VPN. Способ обмена гарантирует, что маршруты из разных VPN остаются изолированными, даже если адресные пространства VPN перекрываются. Маршрутизаторы PE распространяют маршрутизаторам CE в определённой VPN маршруты от других CE той же сети VPN. Маршрутизаторы CE не являются партнёрами, поэтому здесь не возникает «наложения» видимого алгоритму маршрутизации VPN. Термин IP в «IP VPN» используется для индикации того, что PE получает дейтаграммы IP от CE, проверяет в них заголовки IP и соответствующим образом маршрутизирует.

Каждому маршруту в VPN назначается метка MPLS[2] [MPLS-ARCH, MPLS-BGP, MPLS-ENCAPS]. При распространении протоколом BGP маршрута VPN вместе с ним распространяется и метка MPLS. До того, как абонентский пакет данных попадёт в опорную сеть SP, он инкапсулируется с меткой MPLS, которая соответствует абонентской VPN для маршрута, наиболее точно соответствующего адресу получателя пакета. Затем пакет MPLS инкапсулируется ещё раз (например, с другой меткой MPLS или в заголовок туннеля IP или GRE[3] [MPLS-in-IP-GRE]), чтобы его можно было туннелировать через опорную сеть нужному маршрутизатору PE. Поэтому маршрутизаторам ядра SP не нужно знать маршруты VPN.

Основной целью этого метода является поддержка случаев, когда абонент получает обслуживание в IP-магистрали от одного или нескольких SP, с которыми у него имеются контрактные отношения. Клиентами могут быть предприятия, группы предприятий, которым нужны услуги extranet, поставщики услуг Internet (ISP[4]), поставщики услуг приложений, другие провайдеры VPN, использующие такой же метод для предоставления VPN своим клиентам и т. д. Метод обеспечивает клиентам простоту использования услуг опорной сети. Метод обеспечивает расширяемость и гибкость для SP, а также позволяет им предлагать дополнительные услуги.

1.1. Виртуальные частные сети

Рассмотрим множество «сайтов», подключённых к общей сети, которую будет называть опорной или магистральной (backbone). Применим некое правило для создания набора подмножеств в этом множестве и введём правило, в соответствии с которым два сайта могут быть связаны по протоколу IP через магистральную сеть, если хотя бы одно из подмножеств включает оба сайта.

Эти подмножества будут сетями VPN. Два сайта имеют IP-связность через общую опорную сеть лишь в том случае, когда они входят в одну сеть VPN. Два сайта, не входящие в одну сеть VPN, не будут иметь связности через опорную сеть.

Если все сайты в VPN относятся к одному предприятию, VPN можно рассматривать как корпоративную сеть intranet. Если сайты VPN относятся к разным предприятиям, VPN можно считать сетью extranet. Сайт может входить в несколько VPN, например, в корпоративную сеть и несколько сетей extranet. В общем случае термин VPN здесь используется без разделения сетей на intranet и extranet.

Владельцев сайтов будет называть абонентами (customer), а владельце (операторов) опорных сетей - сервис-провайдерами (SP). Абоненты получают «услуги VPN» от SP.

Абонентом может быть отдельная организация, группа организаций, ISP, провайдер приложений, другой SP, предлагающий услуги VPN своим абонентам и т. п.

Правила определения принадлежности конкретного набора сайтов к VPN задаются политикой абонентов. Некоторые абоненты предпочитают целиком передать выполнение этих правил SP, другие могут захотеть совместного с SP исполнения правил. Описанные здесь механизмы обладают достаточной общностью для поддержки обоих вариантов. Однако большая часть обсуждения относится к варианту совместной реализации правил абонентом VPN и SP.

Описанные в этом документе механизмы позволяют множество вариантов политики. Например, в рамках данной VPN можно разрешить каждому сайту прямой маршрут к каждому из других сайтов (полная связность - full mesh). А можно потребовать маршрутизации трафика между парой сайтов через третий сайт. Это может быть полезно, например, для пропускания трафика через межсетевой экран, расположенный на третьем сайте.

В этом документе рассмотрение ограничено случаем, когда абонент явно покупает услуги VPN у SP или группы SP, которые готовы предоставлять услуги совместно. Т. е. абонент не просто покупает у SP услуги доступа в Internet, самостоятельно организуя VPN через множество соединённых между собой сетей SP, а покупает «сквозной» сервис.

Рассмотрение также ограничено ситуациями, где абонентам предоставляются услуги магистральной сети IP, а не сервис L2, такой как Frame Relay, ATM[5], Ethernet, HDLC[6] или PPP[7]. Абонент может подключаться к магистрали с использованием этих (или иных) услуг L2, но этот сервис завершается на «краю» магистральной сети, где абонентские дейтаграммы IP извлекаются из той или иной инкапсуляции L2.

---

[1] Border Gateway Protocol - протокол граничного шлюза.
[2] Multiprotocol Label Switching - многопротокольная коммутация по меткам.
[3] Generic Routing Encapsulation - базовая инкапсуляция маршрутных данных.
[4] Internet Service Provider.
[5] Asynchronous Transfer Mode - асинхронный режим передачи.
[6] High Level Data Link Control - управление каналом данных на высоком уровне.
[7] Point-to-Point Protocol - протокол «точка-точка».

В оставшейся части раздела описаны свойства, которые следует иметь VPN. Далее описано множество механизмов, которые могут быть применены в модели VPN для достижения этих свойств. В этом разделе также даны определения некоторых технических терминов, используемых в документе.

1.2. Периметр абонента и периметр провайдера

Маршрутизаторы могут быть подключены один к другому или к конечным системам разными способами - соединения PPP, виртуальные устройства ATM (VC[1]) или Frame Relay, интерфейсы Ethernet, виртуальные ЛВС (VLAN[2]) на интерфейсах Ethernet, туннели GRE, L2TP[3] или IPsec и т. п. Термин «устройство присоединения» будет обозначать в общем случае любое из таких соединений. Устройством подключения может быть каналом данных или тем или иным туннелем, позволяющим двум устройствам организовать партнерские отношения на сетевом уровне.

Каждый сайт VPN должен содержать хотя бы одно абонентское краевое устройство (CE[4]). Каждое устройство CE через то или иное устройство (канал) присоединения подключено к одному или нескольким краевым маршрутизаторам провайдера (PE[5]).

Маршрутизаторы в сети SP, не соединённые с устройствами CE, называют провайдерскими (P) маршрутизаторами.

Устройства CE могут быть хостами или маршрутизаторами. В типичной ситуации сайт содержит один или несколько маршрутизаторов, подключённых к маршрутизаторам PE. Подключённые к PE маршрутизаторы сайта будут маршрутизаторами CE. Однако ничто не препятствует подключению хоста без функций маршрутизации непосредственно к PE. Такой хост называют устройством CE.

Иногда физически подключённым к маршрутизатору PE устройством является коммутатор L2. В таких случаях мы **не** называем этот коммутатор устройством CE. Устройствами CE в этом случае будут хосты и маршрутизаторы, взаимодействующие с маршрутизатором PE через такой коммутатор, поскольку инфраструктура уровня 2 прозрачна. Если эта инфраструктура обеспечивает многоточечный сервис, к маршрутизатору PE может быть подключено множество CE через одно устройство присоединения.

Устройства CE логически являются частью абонентской сети VPN, а маршрутизаторы PE и P - частью сети SP.

Устройство присоединения, через которое пакеты проходят от CE к PE, называют «входным устройством присоединения» (ingress attachment circuit) для пакета, а маршрутизатор PE называют «входным PE» для пакета. Устройство присоединения, через которое пакеты проходят от PE к CE называют «выходным устройством присоединения» (egress attachment circuit) для пакетов, а маршрутизатор PE - «выходным PE» для пакета.

Будем говорить, что маршрутизатор PE подключён к определённой сети VPN, если он подключён к устройству CE, расположенному на сайте этой сети VPN. Аналогично, будем называть PE подключённым к определённому сайту, если он подключён к CE на этом сайте.

Когда CE является маршрутизатором, это устройство будет партнёром маршрутизатора (маршрутизаторов) PE, к которому оно подключено, но не будет партнером маршрутизаторов CE на других сайтах. Маршрутизаторы разных сайтов не обмениваются напрямую маршрутными данными и им даже не нужно знать о существовании друг друга. В результате абоненту не нужно управлять опорной сетью или «виртуальной магистралью» и решать вопросы маршрутизации между сайтами. Иными словами, VPN **не** является «наложенной» на сеть SP.

В части управления граничными устройствами поддерживаются чёткие административные границы между SP и его заказчиками. Абонентам не нужен доступ к маршрутизаторам PE и P для управления ими, а SP не требуется доступ для управления устройствами CE.

1.3. VPN с перекрывающимися адресами

Если две сети VPN не имеют общих сайтов, они могут использовать перекрывающиеся пространства адресов. Т. Е данный адрес может служить в VPN V1 адресом системы S1, а в VPN V2 - адресом совершенно другой системы S2. Это обычная ситуация при использовании в нескольких VPN адресов из приватных блоков RFC 1918. В каждой сети VPN адреса, естественно, должны быть однозначными.

Даже в двух VPN с общими сайтами адресные блоки могут перекрываться, пока не возникает необходимости взаимодействия между перекрывающимися адресами и системами на общих сайтах.

1.4. VPN с разными маршрутами в одну систему

Хотя сайт может быть во множестве VPN, не требуется, чтобы маршрут к данной системе на данном сайте был одинаковым во всех этих VPN. Предположим, например, сеть intranet, состоящую из сайтов A, B, C, и сеть extranet из сайтов A, B, C и «чужого» сайта D. Предположим, что на сайте A размещаются серверы, к которым мы хотим предоставить доступ клиентам с сайтов B, C, D. Пусть сайт B служит межсетевым экраном и мы хотим, чтобы весь трафик с сайта D на серверы проходил через межсетевой экран с целью контроля доступа из extranet. Однако трафик с сайта C не нужно пропускать через межсетевой экран, поскольку этот сайт является внутренним.

Можно организовать два маршрута к серверу. Один маршрут (для сайтов B и C) будет направлять трафик напрямую в A. Второй маршрут (для D) будет направлять трафик на межсетевой экран. Если межсетевой экран пропустит трафик, он будет выглядеть как трафик сайта B и попадёт по первому маршруту на сайт A.

1.5. Магистральные маршрутизаторы SP

Магистраль SP состоит из PE, а также других маршрутизаторов (P), которые не соединены с устройствами CE.

---

[1]Virtual Circuit.
[2]Virtual Local Area Network.
[3]Layer 2 Tunneling Protocol - протокол туннелирования L2.
[4]Customer Edge.
[5]Provider Edge.

Если каждый маршрутизатор в магистрали SP будет поддерживать маршрутные данные для всех VPN сервис-провайдера, это приведёт к сложностям при расширении, поскольку число поддерживаемых SP сайтов будет ограничиваться объёмом маршрутной информации, которую способен хранить один маршрутизатор. Поэтому важно, чтобы маршрутная информация об отдельной VPN была представлена лишь на маршрутизаторах PE, служащих для подключения этой VPN. В частности, маршрутизаторам P не нужна **вся** информация для каждой VPN (ситуация может меняться при рассмотрении групповой маршрутизации, подробно рассмотренной в [VPN-MCAST].)

Поскольку владельцам VPN не приходится администрировать магистраль или «виртуальную магистраль», SP не выделяет таких магистралей для сетей VPN. Маршрутизация между сайтами в магистрали выполняется оптимально (с ограничениями в политике, применяемыми для формирования VPN) и не ограничивается искусственной «виртуальной топологией» туннелей.

В разделе 10 рассмотрены некоторые специальные вопросы, связанные с работой через несколько SP.

1.6. Безопасность

Описываемые здесь решения VPN даже без криптографической защиты предназначены для обеспечения уровня безопасности, эквивалентного работе на основе магистралей L2 (например, Frame Relay). Т. е. при отсутствии ошибок в конфигурации и преднамеренного соединения между разными VPN системы одной сети VPN не могут получить доступа к системам другой VPN. Описанные здесь методы не шифруют данные для обеспечения конфиденциальности и не поддерживают способов обнаружения перехвата данных в пути. Если это требуется, можно воспользоваться криптографическими мерами (см., например, [MPLS/BGP-IPsec]). Вопросам безопасности посвящено раздел 13.

2. Сайты и CE

С точки зрения отдельной магистральной сети множество IP-систем можно считать сайтом, если эти системы имеют связность IP, не требующую использования магистрали. В общем случае сайт будет состоять из набора систем, которые находятся географически близко. Однако это верно не всегда. Если две географических точки соединены арендованной линией, на которой работает протокол OSPF[1] [OSPFv2], и эта линия является предпочтительным способом связи между двумя точками, эти точки можно считать одним сайтом даже при наличии в каждой из них своего маршрутизатора CE. Такое толкование сайта является скорей топологическим, нежели географическим. Если арендованный канал перестанет работать или не будет предпочтительным путём, но две точки продолжат взаимодействие с использованием магистрали VPN, сайт разделится на два сайта.

Устройство CE всегда относится к одному сайту (хотя в параграфе 3.2, сказано, что сайт может состоять из множества «виртуальных сайтов»). Однако сайт может относиться к нескольким VPN.

Маршрутизатор PE может подключаться к устройствам CE любого множества разных сайтов в одной или множестве VPN. Устройство CE для надёжности может подключаться к нескольким маршрутизаторам PE одного или разных сервис-провайдеров. Если CE является маршрутизатором, PE и CE будут смежными маршрутизаторами.

Хотя мы в основном считаем сайт базовым блоком взаимодействия, ничто не мешает рассматривать сайты более детально для управления связностью между ними. Например, некоторые системы на сайте могут относиться к внутренней сети (intranet), а также входить в одну или несколько внешний сетей (extranet), тогда как другие входят лишь во внутреннюю сеть. Это может потребовать наличия на сайте двух устройств присоединения к магистрали (одно для intranet, другое для extranet), а также организации межсетевого экрана для устройства присоединения extranet.

3. VRF - множество таблиц пересылки в PE

Каждый маршрутизатор PE поддерживает множество таблиц пересылки. Одна из таблиц является используемой по умолчанию, а другие относятся к маршрутизации VPN и называются VRF[2].

3.1. VRF и устройства подключения

Каждое устройство присоединения PE/CE связано конфигурацией с одной или множеством таблиц VRF. Такие устройства называют устройствами присоединения VRF (VRF attachment circuit).

В простейшем и наиболее распространённом случае устройство присоединения PE/CE связывается с единственной таблицей VRF. Когда пакет IP принимается через определённое устройство присоединения, адрес получателя IP для него отыскивается в связанной таблице VRF. Результат этого поиска определяет маршрутизацию пакета. VRF, используемая входным PE для маршрутизации определённого пакета, называется входной VRF (для пакета имеется и выходная таблица VRF, размещающаяся на выходном PE, как описано в разделе 5).

Если пакет IP приходит через устройство подключения, не связанное с VRF, адрес его получателя отыскивается в принятой по умолчанию таблице пересылки и пакет маршрутизируется в соответствии с ней. Пакеты, маршрутизируемые в соответствии с принятой по умолчанию таблицей, включают пакеты от соседних маршрутизаторов P и PE, а также от обращённых к абонентам устройств присоединения, не связанных с VRF.

Интуитивно можно догадаться, что принятая по умолчанию таблица пересылки является «публичной», а таблицы VRF содержат «приватные» маршруты. Аналогично можно рассматривать устройства присоединения VRF как приватные, а не связанные с VRF устройства присоединения - как публичные.

Если определённое устройство подключения VRF соединяет сайт S с маршрутизатором PE, связность для S (через устройство подключения) может ограничиваться контролируемым набором маршрутов, заданным в соответствующей VRF. Набор маршрутов в этой VRF следует ограничивать маршрутами, ведущими к остальным сайтам VPN, в которую входит S. Когда пакет передаётся из S через устройство присоединения VRF, он может маршрутизироваться PE на другой сайт S', если этот сайт относится к одной сети VPN с сайтом S. Т. е. предотвращается взаимодействие (через PE) между сайтами, не входящими в одну VPN. Коммуникации между сайтами VPN и не входящими в VPN сайтами также пресекаются, поскольку маршруты к сайтам VPN не включаются в принятую по умолчанию таблицу пересылки.

[1] Open Shortest Path First - сначала кратчайший путь.
[2] VPN Routing and Forwarding table - таблица маршрутизации и пересылки VPN.

При наличии множества устройств присоединения, ведущих от сайта S к одному или множеству маршрутизаторов PE может существовать множество таблиц VRF, которые служат для маршрутизации трафика с сайта S. Для подобающего ограничения связности S во всех этих VRF должен быть один набор маршрутов. Другим вариантом является установка разных ограничений связности через различные устройства подключения сайта S. В этом случае маршруты в некоторых VRF, связанных с устройствами подключения S, могут различаться.

Возможно связывание одного устройства присоединения с множеством VRF. Это может быть полезно в тех случаях, когда нужно разделить VPN на несколько суб-VPN, для каждой из которых устанавливаются свои ограничения связности, а выбор между суб-VPN выполняется на основе тех или иных характеристик пакетов. Здесь для простоты всегда предполагается, что устройство присоединения связано с единственной таблицей VRF.

3.2. Связывание пакетов IP с VRF

Когда маршрутизатор PE получает пакет от устройства CE, он должен определить через какое устройство присоединения поступил пакет, поскольку это нужно для выбора таблицы VRF (или набора VRF), которая будет применяться для пересылки пакета. В общем случае для определения доставившего пакет устройства присоединения PE отмечает физический интерфейс, принявший пакет, и может также отмечать некоторые аспекты заголовка L2 в пакете. Например, если входным устройством присоединения для пакета служит Frame Relay VC, отождествление устройства присоединения может быть выполнено по физическому интерфейсу Frame Relay и полю DLCI[1] в заголовке Frame Relay.

Хотя заключение PE о прибытии пакета через определённое устройство подключения может частично определяться заголовком L2 в пакете, у абонента не должно быть возможности ввести провайдера в заблуждение путём создания специальных заголовков L2 с целью привязки пакета к другому устройству присоединения. В приведённом выше примере устройство присоединения частично определяется проверкой поля DLCI в заголовке Frame Relay, которое не может быть установлено абонентом. Это должно быть значение, назначенное SP, поскольку в ином случае пакет просто не попадёт в маршрутизатор PE.

В некоторых случаях абонент может делить сайт на несколько «виртуальных сайтов». SP может обозначить определённый набор таблиц VRF, используемых для маршрутизации пакетов с такого сайта, и позволить абоненту устанавливать некоторые характеристики пакетов, которые позволят выбрать нужную таблицу VRF из этого набора.

Например, каждый виртуальный сайт может быть реализован в форме VLAN, а SP и абонент могут договориться, что пакеты, приходящие из определённого CE, содержат значения VLAN, которые будут применяться для идентификации некоторых VRF. Пакеты от этого CE будут отбрасываться устройством PE, если в них будут указаны несогласованные теги VLAN. Другим способом решения задачи является использование IP-адресов отправителей. В этом случае PE использует адреса отправителей в пакетах, полученных от CE, вместе с принявшим пакет интерфейсом для привязки пакета к определённой таблице VRF. И снова абоненту предоставляется лишь выбор из определённого набора VRF, разрешённого для этого абонента.

Если желательно поместить определённых хост в несколько виртуальных сайтов, этот хост должен будет для каждого пакета определить виртуальный сайт, с которым пакет будет связан. Это можно сделать, например, путём передачи пакетов из разных виртуальных сайтов в разные VLAN или через разные сетевые интерфейсы.

3.3. Заполнение VRF

Какими маршрутами заполняются таблицы VRF?

Рассмотрим пример с тремя маршрутизаторами PE (PE1, PE2, PE3) и тремя маршрутизаторами CE (CE1, CE2, CE3). Предположим, что PE1 узнает от CE1 маршруты, доступные на сайте CE1. Если PE2 и PE3 подключены к CE2 и CE3, соответственно, и имеется VPN V, включающая CE1, CE2 и CE3, маршрутизатор PE1 использует BGP для распространения маршрутизаторам PE2 и PE3 маршрутов, полученных от CE1. PE2 и PE3 используют эти маршруты для заполнения своих таблиц VRF, которые связаны с сайтами CE2 и CE3 соответственно. Маршруты от сайтов, не входящих в VPN V, не будут присутствовать в этих VRF, в результате чего пакеты от CE2 и CE3 не могут быть переданы на сайты, не входящие в VPN V.

Когда мы говорим, что PE «узнает» маршруты от CE, не предполагается какой-либо определённый метод обучения. PE может узнать маршруты с помощью алгоритма динамической маршрутизации, но может также получить их из конфигурации (статические маршруты).

Маршрутизаторам PE нужно также узнать от других PE маршруты, относящиеся к данной сети VPN. Процедуры, используемые для заполнения таблиц VRF нужными маршрутами, описаны в разделе 4.

При наличии множества устройств присоединения между маршрутизатором PE и определенным сайтом эти устройства могут отображаться на одну таблицу пересылки. Но политика может требовать отображения таких устройств на разные таблицы. Например, в соответствии с политикой определённое устройство подключения можно использовать лишь для внутреннего трафика (intranet), а другое - только для extranet (это может быть связано с наличием межсетевого экрана на CE, подключённом к extranet). В этом случае два устройства присоединения будут связаны с разными VRF.

Отметим, что в случае связывания двух устройств присоединения с одной таблицей VRF пакеты, которые PE получает через одно устройство, могут быть доставлены тому же набору адресатов, что и пакеты, принятые PE через другое устройство. Поэтому два устройства подключения не могут быть связаны с одной VRF, пока каждой их устройств CE не имеет одинакового набора VPN.

Если устройство присоединения ведёт на сайт со множеством VPN, это устройство может оставаться связанным с одной таблицей VRF, которая в этом случае будет включать маршруты для всех VPN данного сайта.

4. Распространение маршрутов VPN по протоколу BGP

Маршрутизаторы PE используют BGP для распространения маршрутов VPN между собой.

[1]Data Link Connection Identifier - идентификатор соединения на канальном уровне.

Каждая сеть VPN может использовать своё адресное пространство, поэтому определённый адрес может указывать разные системы в разных VPN. Если два маршрута к одному адресному префиксу IP являются маршрутами в разные системы, важно обеспечить, чтобы протокол BGP не считал эти маршруты сравнимыми. Иначе BGP может выбрать лишь один из таких маршрутов, сделав другую систему недоступной. Кроме того, нужно убедиться, что **политика** применяется для определения маршрутов, по которым нужно отправлять пакеты. С учётом того, что BGP устанавливает несколько маршрутов для одного префикса, в каждой конкретной таблице VRF может присутствовать лишь один такой маршрут.

Для решения этой задачи далее определяется новое семейство адресов.

4.1. Семейство адресов VPN-IPv4

Многопротокольные расширения BGP [BGP-MP] позволяют протоколу BGP передавать маршруты для множества семейств адресов. Здесь вводится новое семейство VPN-IPv4, которое включает 12-байтовое значение, начинающееся с 8-байтового поля RD[1] и заканчивающееся 4 байтами адреса IPv4. Если несколько VPN используют один префикс IPv4, маршрутизаторы PE будут транслировать этот префикс в уникальные адресные префиксы VPN-IPv4. Это позволяет применять один и тот же адрес в разных VPN и обеспечивает BGP поддерживать разные маршруты к одинаковым префиксам в разных VPN.

Поскольку адреса VPN-IPv4 и IPv4 относятся к разным семействам, BGP никогда не сравнивает такие адреса.

RD является лишь числом и не содержит какой-либо специальной информации, не указывая источник маршрута или множества VPN, в которые распространяется маршрут. Назначение RD заключается лишь в создании различаемых маршрутов для одинаковых префиксов IPv4. Для определения области распространения маршрута применяются иные средства (параграф 4.3).

RD можно также использовать для создания множества разных маршрутов в одну и ту же систему. Выше уже рассматривалась ситуация, когда маршруты в одну систему должны различаться для трафика intranet и extranet. Это можно обеспечить путём организации двух разных маршрутов VPN-IPv4 с совпадающей частью IPv4, но разными RD. В результате BGP сможет установить множество разных маршрутов в одну систему и позволит применять правила (параграф 4.3.5) для выбора маршрута конкретных пакетов.

Значения RD структурированы так, что каждый SP может администрировать своё собственное «пространство номеров» (т. е. самостоятельно назначать RD) без конфликтов со значениями RD, выделенными другими SP. Значение RD состоит из трёх полей - двухбайтовое поле типа, поле администратора и поле назначенного номера. Значение поля типа определяет размеры двух оставшихся полей и семантику поля администратора. Поле administrator указывает выделившую номер организацию, а поле assigned number содержит номер, назначенный этой организацией для определённой цели. Например, можно использовать RD, где в поле администратора будет указан номер автономной системы (ASN[2]), а (4-байтовое) поле номера будет содержать значение, выделенное SP, которому принадлежит ASN (номера автономных систем выделяются для SP специальными регистраторами).

Структура RD выбрана так, чтобы SP, предоставляющие магистральные услуги VPN, могли создавать уникальные значения RD при возникновении потребности. Однако структура RD не имеет значения в BGP и не принимается во внимание при сравнении адресных префиксов.

PE нужно настраивать так, чтобы маршруты к определённому CE были связаны с определенным RD. Конфигурация может связывать все маршруты к одному CE с одним значением RD или различать маршруты по разным RD, даже если они приводят к одному CE.

4.2. Кодирование RD

Как было отмечено, адрес VPN-IPv4 состоит из 8-байтового идентификатора маршрута RD, за которым следует 4 байта адреса IPv4. Кодирование RD показано ниже.

- поле Type - 2 байта
- поле Value - 6 байтов

Интерпретация поля Value зависит от значения поля Type. В настоящее время для типа определены значения 0, 1, 2.

- Type 0 - поле Value состоит из двух субполей:
  - Administrator - 2 байта;
  - Assigned Number - 4 байта.

  Субполе Administrator должно содержать номер автономной системы. Если ASN относится к публичному пространству номеров, его значение должно быть выделено соответствующим регистратором (использование ASN из частного пространства настоятельно не рекомендуется). Субполе Assigned Number содержит значение из пространства номеров, администрируемого владельцем ASN, выделенного регистратором.
- Type 1 - поле Value состоит из двух субполей:
  - Administrator - 4 байта;
  - Assigned Number - 2 байта.

  Субполе Administrator должно содержать IP-адрес. Если это адрес из публичного пространства IP, он должен быть выделен соответствующим регистратором (использование IP из частных пространство настоятельно не рекомендуется). Субполе Assigned Number содержит значение из пространства номеров, администрируемого владельцем IP-адреса.
- Type 2 - поле Value состоит из двух субполей:

[1] Route Distinguisher - отличие (обозначение) маршрута.
[2] Autonomous System number.

- Administrator - 4 байта;
- Assigned Number - 2 байта.

Субполе Administrator должно содержать 4-байтовый номер автономной системы [BGP-AS4]. Если ASN относится к публичному пространству номеров, его значение должно быть выделено соответствующим регистратором (использование ASN из частного пространства настоятельно не рекомендуется). Субполе Assigned Number содержит значение из пространства номеров, администрируемого владельцем ASN, выделенного регистратором.

4.3. Контроль распространения маршрутов

В этом параграфе рассматривается способ контроля за распространением маршрутов VPN-IPv4.

Если маршрутизатор PE подключён к некой VPN (будучи подключённым к CE из этой VPN), он узнает некоторые из IP-маршрутов этой VPN от подключённого маршрутизатора CE. Маршруты, полученные от партнерского CE через определённое устройство присоединения, могут быть включены в таблицу VRF, связанную с этим устройством присоединения. Точный выбор устанавливаемых маршрутов определяется способом получения устройством PE маршрутов от CE. В частности, когда PE и CE являются партнёрами для протокола маршрутизации, это определяется решением протокола маршрутизации, как описано в разделе 7.

Полученные маршруты преобразуются с маршруты VPN-IP4 и «экспортируются» в BGP. Если для определённого префикса VPN-IP4 имеется несколько маршрутов, BGP выбирает «лучший», используя процесс решения BGP. Этот маршрут затем распространяется по протоколу BGP другим устройствам PE, которые должны его знать. На этих PE протокол BGP снова выбирает лучший маршрут для конкретного префикса VPN-IP4. Затем выбранные маршруты VPN-IP4 конвертируются обратно в маршруты IP и «импортируются» в одну или несколько таблиц VRF. Выбор маршрутов для установки в VRF зависит от процесса решения метода маршрутизации, используемого между PE и теми CE, которые связаны с соответствующей таблицей VRF. В заключение все маршруты, установленные в VRF, могут быть распространены соответствующим устройствам CE.

4.3.1. Атрибут RT

Каждая таблица VRF связана с одним или несколькими атрибутами RT[1].

При создании маршрута VPN-IPv4 (из маршрута IPv4, который PE узнал от CE) маршрутизатором PE, с ним связывается один или несколько атрибутов RT, которые передаются BGP как атрибуты маршрута.

Любой маршрут, связанный с RT T, должен распространяться каждому маршрутизатору PE, который имеет VRF, связанную с RT T. При получении такого маршрута устройством PE он может быть выбран для включения в таблицы VRF этого PE, которые связаны с RT T (реальная установка зависит от процесса решения BGP и IGP на интерфейсе PE/CE).

Атрибут RT можно считать идентификатором набора сайтов (хотя точнее будет считать идентификатором набора VRF). Связывание определённого атрибута RT с маршрутом позволяет поместить этот маршрут в таблицы VRF, которые применяются для маршрутизации трафика, полученного от соответствующих сайтов.

Имеется набор атрибутов RT, которые маршрутизатор PE присоединяет к маршруту, полученному от сайта S, его можно назвать «целями экспорта» (Export Target). Имеется также набор атрибутов RT, которые PE использует для решения вопроса о размещении полученных от другого PE маршрутов в таблицу VRF, связанную с сайтом S, их можно назвать «целями импорта» (Import Target). Эти два набора различаются и не должны быть одинаковыми. Отметим, что определённый маршрут VPN-IPv4 может быть выбран для установку в ту или иную таблицу VRF лишь при наличии некого RT одновременно в атрибутах RT этого маршрута и целях импорта VRF.

Функция, выполняемая атрибутом RT, похожа на функцию атрибута BGP Communities. Однако формат последнего не подходит для рассматриваемых здесь задач, поскольку он поддерживает лишь 2-байтовые значения. Желательно структурировать формат подобно формату RD (параграф 4.2), чтобы поле типа определяло размер поля администратора, а оставшаяся часть атрибута включала выделенное администратором числовое значение. Это можно сделать с помощью BGP Extended Communities. Рассматриваемые здесь атрибуты RT кодируются как BGP Extended Community Route Targets [BGP-EXTCOMM] и имеют структуру, подобную RD.

Когда узел BGP получает более одного маршрута для одного префикса VPN-IPv4, используются правила предпочтения маршрутов BGP для выбора маршрута VPN-IPv4, устанавливаемого BGP.

Отметим, что маршрут может иметь лишь один идентификатор RD, но множество атрибутов RT. Расширяемость BGP улучшается, если можно использовать один маршрут со множеством атрибутов вместо множества маршрутов. Можно отказаться от атрибутов RT за счёт создания множества маршрутов (т. е. применять больше RD), но это ухудшит расширяемость.

PE может определить атрибуты RT для связывания с маршрутом множеством разных способов. Можно настроить на PE привязку всех маршрутов, ведущих на определённый сайт, с конкретным значением RT. Можно настроить на PE привязку некоторых маршрутов, узнанных для заданного сайта к одному RT, а иных - к другому.

Если PE и CE являются партнёрами BGP (раздел 7), SP может разрешить абоненту в определённых пределах управлять распространением своих маршрутов. SP и абонент должны заранее согласовать набор RT, которые можно присоединять к маршрутам абонентских VPN. Тогда CE сможет присоединять один или несколько таких атрибутов RT к каждому маршруту IP, который он распространяет PE. Это даёт абоненту свободу управлять в реальном масштабе времени правилами распространения своих маршрутов в согласованных рамках. Если маршрутизатору CE разрешено присоединять атрибуты RT к своим маршрутам, PE **должен** отфильтровывать все маршруты, с неразрешенными для абонента атрибутами RT. Если CE не разрешено присоединять RT к своим маршрутам, но он все равно это делает, PE **должен** удалять атрибуты RT перед конвертированием абонентского маршрута в VPN-IPv4.

[1]Route Target - цель маршрута.

4.3.2. Распространение маршрутов между PE по протоколу BGP

Если два сайта VPN подключены к маршрутизаторам PE в одной автономной системе, эти PE могут обмениваться маршрутами VPN-IPv4 через соединение IBGP между ними (термином IBGP обозначают набор протоколов и процедур, используемых для соединений между двумя узлами BGP в одной AS, а термином EBGP - набор процедур, используемых между двумя узлами BGP в разных AS). В дополнение к этому каждый PE может иметь соединение IBGP с рефлектором маршрутов [BGP-RR].

При распространении маршрутизатором PE маршрутов VPN-IPv4 по протоколу BGP он указывает свой адрес в качестве BGP next hop, указываемый в форме VPN-IPv4 с RD = 0 ([BGP-MP] требует использования для next hop того же семейства адресов, что и NLRI[1]). Маршрутизатор также назначает и распространяет метку MPLS (важно отметить, что PE распространяют не просто маршруты VPN-IPv4, а Labeled VPN-IPv4 [MPLS-BGP]). Когда PE получает пакет с такой меткой на вершине стека, он выталкивает метку из стека и обрабатывает пакет соответствующим образом.

PE может распространять напрямую маршруты из VRF, объединять маршруты и распространять агрегаты или распространять часть маршрутов напрямую, а другие агрегировать.

Предположим, что PE назначил метку L для маршрута R и распространяет это отображение по протоколу BGP. Если R является агрегатом набора маршрутов из VRF, PE будет знать, для каких пакетов, пришедших из магистрали с такой меткой, нужно искать адреса получателей в VRF. Когда PE выполняет поиск метки в Label Information Base, он узнает какую VRF нужно использовать. С другой стороны, если R не является агрегатом, то при поиске метки в базе PE узнает выходное устройство присоединения, а также заголовок инкапсуляции для пакета. В этом случае поиск в VRF не нужен.

Предполагается, что в большинстве случаев маршруты **не** будут агрегатами. Агрегированные маршруты могут быть очень полезны, если VRF содержит большое число маршрутов к хостам (например, при доступе по телефонным линиям) или VRF связана с интерфейсом ЛВС, где используется свой заголовок L2 для каждой системы, но маршруты не распространяются в каждую системы.

Наличие своей метки у каждого маршрута зависит от реализации. Существует большое число алгоритмов, которые позволяют определить, присваиваются ли одинаковые метки разным маршрутам.

- Можно выбрать одну метку для таблицы VRF и использовать её во всех маршрутах. Когда выходной PE получит пакет с такой меткой, он должен будет найти IP-адрес получателя пакета в таблице VRF («выходная VRF» для пакета), чтобы определить выходное устройство подключения и соответствующую инкапсуляцию L2.
- Можно выбрать одну метку для каждого устройства присоединения, чтобы она применялась на всех маршрутах с одним «выходным устройством подключения». Это позволяет избежать поиска в выходной VRF, хотя тот или иной поиск может потребоваться для определения инкапсуляции L2 (например, поиск ARP[2]).
- Можно выбрать свою метку для каждого маршрута. Тогда при доступности через несколько устройств подключения маршрутизация PE/CE сможет менять предпочтительный путь с одного устройство подключения на другое без необходимости распространения новой метки для этого маршрута.

Возможны и другие алгоритмы и выбор остаётся за выходным PE.

При таком использовании распространённых по BGP меток MPLS мы предполагаем, что пакет MPLS с такой меткой может быть туннелировать от маршрутизатора, установившего полученный по протоколу BGP маршрут, в маршрутизатор, который служит BGP next hop для этого маршрута. Это требует наличия между этими маршрутизаторами пути с коммутацией по меткам или применения какой-либо иной технологии туннелирования (например, [MPLS-in-IP-GRE]).

Этот туннель может проходить по обычному (best effort) маршруту или следовать маршруту с организацией трафика (traffic-engineered). Между данной парой маршрутизаторов может быть один такой туннель или несколько туннелей с разными характеристиками QoS[3]. Для архитектуры VPN важно лишь наличие такого туннеля. Для обеспечения взаимодействия между системами, реализующими эту архитектуру VPN с использованием путей с коммутацией по меткам MPLS в качестве технологии туннелирования, эти системы **должны** поддерживать протокол LDP[4] [MPLS-LDP]. В частности, **должен** поддерживаться режим Downstream Unsolicited на интерфейсах, которые не являются LC-ATM[5] [MPLS-ATM] или LC-FR[6] [MPLS-FR], и режим Downstream on Demand на интерфейсах LC-ATM и LC-FR.

Если туннель проходит по маршруту best-effort, PE находит маршрут к удалённой точке путём поиска её IP-адреса в принятой по умолчанию таблице пересылки.

Маршрутизатору PE, **пока** он не является рефлектором маршрутов (параграф 4.3.3) или граничным маршрутизатором AS (ASBR) для межпровайдерской VPN (раздел 10), не следует устанавливать маршрут VPN-IPv4, если у него нет хотя бы одной VRF с целью импорта, идентичной одному из атрибутов RT в маршруте. Для отбрасывания маршрутов следует применять входную фильтрацию. Если позднее будет добавлена новая цель импорта в одну из таблиц VRF маршрутизатора PE (операция VPN Join), маршрутизатор должен будет принять маршруты, которые ранее могли отбрасываться. Это можно сделать с помощью механизма обновления, описанного в [BGP-RFSH]. Можно также использовать механизм выходной фильтрации [BGP-ORF], чтобы сделать фильтрацию более динамичной.

Аналогично, если конкретная цель импорта больше не присутствует в таблицах VRF маршрутизаторов PE (в результате операций VPN Prune), PE может отбросить все маршруты и в результате не будет иметь каких-либо целей импорта в своих VRF как одного из атрибутов RT.

Маршрутизатор, не подключенный к VPN и не являющийся рефлектором маршрутов (т. е. P) не устанавливает маршрутов VPN-IPv4.

---

[1] Network Layer Reachability Information - информация о доступности на сетевом уровне.
[2] Address Resolution Protocol - протокол преобразования адресов.
[3] Quality of Service - качество обслуживания.
[4] Label Distribution Protocol - протокол распространения меток.
[5] Label Controlled ATM - ATM с управлениям по меткам.
[6] Label Controlled Frame Relay - Frame Relay с управлениям по меткам.

Обметим, что операции VPN Join и VPN Prune не нарушают работу и не требуют разрыва соединений BGP, если используется механизм обновления [BGP-RFSH].

В результате применения этих правил ни одному маршрутизатору PE не требуется поддерживать маршруты во все VPN - это важно для расширяемости.

4.3.3. Использование рефлекторов

Вместо организации полносвязных соединений IBGP между всеми PE, можно воспользоваться рефлекторами маршрутов BGP RR[1] [BGP-RR] для более эффективной расширяемости. Доступны все обычные методы использования рефлекторов для улучшения расширяемости (например, иерархии рефлекторов).

Рефлекторы маршрутов являются единственными системами, которым нужна маршрутная информация для VPN, не подключённых непосредственно. Однако не требуется, чтобы какой-либо из рефлекторов знал все маршруты VPN-IPv4 для всех VPN, поддерживаемых в магистральной сети.

Ниже описаны два разных способа разделения маршрутов VPN-IPv4 между двумя наборами рефлекторов.

1. В каждом рефлекторе заранее настраивается список целей маршрутов (RT). Для резервирования один и тот же список может быть задан в нескольких рефлекторах. Рефлектор применяет заданный список RT для создания входных фильтров маршрутов. Рефлектор может использовать методы [BGP-ORF], чтобы организовать для каждого из своих партнёров (независимо от того, является партнёр другим рефлектором или PE) набор выходных фильтров маршрутов (ORF[2]) содержащий список заранее заданных RT. Отметим, что рефлекторам следует воспринимать фильтры ORF от других рефлекторов, что означает анонсирование возможности поддержки ORF другим рефлекторам.

   Сервис-провайдер может изменить список заранее настроенных RT на рефлекторе. Когда это сделано, рефлектор меняет таблицы ORF, установленные на всех его партнёрах IBGP. Для снижения частоты изменения конфигурации на рефлекторах маршрутов в каждом рефлекторе можно заранее настроить блок RT. Таким образом, при возникновении потребности в новом атрибуте RT для новой сети VPN на одном или нескольких рефлекторах уже будет (заранее) заданный в настройке атрибут RT.

   Если данный маршрутизатор PE не является клиентом всех рефлекторов, при добавлении на нем новой сети VPN[3] (VPN Join) маршрутизатору нужно будет стать клиентом рефлекторов, которые поддерживают маршруты для этой VPN. Аналогично, удаление имеющейся VPN из PE (VPN Prune) может приводить к ситуации, когда PE уже не нужно быть клиентом некоторых рефлекторов. В любом случае операция Join или Prune не нарушает работу (поскольку используется [BGP-RFSH]) и никогда не требуется прерывать соединение BGP, нужно лишь правильно резервировать его.

2. Другой метод заключается в том, чтобы каждый маршрутизатор PE был клиентом некоторого подмножества рефлекторов. На рефлекторах не настраивается заранее список RT и не выполняется входной фильтрации маршрутов и принимаются все маршруты от клиентов (PE). Рефлекторы отслеживают набор атрибутов RT во всех принимаемых маршрутах. При получении рефлектором от клиента маршрута с отсутствующим в наборе атрибутом RT, этот атрибут сразу же добавляется в набор. С другой стороны, при отсутствии у рефлектора каких-либо маршрутов с конкретным RT из набора, рефлектору следует задерживать (на несколько часов) удаление RT из набора.

   Рефлектор маршрутов использует этот набор для формирования фильтров входящих от других рефлекторов маршрутов. Рефлектор может также использовать фильтры ORF для установки подходящей выходной фильтрации маршрутов на других рефлекторах. Подобно первому варианту, рефлектору маршрутов следует воспринимать фильтры ORF от других рефлекторов. Для этого рефлектор анонсирует поддержку ORF другим рефлекторам маршрутов.

   При изменении маршрутизатором набора атрибутов следует сразу же изменить входную фильтрацию. Кроме того, при использовании рефлектором ORF фильтры ORF тоже следует сразу же изменить в соответствии с изменением набора атрибутов. Если рефлектор не применяет ORF и добавлен новый атрибут RT, после изменения входной фильтрации рефлектор должен передать другим рефлекторам сообщение BGP Refresh.

   Указанная выше задержка в несколько часов позволяет рефлектору удерживать маршруты с данным RT даже после потери своих клиентов, заинтересованных в таких маршрутах. Это избавляет от необходимости заново получать эти маршруты, если «исчезновение» клиента было лишь временным.

   В этой процедуре операции VPN Join и VPN Prune также не нарушают работы.

   Отметим, что этот метод не будет корректно работать, если какой-либо клиент PE имеет таблицу VRF с импортируемым атрибутом RT, который является одним из экспортируемых RT.

В этих процедурах маршрутизатор PE, подключаемый к определённой сети VPN, автоматически обнаруживает другие PE, подключённые к той же VPN. При добавлении нового PE или подключении имеющегося PE к новой сети VPN не требуется менять конфигурацию других маршрутизаторов PE.

Как нет ни одного PE, которому требуется знать все маршруты VPN-IPv4, поддерживаемые в магистральной сети, при использовании этих правил гарантированно не будет требоваться и рефлектор маршрутов (RR), которому нужно знать все поддерживаемые в магистральной сети маршруты VPN-IPv4. В результате число таких маршрутов в магистральной сети не ограничивается возможностями одного устройства и поэтому является практически безграничным.

[1] Route Reflector.
[2] Outbound Route Filter.
[3] Добавление новой сети VPN на PE реально означает добавление нового импорта RT в одну из его таблиц VRF или добавление новой VRF с импортом RT, которого нет в других VRF этого PE.

4.3.4. Передача VPN-IPv4 NLRI в BGP

Для кодирования NLRI применяются многопротокольные расширения BGP [BGP-MP]. Если поле AFI[1] имеет значение 1, а SAFI[2] - 128, поле NLRI будет содержать адрес VPN-IPv4 с меткой MPLS. AFI 1 применяется потому, что протоколом сетевого уровня, связанным с NLRI остаётся IP. Отметим, что эта архитектура VPN не требует поддержки распространения адресов VPN-IPv4 без меток.

Для того, чтобы два узла BGP могли обмениваться помеченными VPN-IPv4 NLRI, они должны использовать BGP Capabilities Advertisement для подтверждения обработки таких NLRI обеими сторонами. Эта процедура выполняется в соответствии с [BGP-MP] и использует код возможности 1 (multiprotocol BGP) с AFI = 1 и SAFI = 128.

Кодирование VPN-IPv4 NLRI с метками задано в [MPLS-BGP] - префикс содержит 8-байтовое поле RD, за которым следует префикс IPv4.

4.3.5. Построение VPN с использованием RT

Путём установки нужных целей импорта и экспорта можно создавать разные типы VPN.

Предположим, что нужно создать полносвязную закрытую группу пользователей, т. е. набор сайтов, где каждый может напрямую передавать трафик другому, но нет возможности обмена трафиком с не входящими в группу сайтами. Каждый сайт связывается с таблицей VRF, выбирается один атрибут RT, затем этот атрибут назначается VRF в качестве цели экспорта и импорта. Никакой другой таблице VRF данный атрибут RT не назначается ни в качестве цели экспорта, ни в качестве цели импорта.

В качестве другого варианта предположим создание VPN типа «звезда» (hub and spoke[3]). Это можно сделать, применяя два значения RT - одно для Hub, другое для Spoke. Для таблиц VRF, связанных с сайтами-концентраторами, Hub будет целью экспорта, а Spoke - целью импорта. Для VRF, связанных с лучами, целью импорта будет Hub, а целью экспорта - Spoke.

Таким образом, методы управления распространением маршрутной информации между разными наборами сайтов обеспечивают гибкость при создании VPN.

4.3.6. Распространение маршрутов между VRF в одном PE

Возможно распространение маршрутов между таблицами VRF, даже находящимися в одном PE, хотя в этом случае нельзя сказать, что они распространяются протоколом BGP. Тем не менее, решение о распространении маршрута из одной VRF в другую в рамках одного PE не отличается от решения о распространении в VRF на других PE. Т. е. это зависит от атрибута RT, связанного с маршрутом (или который был бы связан при распространении по протоколу BGP), и целью импорта другой таблицы VRF.

5. Пересылка

Если промежуточные маршрутизаторы ничего не знают о маршрутах в сети VPN, как же пакеты передаются из одной VPN в другую?

Когда маршрутизатор PE получает пакет IP от устройства CE, он выбирает определённую таблицу VRF для поиска адреса получателя пакета. Этот выбор обусловлен входным устройством присоединения, доставившим пакет. Предположим, что адрес найден в таблице. В результате вы узнаем следующий интервал пересылки (next hop).

Если этот интервал доступен напрямую через устройство присоединения VRF с этого PE (т. е. выходное устройство присоединения находится на том же PE, что и входное устройство присоединения), пакет передаётся в выходное устройство без вталкивания метки MPLS в стек меток пакета.

Если входное и выходное устройства присоединения находятся на одном PE, но связаны с разными таблицами VRF и лучший маршрут для адреса получателя в VRF входного устройства присоединения объединяет несколько маршрутов из VRF выходного устройства присоединения, может потребоваться также поиск адреса получателя в выходной VRF.

Если следующий интервал пересылки (next hop) **не** доступен через устройство присоединения VRF, пакет должен пройти по меньшей мере один интервал пересылки через магистральную сеть. Пакет, таким образом, имеет интервал пересылки BGP Next Hop, с которым связана метка MPLS для маршрута, лучше всего соответствующего адресу получателя. Эту метку называют «меткой маршрута VPN» (VPN route label). Пакет IP помещается в пакет MPLS с меткой маршрута VPN в качестве единственной метки в стеке.

Затем пакет должен туннелироваться по адресу BGP Next Hop.

Если магистральная сеть поддерживает MPLS, пересылка происходит в соответствии с приведённым ниже описанием.

- Маршрутизаторы PE (и любые граничные маршрутизаторы AS[4]), которые распространяют адрес VPN-IPv4, должны добавлять префикс /32 своего адреса в таблицы маршрутизации IGP магистральной сети. Это позволяет MPLS на каждом узле магистральной сети назначать метку, соответствующую маршруту к каждому PE. Для обеспечения взаимодействия разных реализаций требуется поддержка протокола LDP для организации путей с коммутацией по меткам через магистральную сеть. Однако возможны и другие методы организации таких путей (некоторые из таких методов не требуют наличия адресных префиксов /32 в IGP).
- При наличии каких-либо туннелей с организацией трафика (traffic engineering) к BGP next hop и доступности одного или нескольких таких туннелей для рассматриваемого пакета выбирается один из этих туннелей. Туннели будут связаны с меткой MPLS (метка туннеля), которая вталкивается в стек меток MPLS и пакет пересылается на следующий интервал туннеля.
- В остальных случаях выполняются перечисленные ниже действия.

[1] Address Family Identifier - идентификатор семейства адресов.
[2] Subsequent Address Family Identifier - идентификатор последующего семейства адресов.
[3] Концентратор и лучи.
[4] Autonomous System - автономная система.

- Пакет будет иметь IGP Next Hop, указывающий следующий интервал пересылки по маршруту IGP к BGP Next Hop.
- Если BGP Next Hop и IGP Next Hop совпадают и применяется выталкивание метки на предпоследнем интервале, пакет передаётся по адресу IGP Next Hop с единственной меткой маршрута VPN.
- В остальных случаях IGP Next Hop будет иметь метку, назначенную для маршрута, наиболее соответствующего адресу BGP Next Hop. Эту метку называют меткой туннеля и она помещается на вершину стека меток пакета. Затем пакет пересылается по адресу IGP Next Hop.

- MPLS будет предавать пакет через магистральную сеть по адресу BGP Next Hop, где проверяется метка VPN.

Если магистральная сеть не поддерживает MPLS, пакет MPLS с единственной меткой туннеля VPN может быть туннелирован до BGP Next Hop с использованием методов [MPLS-in-IP-GRE]. Пакет выйдет из туннеля на BGP Next Hop, где метка туннеля VPN будет проверяться.

На BGP Next Hop трактовка пакета зависит от метки маршрута VPN (параграф 4.3.2). Во многих случаях по этой метке PE сможет определить устройство присоединения, через которое следует передать пакет (устройству CE), а также подходящий заголовок канального уровня для этого интерфейса. В других случаях PE сможет определить только адрес получателя, по которому нужно выполнить поиск в соответствующей таблице VRF до пересылки пакета устройству CE. Имеются также промежуточные варианты, в которых метка маршрута VPN может определять выходное устройство присоединения для пакета, но придётся выполнять тот или иной поиск (например, ARP) для определения заголовка канального уровня на этом устройстве.

Информация в самом заголовке MPLS и/или информация, связанная с меткой, могут также служить для обеспечения QoS на интерфейсе с устройством CE.

В любом случае прибывший на входное устройство PE пакет IP без метки останется не помеченным и на выходе из выходного устройства PE.

Туннелирование пакетов с метками маршрутов VPN через магистральную сеть позволяет изолировать все маршруты VPN от маршрутизаторов P. Это важно для расширяемости схемы. Магистральной сети не нужно знать и маршрутов к CE, достаточно знать маршруты к PE.

Применительно к туннелям данная спецификация:

- **не** требует туннелей «точка-точка» и позволяет применять туннели «множество с одним» (multipoint-to-point);
- **не** требует какой-либо явной организации туннелей с помощью протоколов сигнализации или вручную;
- **не** требует какой-либо конкретной сигнализации для туннелей;
- **не** требует каких-либо связанных с туннелями состояний в маршрутизаторах P или PE кроме тех, которые нужны для поддержки маршрутной информации и меток MPLS, если они применяются.

Спецификация совместима с использованием туннелей «точка-точка», которые должны явно настраиваться с помощью сигнализации или вручную, и в некоторых случаях могут быть причины для применения таких туннелей.

Вопросы выбора конкретной технологии туннелирования выходят за рамки этого документа.

6. Поддержка подобающей изоляции VPN

Для обеспечения требуемой изоляции VPN важно, чтобы маршрутизаторы магистральной сети не принимали туннелированных пакетов извне, если нет уверенности, что обе конечных точки туннеля находятся за пределами магистрали.

При использовании MPLS в качестве технологии туннелирования маршрутизаторам магистральной сети **недопустимо** воспринимать пакеты с метками от любых смежных устройств, не относящихся к магистрали, если не выполняются оба приведённых ниже условия.

1. Верхняя метка в стеке была действительно распространена данным маршрутизатором магистральной сети не относящемуся к магистрали устройству.
2. Магистральный маршрутизатор может убедиться в том, что использование этой метки будет приводить к выходу пакета из магистральной сети до того, как будут проверены нижележащие метки или заголовок IP.

Первое условие гарантирует, что помеченные пакеты от устройства, не относящегося к магистральной сети, имеют легитимную и корректно установленную метку в своём стеке. Второе условие гарантирует, что магистральные маршрутизаторы не будут «заглядывать» ниже этой метки. Простейшим способом выполнения этих условий является отказ магистрального маршрутизатора воспринимать помеченные пакеты от немагистральных устройств.

Если MPLS не применяется в качестве технологии туннелирования, фильтрация должна быть организована так, чтобы магистральные маршрутизаторы воспринимали пакеты MPLS-in-IP и MPLS-in-GRE лишь в том случае, когда адрес получателя в них будет вызывать передачу за пределы магистральной сети.

7. Как PE узнают маршруты от CE

Маршрутизаторы PE, подключённые к определённой сети VPN, должны знать для каждого ведущего в VPN устройства присоединения, какие из адресов VPN должны быть доступны через данное устройство присоединения.

PE транслирует эти адреса в адреса VPN-IPv4, используя настроенные идентификаторы RD. Затем эти маршруты VPN-IPv4 устройство PE трактует как входную информацию BGP. Маршруты из сайта VPN **не** передаются в IGP магистральной сети.

Конкретный метод распространения маршрутов PE/CE зависит от того, находится ли конкретное устройство CE в «транзитной VPN». Транзитной считается сеть VPN, которая содержит маршрутизатор, получающий маршруты от «третьей стороны» (т. е. от маршрутизатора, не входящего в VPN и не являющегося PE) и распространяющий их

маршрутизатору PE. VPN, не являющиеся транзитными, называют оконечными VPN (stub VPN). Подавляющее большинство сетей VPN, включая практически все корпоративные сети, в этом смысле являются оконечными.

Возможные методы распространения маршрутов PE/CE перечислены ниже.

1. Может использоваться статическая маршрутизация (т. е. настройка). Очевидно, что это подходит лишь для оконечных VPN.

2. Маршрутизаторы PE и CE могут быть партнёрами RIP[1] [RIP] и CE может использовать RIP для информирования PE о наборе адресных префиксов, доступных на сайте CE. При настройке RIP на CE следует принять меры, чтобы префиксы других сайтов (т. е. префиксы, полученные CE от маршрутизатора PE) никогда не анонсировались PE. Точнее говоря, если маршрутизатор PE (скажем, PE1) получает маршрут VPN-IPv4 R1 и распространяет маршрут IPv4 R2 устройству CE, маршрут R2 не должен распространяться обратно с сайта CE маршрутизатору PE (скажем, PE2, где PE1 и PE2 могут быть одним или разными маршрутизаторами), если PE2 не отображает R2 на маршрут VPN-IPv4, который отличается (т. е. имеет иной идентификатор RD) от R1.

3. Маршрутизаторы PE и CE могут быть партнёрами OSPF. PE, являющийся OSPF-партнером маршрутизатора CE с точки зрения CE будет находиться в области 0. Если PE является партнёром OSPF маршрутизаторов CE из разных VPN, PE должен поддерживать отдельные экземпляры OSPF для каждой VPN.

   Маршруты IPv4, которые PE получает от CE по протоколу OSPF, распространяются в BGP как маршруты VPN-IPv4. Атрибуты Extended Community служат для передачи вместе с маршрутом всех информации, требуемой для распространения маршрута другим CE в сети VPN в форме подходящих анонсов OSPF LSA[2]. Маршруты OSPF помечаются, чтобы принятые из магистрали MPLS/BGP маршруты не передавались обратно.

   Спецификация полного набора процедур использования OSPF между PE и CE приведена в [VPN-OSPF] и [OSPF-2547-DNBIT].

4. Маршрутизаторы PE и CE могут быть партнёрами BGP и CE может применять BGP (в частности, EBGP) для передачи PE набора адресных префиксов на сайте CE (это можно применять в транзитных и оконечных VPN).

   Этот метод обеспечивает многочисленные преимущества, перечисленные ниже.

   a) В отличие от IGP это не требует от PE использования множества экземпляров протокола маршрутизации для взаимодействия с множеством CE.

   b) Протокол BGP создан для решения таких задач - передачи маршрутной информации между системами с раздельным администрированием.

   c) Если сайт имеет «закулисные соединения BGP», т. е. маршрутизаторы с соединениями BGP с сайтами, не являющимися PE, эта процедура будет работать корректно при любых обстоятельствах, в отличие от других процедур, которые могут работать не во всех случаях.

   d) Использование BGP позволяет CE простым путём передавать атрибуты маршрутов маршрутизатору PE. Полная спецификация набора передаваемых атрибутов выходит за рамки документа. Однако некоторые примеры приводятся ниже.

      - CE может предложить конкретный атрибут RT для каждого маршрута из числа RT, которые PE уполномочен присоединять к маршрутам. PE будет тогда добавлять предложенный атрибут RT, а не полный набор атрибутов. Это даёт администратору CE возможность динамического управления распространением маршрутов от CE.

      - Могут быть определены дополнительные типы атрибутов Extended Community для передачи между маршрутизаторами CE без изменения атрибутов устройствами PE. Это позволит администраторам CE реализовать дополнительные правила фильтрации маршрутов сверх тех, которые выполняются PE. Эти дополнительные правила не требуется согласовывать с SP.

      С другой стороны, у администраторов CE может не оказаться опыта работы с BGP.

      Если сайт не является транзитной VPN, ему не нужен уникальный номер автономной системы (ASN). Каждый CE, чей сайт не является транзитной VPN, может применять один и тот же номер ASN, который можно выбрать из приватных номеров ASN, поскольку он будет исключаться маршрутизатором PE. Маршрутные петли предотвращаются использованием атрибута Site of Origin (см. ниже).

      Что делать в случаях, когда набор сайтов образует транзитную сеть VPN? Это обычно возникает лишь в тех случаях, когда VPN является сетью сервис-провайдера ISP[3], который сам покупает магистральные услуги у другого SP (оператора для операторов). В этом случае лучшим вариантом организации VPN будет поддержка MPLS маршрутизаторами CE и применение метода, описанного в разделе 9.

Когда нам не нужно различать способы информирования маршрутизаторов PE о существующих на сайте адресных префиксах, можно просто говорить, что PE «узнает» маршруты от данного сайта. Это включает случай задания маршрутов на PE вручную.

До того, как PE сможет распространять маршруты VPN-IPv4, узнанные от сайта, он должен назначить атрибут RT (параграф 4.3.1) для маршрута и может назначить также атрибут Site of Origin.

При использовании атрибута Site of Origin он кодируется в форме Route Origin Extended Community [BGP-EXTCOMM]. Назначение этого атрибута заключается в однозначном указании набора маршрутов, полученного от конкретного сайта. Этот атрибут требуется в некоторых случаях для обеспечения гарантии того, что маршрут, полученный от конкретного сайта через определённое соединение PE - CE, не будет распространяться на этот же сайт через другое соединение PE - CE. Это полезно, в частности, при использовании протокола BGP между PE и CE с различными номерами ASN на разных сайтах.

---

[1] Routing Information Protocol - протокол маршрутной информации.
[2] Link State Advertisement - анонс состояния канала.
[3] Internet Service Provider - поставщик услуг доступа в Internet.

8. Как CE узнают маршруты от PE

В этом разделе предполагается, что устройство CE является маршрутизатором.

Если PE помещает определённый маршрут в таблицу VRF и использует его для маршрутизации пакетов, полученных от конкретного CE, в общем случае PE может распространять этот маршрут CE (если это разрешают правила протокола, используемого между CE и PE). Например, если отдельный протокол PE - CE использует «расщепление горизонта» (split horizon), некоторые маршруты из VRF невозможно распространять устройству CE). Можно добавить одно ограничение на распространение маршрутов от PE к CE - если атрибут Site of Origin указывает определённый сайт, маршрут недопустимо распространять какому-либо CE на этом сайте.

Однако в большинстве случае для PE достаточно просто распространять устройству CE используемый по умолчанию маршрут (в некоторых случаях достаточно просто задать на CE использование по умолчанию маршрута к PE). Это обычно будет работать на любом сайте, которому не требуется распространять используемый по умолчанию маршрут другим сайтам (например, если один сайт в корпоративной VPN обеспечивает доступ в Internet для всей компании, этот сайт должен распространять принятый по умолчанию маршрут другим сайтам, но не может распространять его себе).

Какая бы процедура ни использовалась для распространения маршрутов из CE в PE, она же будет применяться для распространения маршрутов из PE в CE.

9. Операторы для операторов

В некоторых случаях VPN может быть сетью ISP с его партнерскими связями (peering) и политикой маршрутизации, а иногда VPN может быть сетью SP, предоставляющего услуги VPN своим абонентам. Такие сети VPN также могут использовать магистральные услуги других SP (оператора для операторов) на основе методов, описанных в этом документе. Однако в таких случаях требуется поддержка MPLS маршрутизаторами CE. Требования приведены ниже.

- Маршрутизаторам CE следует распространять в PE **только** внутренние маршруты VPN. Это позволит VPN работать в режиме оконечной VPN.
- Маршрутизаторам CE следует поддерживать MPLS в части возможности принимать метки от маршрутизаторов PE и передавать тем пакеты с метками. Эти маршрутизаторы не распространяют своих меток.
- PE следует распространять маршрутизаторам CE метки для маршрутов, которые они распространяют CE.

  PE недопустимо распространять одну метку двум разным CE, если не выполняется одно из условий:

  - два CE связаны с одним набором таблиц VRF;
  - PE поддерживает разные отображения входящих меток [MPLS-ARCH] для каждого CE.

  Кроме того при получении PE пакета с меткой от CE он должен убедиться, что эта метка распространялась данному CE.
- Маршрутизаторам разных сайтов следует организовать между собой соединения BGP для обмена внешними маршрутами (т. е. маршрутами, ведущими за пределы VPN).
- Все внешние маршруты должны быть известны устройствам CE.

Когда маршрутизатор CE выполняет поиск по адресу получателя, он будет находить внутренний адрес, который обычно указывает BGP next hop для пакета. CE соответствующим образом помечает пакет и передаёт его маршрутизатору PE, который вместо поиска IP-адреса получателя в таблице VRF использует верхнюю метку MPLS для выбора BGP next hop. В результате, если до BGP next hop более одного интервала пересылки (hop), верхняя метка стека будет заменяться двумя метками - для туннеля и маршрута VPN. Если до BGP next hop лишь один интервал, верхняя метка может быть просто заменена меткой маршрута VPN. Если входной PE является одновременно выходным PE, верхняя метка будет просто выталкиваться из стека. При передаче пакета от выходного PE маршрутизатору CE, пакет будет иметь на одну метку MPLS меньше, чем было получено на входном PE.

В приведённой выше процедуре лишь маршрутизаторам CE в VPN требуется поддержка MPLS. С другой стороны, если все маршрутизаторы определённого сайта VPN поддерживают MPLS, от CE не требуется больше знать все внешние маршруты. Достаточно, чтобы внешние маршруты были известны маршрутизаторам, отвечающим за добавление меток в стек пакетов без меток и наличие пути с коммутацией по меткам от этих маршрутизаторов к их BGP-партнерам на других сайтах. В этом случае для каждого внутреннего маршрута, который CE распространяет маршрутизатору PE, должна добавляться метка.

10. Магистрали с множеством AS

Что будет, если два сайта VPN подключены к разным автономным системам (например, через разных SP)? Маршрутизаторы PE, связанные с данной VPN с этом случае не смогут организовать соединения IBGP между собой или с общим рефлектором маршрутов. Здесь потребуется организация соединений EBGP для распространения адресов VPN-IPv4.

Для решения таких задач имеется много способов, которые перечислены ниже в порядке роста возможностей расширения.

a) Соединения между VRF на граничных маршрутизаторах AS.

   В этом случае маршрутизатор PE одной AS соединяется напрямую с PE другой AS. Соединения между PE организуются через множество субинтерфейсов - по меньшей мере один интерфейс на каждую сеть VPN, для которой требуется работа через разные AS. Каждый маршрутизатор PE будет рассматривать другой PE как маршрутизатор CE. Т. е. маршрутизаторы PE связывают каждый субинтерфейс с VRF и используют EBGP для распространения адресов IPv4 без меток другому PE.

   Эта процедура не требует поддержки MPLS на границе между AS. Однако она не обеспечивает достаточных возможностей расширения по сравнению с другими вариантами.

b) EBGP служит для распространения маршрутов VPN-IPv4 с метками из одной AS в соседнюю.

В этом варианте маршрутизаторы PE применяют IBGP для распространения маршрутов VPN-IPv4 граничному маршрутизатору ASBR[1] или рефлектору маршрутов, для которого ASBR является клиентом. ASBR затем использует EBGP для распространения маршрутов VPN-IPv4 с метками маршрутизатору ASBR в другой AS, который в свою очередь распространяет из маршрутизаторам PE в своей AS и возможно другим ASBR и т. д.

При использовании такой процедуры маршруты VPN-IPv4 следует воспринимать лишь на соединениях EBGP в приватных точках партнёрства (peering point) как часть доверенных отношений между SP. Маршруты VPN-IPv4 не следует воспринимать из публичной сети Internet (и распространять туда) или от недоверенных партнёров BGP. Маршрутизатору ASBR не следует воспринимать помеченные пакеты от партнёра EBGP, если он не распространял этому партнёру верхнюю метку из пакета.

При наличии множества VPN с подключениями к разным AS не требуется наличие между автономными системами одного ASBR, которому известны маршруты для всех VPN между этими двумя AS, и может использоваться множество ASBR, каждый из которых поддерживает маршруты для определённого подмножества VPN.

Эта процедура требует наличия пути с коммутацией по меткам от входного PE для пакета до его выходного PE. Поэтому нужны соответствующие отношения доверия между AS на пути пакетов. Также между SP должно иметься соглашения о выборе граничных маршрутизаторов для получения маршрутов с разными атрибутами RT.

c) Распространение EBGP для помеченных маршрутов VPN-IPv4 между AS через множество этапов пересылки с распространением EBGP для помеченных маршрутов IPv4 из одной AS в соседнюю.

В этом варианте маршруты VPN-IPv4 не поддерживаются и не распространяются маршрутизаторами ASBR, которые однако должны поддерживать помеченные маршруты IPv4 с префиксом /32 для маршрутизаторов PE внутри AS. Для распространения маршрутов в другие AS применяется EBGP. маршрутизаторы ASBR в транзитных AS будут также передавать по EBGP помеченные маршруты /32. Это обеспечивает создание пути с коммутацией по меткам от входного маршрутизатора PE к выходному PE. Маршрутизаторы PE в разных AS могут организовывать между собой соединения EBGP с множеством этапов пересылки и обмениваться через них маршрутами VPN-IPv4.

Если маршруты /32 для устройств PE известны маршрутизаторам P в каждой AS, все работает хорошо. Если маршруты /32 для PE **не** известны маршрутизаторам P (кроме ASBR), эта процедура требует от входного PE помещать в пакет стек из 3 меток. Нижняя метка назначается выходным PE, соответствующим адреса получателя пакета в определённой таблице VRF. Среднюю метку назначает маршрутизатор ASBR, соответствующий маршруту /32 к входному PE. Верхняя метка назначается узлом IGP Next Hop для PE, соответствующим маршруту /32 к ASBR.

Для улучшения расширяемости можно использовать многоэтапные соединения EBGP лишь между рефлекторами маршрутов в разных AS (однако при распространении рефлекторами маршрутов через такое соединение они не меняют атрибут BGP next hop). Тогда маршрутизаторы PE будут иметь лишь соединения IBGP с рефлекторами маршрутов в своей AS.

Эта процедура похожа на процедуры, описанные в разделе 9. Подобно предыдущей процедуре, она требует наличия пути с коммутацией по меткам от входного PE к выходному PE.

11. Доступ в Internet из VPN

Многим сайтам VPN нужен доступ в публичную сеть Internet, а также к другим сайтам VPN. Ниже описано несколько вариантов решения этих задач.

1. В некоторых VPN один или несколько сайтов могут получать доступ в Internet с помощью «шлюза Internet» (возможно, межсетевого экрана), подключённого к интерфейсу в сеть ISP без таблицы VRF. Это может быть тот же провайдер, который обеспечивает услуги VPN, или другой ISP. Трафик на шлюз и от него будет маршрутизироваться с использованием принятой по умолчанию таблицы пересылки PE.

   В этом случае сайты с доступом в Internet могут распространять принятый по умолчанию маршрут своим PE, которые затем распространят его другим PE на других сайтах VPN. Это обеспечит доступ в Internet на всех сайтах VPN.

   Для корректной обработки трафика из Internet провайдер ISP должен распространить в сеть Internet маршруты, ведущие к адресам VPN. Это никак не связано с процедурами распространения маршрутов, описанными в этом документе. Внутренняя структура VPN в общем случае не видна из Internet и эти маршруты будут просто вести к интерфейсу без VRF, подключённому к шлюзу Internet в сети VPN.

   В этой модели не происходит обмена маршрутами между принятой по умолчанию таблицей пересылки PE и какими-либо из его VRF. Процедуры распространения маршрутов VPN и Internet полностью независимы.

   Отметим, что хотя некоторые сайты VPN используют интерфейс VRF для обмена данными с Internet, в конечном итоге все пакеты в сеть Internet и из неё проходят через интерфейс без VRF перед выходом из VPN или входом в неё, поэтому этот вариант называется доступом в Internet без VRF (non-VRF Internet access).

   Отметим, что маршрутизатор PE, к которому подключён интерфейс без VRF, не обязан поддерживать все маршруты Internet в принятой по умолчанию таблице пересылки. В этой таблице достаточно используемого по умолчанию маршрута (default), который ведёт к другому маршрутизатору (возможно, смежному), имеющему маршруты Internet. Вариантом этой схемы является туннелирование пакетов, принятых интерфейсом без VRF, от маршрутизатора PE к другому маршрутизатору, где поддерживается полная таблица маршрутов Internet.

[1] Autonomous System Border Router - граничный маршрутизатор автономной системы.

2. Некоторые VPN могут получать доступ в Internet через интерфейс VRF (VRF Internet access). Если пакет принят PE через интерфейс VRF и адрес получателя в нем не соответствует ни одному маршруту из VRF, этот пакет может быть сопоставлен с принятой по умолчанию таблицей пересылки PE. При обнаружении соответствия пакет можно переслать обычным путём через магистральную сеть в Internet, не применяя MPLS.

   Для трафика в обратном направлении (из Internet на интерфейс VRF), некоторые из маршрутов VRF должны экспортироваться в таблицу пересылки Internet. Нет нужды говорить о том, что это должны быть маршруты к уникальным в глобальном масштабе адресам.

   В этой схеме используемая по умолчанию таблица пересылки может иметь полный набор маршрутов Internet или содержать лишь принятый по умолчанию маршрут к другому маршрутизатору, имеющему полный набор.

3. Предположим, что PE имеет возможность сохранять не связанные с VPN маршруты в таблице VRF. Если пакет имеет адрес получателя, соответствующий маршруту «не VPN», он будет передаваться естественным путём, а не через MPLS. Если VRF содержит принятый по умолчанию маршрут «не VPN», все пакеты для публичной сети Internet будут соответствовать ему и передаваться естественным путём по принятому по умолчанию маршруту на следующий интервал. Там будет выполняться поиск по адресу получателя в принятой по умолчанию таблице пересылке и может быть выбран более точный маршрут.

   Этот метод доступен лишь в том случае, когда ни один из маршрутизаторов CE не распространяет принятый по умолчанию маршрут.

4. Можно также получать доступ в Internet через интерфейс VRF путём включения в таблицу VRF маршрутов Internet. По сравнению с моделью 2 это избавляет от второй операции поиска, но требует репликации маршрутов Internet в каждую таблицу VRF.

   При использовании этого метода SP может захотеть использовать в качестве интерфейса в Internet интерфейс VRF и применять методы раздела 4 для распространения маршрутов Internet как маршрутов VPN-IPv4 в другие VRF.

Следует понимать, что по умолчанию не происходит обмена маршрутами между VRF и принятой по умолчанию таблицей пересылки. Такой обмен выполняется **лишь** по соглашению между абонентом и SP и только при соответствии правилам абонента.

12. VPN для управления

Эта спецификация не требует наличия адреса у субинтерфейса, соединяющего маршрутизаторы PE и CE. Если этот интерфейс имеет адрес, спецификация позволяет использовать адресное пространство VPN или SP.

Если маршрутизатор CE управляется сервис-провайдером, системе управления сетью SP будет нужен доступ к маршрутизатору CE. В этом случае субинтерфейсу между CE и PE следует назначать адрес из пространства SP и этот адрес должен быть уникальным в рамках провайдера. Системе управления сетью следует самостоятельно подключиться к маршрутизатору PE (точнее, присутствовать на сайте, подключённом к PE) через интерфейс VRF. Адрес системы управления будет экспортироваться во все таблицы VRF, связанные с интерфейсами к маршрутизаторам CE, которыми управляет SP. Адреса маршрутизаторов CE будут экспортироваться в VRF, связанные с системой сетевого управления, но не в остальные VRF.

Это разрешает взаимодействие между CE и системой сетевого управления но препятствует ненужному взаимодействию между маршрутизаторами CE.

Одним из способов организации нужного экспорта и импорта маршрутов является использование двух атрибутов RT, например T1 и T2. Если конкретный интерфейс VRF подключён к маршрутизатору CE, управляемому SP, для этой таблицы VRF настраиваются:

- импорт маршрутов с атрибутом T1;
- добавление атрибута T2 к адресам, настроенным на каждой стороне интерфейсов VRF.

Если конкретный интерфейс VRF подключён к системе управления SP, для таблицы VRF настраивается добавление атрибута T1 к адресу этой системы и импорт маршрутов с атрибутом T2.

13. Вопросы безопасности

13.1. Уровень данных

Под безопасностью уровня данных мы понимаем защиту от указанных ниже угроз.

- Передача пакетов из туннеля VPN на сайт за пределами VPN, не разрешённых политикой VPN.
- Передача на один из сайтов VPN пакетов извне VPN, не разрешённых политикой VPN.

Ниже перечислено несколько условий, при соблюдении которых защита уровня данных, обеспечиваемая этой архитектурой, виртуально идентична защите VPN в магистралях Frame Relay или ATM. Если находящиеся под управлением SP устройства настроены корректно, данные не смогут попадать в сеть VPN или уходить их неё без должных полномочий.

1. Магистральный маршрутизатор не воспринимает помеченные пакеты через конкретный канал данных, пока он не уверен, что к каналу подключены лишь доверенные системы или не знает, что такие пакеты покинут магистральную сеть до того, как будет просматриваться заголовок IP или нижележащие метки в стеке.
2. Помеченные маршруты VPN-IPv4 не воспринимаются из недоверенных или ненадёжных параненов по маршрутизации.
3. Не было успешных атак на уровень управления.

Условие 1 можно выразить более точно. Следует отбрасывать помеченные пакеты, полученные от конкретного соседа, пока не выполняется одно из приведённых ниже условий.

- Верхняя метка в пакете имеет значение, которое принимающая система распространила данному соседу.
- Верхняя метка в пакете имеет значение, которое принимающая система распространила системам, находящимся за этим соседом (т. е. известно, что путь от системы, которой распространена метка, до принимающей системы проходит через этого соседа).

Условие 2 наиболее интересно для случая организации VPN с участием нескольких провайдеров (раздел 10). Для межпровайдерских VPN, создаваемых по схеме b) из раздела 10, условие 2 легко проверить (безопасность для схемы c) из раздела 10 требует дополнительного изучения).

Важно подчеркнуть, что использование MPLS значительно упрощает обеспечение безопасности уровня данных по сравнению с использованием той или иной формы туннелей IP вместо внешней метки MPLS. Очень просто сделать так, чтобы граничные маршрутизаторы отказывались воспринимать помеченные пакеты, если для них не выполняется условие 1. Существенно сложнее настроить маршрутизатор так, чтобы он отвергал туннелированные пакеты IP, в которых получателем указан данный маршрутизатор PE, - это можно сделать, но за счёт более сложного управления и снижения производительности.

Туннелирование MPLS-in-IP и MPLS-in-GRE описано в [MPLS-in-IP-GRE]. Если нужно применять такие туннели для передачи пакетов VPN, следует разобраться с вопросами безопасности, упомянутыми в разделе 8 указанного документа. Любая реализация BGP/MPLS IP VPN, разрешающая туннелировать пакеты VPN в соответствии с указанным документом, **должна** поддерживать IPsec. Если туннель не защищён с помощью IPsec, фильтрация по адресам IP на граничных маршрутизаторах, описанная в параграфе 8.2 упомянутого документа, остаётся единственным способом гарантировать, что пакеты, выходящие из туннеля на конкретном выходном PE, действительно были помещены в туннель разрешённым узлом в начале туннеля (т. е. пакет не содержит обманный адрес отправителя). Поскольку граничные маршрутизаторы зачастую фильтруют лишь по адресам отправителей, фильтрация пакетов может быть неэффективной, пока выходной маршрутизатор PE не может проверить IP-адрес отправителя любого туннелированного пакета и сравнить его со списком адресов IP, которые разрешены в начальной точке туннеля. Все реализации, разрешающие туннели MPLS-in-IP и/или MPLS-in-GRE без IPsec **должны** позволять выходному PE такую проверку IP-адресов отправителей для всех принимаемых из туннеля пакетов.

При подключении множества CE к маршрутизатору PE через интерфейс ЛВС для обеспечения безопасности должно выполняться одно из приведённых ниже условий.

1. Все маршрутизаторы CE в ЛВС должны относиться к одной сети VPN.
2. Доверенный и защищённый коммутатор ЛВС делит сеть на несколько VLAN, каждая из которых включает лишь системы одной сети VPN. В этом случае коммутатор будет добавлять подходящий тег VLAN ко всем пакетам, пересылаемым маршрутизатору PE.

Эта архитектура не обеспечивает криптографической защиты конфиденциальности, так же как в Frame Relay или ATM VPN. Эти варианты архитектуры совместимы с использованием криптографии на базе CE-CE, если это нужно.

Использование криптографии между устройствами PE требует дополнительного исследования.

13.2. Уровень управления

Безопасность уровня данных, рассмотренная выше, зависит от защиты уровня управления. Для обеспечения безопасности не следует разрешать соединения BGP или LDP с недоверенными партнёрами. Следует применять опцию проверки подлинности TCP/IP MD5 [TCP-MD5] с обоими протоколами. Протокол маршрутизации в сети SP также следует защищать аналогичным способом.

13.3. Защита устройств P и PE

При нарушении физической защиты этих устройств уровень данных также подвергается рискам.

Следует применять обычные меры защиты от использования трафика IP из публичной сети Internet для изменения конфигурации этих устройств или организации против них атак на службы (Denial of Service).

14. Качество обслуживания

Хотя качество обслуживания не рассматривается в этом документе, вопросы QoS[1] являются важной частью сервиса VPN. В сетях MPLS/BGP VPN имеющиеся возможности L3 QoS могут применяться к помеченным пакетам за счёт использования экспериментальных битов в промежуточном заголовке [MPLS-ENCAPS] или, при работе через магистрали ATM, за счёт использования ATM QoS. Организация трафика, описанная в [MPLS-RSVP], также применима для сетей MPLS/BGP VPN. Организацию трафика можно даже использовать для организации путей с коммутацией по меткам и заданными параметрами QoS между парой конкретных сайтов, если это нужно. При организации MPLS/BGP VPN через несколько SP может быть полезна архитектура, описанная в [PASTE]. SP может применять возможности интегрированного (intserv[2]) или дифференцированного (diffserv[3]) для конкретной сети VPN по своему усмотрению.

15. Расширяемость

В этом документе были рассмотрены вопросы расширяемости и в этом разделе приведено краткое резюме по характеристикам расширяемости этой модели.

Магистральная сеть SP состоит из (a) маршрутизаторов PE, (b) рефлекторов маршрутов BGP, © маршрутизаторов P (не PE и не рефлекторы маршрутов), а в случае межпровайдерских VPN включает (d) граничные маршрутизаторы ASBR.

[1] Quality of Service - качество обслуживания.
[2] Integrated Services - интегрированное обслуживание.
[3] Differentiated Services - дифференцированное обслуживание.

Маршрутизаторы P не поддерживают маршрутов VPN и для правильной пересылки трафика VPN им достаточно поддерживать маршруты к PE и ASBR. Использование двух уровней меток обеспечивает возможность сохранения маршрутов VPN при передаче трафика через маршрутизаторы P.

Маршрутизаторы PE поддерживают маршруты VPN лишь для сетей VPN, к которым они подключены напрямую.

Рефлекторы маршрутов могут быть разделены между VPN так, что каждая часть обслуживает лишь подмножество VPN, обслуживаемых SP. Таким образом, ни одному рефлектору не нужно знать маршрутов для всех VPN.

Для межпровайдерских VPN маршрутизаторы ASBR, поддерживающие и распространяющие маршруты VPN-IPv4, могут быть разделены между VPN как рефлекторы и в результате не одному ASBR не потребуется знать маршруты во все межпровайдерские VPN. При использовании multi-hop EBGP маршрутизаторам ASBR не требуется поддерживать и распространять маршруты VPN-IPv4.

В результате в сети SP не требуется наличия одного узла, знающего все маршруты для всех VPN. Поэтому число поддерживаемых в сети VPN не ограничивается возможностями отдельных компонентов.

16. Взаимодействие с IANA

Агентство IANA[1] организовало новый реестр Route Distinguisher Type Field (параграф 4.2). Это поле является двухбайтовым. Типы 0 - 2 определены в этом документе. Дополнительные значения Route Distinguisher Type Field со старшим битом 0 выделяются IANA в порядке очерёдности запросов (First Come, First Served) в соответствии с [IANA]. Значения со старшим битом 1 выделяются IANA по согласованию с IETF (IETF consensus) в соответствии с [IANA].

Этот документ (параграф 4.3.4) задаёт использование BGP Address Family Identifier (AFI) со значением 1 вместе с BGP Subsequent Address Family Identifier (SAFI) со значением 128 для представления семейства адресов VPN-IPv4 с метками, определённого в этом документе.

Использование AFI = 1 для адресов IP в настоящее время задано реестром IANA Address Family Identifier, поэтому от IANA не требуется дополнительных действий.

Значение SAFI = 128 изначально было выделено для частных применений (Private Use) в реестре IANA Subsequent Address Family Identifier. Агентство IANA изменило для SAFI = 128 запись «private use» на «MPLS-labeled VPN address».

17. Благодарности

Полный список участников работы представлен в разделе 18.

Важный вклад в работу внесли также Ravi Chandra, Dan Tappan и Bob Thomas.

Спасибо Shantam Biswas за его рецензию и вклад в работу.

18. Участники работы

Tony Bogovic
Telcordia Technologies
445 South Street, Room 1A264B
Morristown, NJ 07960
EMail: tjb@research.telcordia.com

Stephen John Brannon
Swisscom AG
Postfach 1570
CH-8301
Glattzentrum (Zuerich), Switzerland
EMail: stephen.brannon@swisscom.com

Marco Carugi
Nortel Networks S.A.
Parc d'activites de Magny-Les Jeunes Bois CHATEAUFORT
78928 YVELINES Cedex 9 - FRANCE
EMail: marco.carugi@nortelnetworks.com

Christopher J. Chase
AT&T
200 Laurel Ave
Middletown, NJ 07748
USA
EMail: chase@att.com

Ting Wo Chung
Bell Nexxia
181 Bay Street
Suite 350
Toronto, Ontario
M5J2T3
EMail: ting_wo.chung@bellnexxia.com

Eric Dean

Jeremy De Clercq
Alcatel Network Strategy Group
Francis Wellesplein 1
2018 Antwerp, Belgium
EMail: jeremy.de_clercq@alcatel.be

Luyuan Fang
AT&T
IP Backbone Architecture
200 Laurel Ave.
Middletown, NJ 07748
EMail: luyuanfang@att.com

Paul Hitchen
BT
BT Adastral Park
Martlesham Heath,
Ipswich IP5 3RE
UK
EMail: paul.hitchen@bt.com

Manoj Leelanivas
Juniper Networks, Inc.
385 Ravendale Drive
Mountain View, CA 94043 USA
EMail: manoj@juniper.net

Dave Marshall
Worldcom
901 International Parkway
Richardson, Texas 75081
EMail: dave.marshall@wcom.com

Luca Martini
Cisco Systems, Inc.

[1]Internet Assigned Numbers Authority.

9155 East Nichols Avenue, Suite 400
Englewood, CO, 80112
EMail: lmartini@cisco.com

Monique Jeanne Morrow
Cisco Systems, Inc.
Glatt-com, 2nd floor
CH-8301
Glattzentrum, Switzerland
EMail: mmorrow@cisco.com

Ravichander Vaidyanathan
Telcordia Technologies
445 South Street, Room 1C258B
Morristown, NJ 07960
EMail: vravi@research.telcordia.com

Adrian Smith
BT
BT Adastral Park
Martlesham Heath,
Ipswich IP5 3RE
UK
EMail: adrian.ca.smith@bt.com

Vijay Srinivasan
1200 Bridge Parkway
Redwood City, CA 94065
EMail: vsriniva@cosinecom.com

Alain Vedrenne
Equant
Heraklion, 1041 route des Dolines, BP347
06906 Sophia Antipolis, Cedex, France
EMail: Alain.Vedrenne@equant.com

19. Нормативные документы

[BGP] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006.

[BGP-MP] Bates, T., Rekhter, Y., Chandra, R., and D. Katz, "Multiprotocol Extensions for BGP-4", RFC 2858, June 2000.

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

[MPLS-ARCH] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[MPLS-BGP] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[MPLS-ENCAPS] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

20. Дополнительная литература

[BGP-AS4] Vohra, Q. and E. Chen, "BGP Support for Four-Octet AS Number Space", Work in Progress[1], March 2004.

[BGP-ORF] Chen, E. and Y. Rekhter, "Cooperative Route Filtering Capability for BGP-4", Work in Progress[2], March 2004.

[BGP-RFSH] Chen, E., "Route Refresh Capability for BGP-4", RFC 2918, September 2000.

[BGP-RR] Bates, T., Chandra, R., and E. Chen, "BGP Route Reflection - An Alternative to Full Mesh IBGP", RFC 2796, April 2000.

[IANA] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 2434, October 1998.

[MPLS-ATM] Davie, B., Lawrence, J., McCloghrie, K., Rosen, E., Swallow, G., Rekhter, Y., and P. Doolan, "MPLS using LDP and ATM VC Switching", RFC 3035, January 2001.

[MPLS/BGP-IPsec] Rosen, E., De Clercq, J., Paridaens, O., T'Joens, Y., and C. Sargor, "Architecture for the Use of PE-PE IPsec Tunnels in BGP/MPLS IP VPNs", Work in Progress, March 2004.

[MPLS-FR] Conta, A., Doolan, P., and A. Malis, "Use of Label Switching on Frame Relay Networks Specification", RFC 3034, January 2001.

[MPLS-in-IP-GRE] Worster, T., Rekhter, Y., and E. Rosen, "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", RFC 4023, March 2005.

[MPLS-LDP] Andersson, L., Doolan, P., Feldman, N., Fredette, A., and B. Thomas, "LDP Specification", RFC 3036, January 2001.

[MPLS-RSVP] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

[OSPFv2] Moy, J., "OSPF Version 2", STD 54, RFC 2328, April 1998.

[PASTE] Li, T. and Y. Rekhter, "A Provider Architecture for Differentiated Services and Traffic Engineering (PASTE)", RFC 2430, October 1998.

[RIP] Malkin, G., "RIP Version 2", STD 56, RFC 2453, November 1998.

[OSPF-2547-DNBIT] Rosen, E., Psenak, P., and P. Pillay-Esnault, "Using an LSA Options Bit to Prevent Looping in BGP/MPLS IP VPNs", Work in Progress[3], March 2004.

[TCP-MD5] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385, August 1998.

[1] Работа опубликована в RFC 4893, который заменён RFC 6793. *Прим. перев.*
[2] Работа опубликована в RFC 5291. *Прим. перев.*
[3] Работа опубликована в RFC 4576. *Прим. перев.*

[VPN-MCAST] Rosen, E., Cai, Y., and J. Wijsnands, "Multicast in MPLS/BGP VPNs", Work in Progress[4], May 2004.

[VPN-OSPF] Rosen, E., Psenak, P., and P. Pillay-Esnault, "OSPF as the PE/CE Protocol in BGP/MPLS VPNs", Work in Progress[2], February 2004.

Адреса авторов

Eric C. Rosen

Cisco Systems, Inc.

1414 Massachusetts Avenue

Boxborough, MA 01719

EMail: erosen@cisco.com

Yakov Rekhter

Juniper Networks

1194 N. Mathilda Avenue

Sunnyvale, CA 94089

EMail: yakov@juniper.net

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru

Полное заявление авторских прав

Интеллектуальная собственность

Подтверждение

Финансирование функций RFC Editor обеспечено IETF Administrative Support Activity (IASA).

[4]Работа опубликована в RFC 6513. *Прим. перев.*
[2]Работа опубликована в RFC 4577. *Прим. перев.*