RFC 4277 — Опыт использования протокола BGP-4

Статус документа

Этот документ содержит информацию для сообщества Internet. Документ не задает каких-либо стандартов. Допускается свободное распространения документа.

Тезисы

Целью настоящего документа является рассмотрение вопроса о том, как требования, предъявляемые к публикации проектов стандартов Internet для протоколов маршрутизации, выполнены для протокола BGP-4.
Этот документ удовлетворяет требованиям ко «второму отчету» (the second report), описанным в параграфе 6.0 документа RFC 1264. Для удовлетворения этих требования данный документ дополняет RFC 1773 и описывает дополнительный опыт, приобретенный в период между выпуском проекта стандарта (Draft Standard) и заявлением его в качестве стандарта.

Оглавление

1. Введение

Целью настоящего документа является рассмотрение вопроса о том, как требования, предъявляемые к публикации проектов стандартов Internet для протоколов маршрутизации, выполнены для протокола BGP-4.

Этот документ удовлетворяет требованиям ко «второму отчету» (the second report), описанным в параграфе 6.0 документа RFC 1264. Для удовлетворения этих требования данный документ дополняет RFC 1773 и описывает дополнительный опыт, приобретенный в период между выпуском проекта стандарта (Draft Standard) и заявлением его в качестве стандарта.

2. Обзор BGP-4

BGP представляет собой протокол маршрутизации между автономными системами, рассчитанный на работу в сетях TCP/IP. Основной задачей поддерживающих BGP систем является обмен информацией о доступности сетей с другими системами BGP. Информация о доступности сети включает список автономных систем (AS), через которые проходит эта информация. Набора информации достаточно для построения графа связности AS с учетом их доступности, из которого удаляются маршрутные петли и могут быть приняты некоторые решения на основе правил, действующие на уровне данной AS.
Первая версия протокола BGP была определена в [RFC1105]. Впоследствии были разработаны версии BGP с номерами 2, 3 и 4, определенные в [RFC1163], [RFC1267] и [RFC1771], соответственно. Изменения BGP-4 после получения статуса Draft Standard [RFC1771], перечислены в Приложении A документа [RFC4271].

2.1. Протокол граничного шлюза

Первая версия протокола BGP была опубликована в [RFC1105], вторая — в [RFC1163], третья — в [RFC1267]. Протокол BGP версии 4 определен в [RFC1771] и [RFC4271]. Приложения A, B, C и D документа [RFC4271] содержат список изменений, внесенных в каждую итерацию спецификации протокола BGP.

3. База MIB

База управляющей информации MIB протокола BGP-4 была опубликована в документе [BGP-MIB]. Эта база была обновлена на основе предыдущих версий, опубликованных в [RFC1657] и [RFC1269].
За исключением нескольких системных переменных BGP MIB разбита на две таблицы — BGP Peer и BGP Received Path Attribute.
Таблица Peer содержит информацию о соединениях с партнерами BGP (состояние и текущие действия). Таблица Received Path Attribute содержит все атрибуты, полученные от всех партнеров без применения к ним правил локальной политики маршрутизации. Атрибуты, используемые для реального определения маршрутов, являются подмножеством этой таблицы.

4. Сведения о реализациях

В настоящее время имеется множество независимых интероперабельных реализаций протокола BGP. Хотя предыдущий вариант этого документа содержал обзор интероперабельности реализаций, использующихся в сети Internet, в настоящее время было предложено выделить эту информацию в отдельный документ — BGP Implementation Report [RFC4276].
Следует отметить, что опыт реализации BGP-4 в продукции Cisco описан в документе [RFC1656]. Дополнительные сведения о реализациях протокола можно найти в [RFC4276].

5. Опыт использования протокола

В этой главе рассматривается опыт использования BGP и BGP-4. Протокол BGP использовался в сети с 1989 г., а BGP-4 — с 1993 г. Эксплуатация BGP подразумевает использование всех значимых функций протокола. Современная сетевая среда, где BGP используется в качестве протокола маршрутизации между автономными системами, является гетерогенной. Полоса каналов меняется в диапазоне от 56 кбит/с до 10 Гбит/с. Маршрутизаторы, на которых используется BGP, существенно различаются по производительности — от маломощных процессоров общего назначения до высокопроизводительных специализированных RISC-процессоров, и включают как специализированные устройства, так и обычные станции, на которых используется тот или иной вариант UNIX или другая операционная система.
С точки зрения топологии сеть меняется от очень малозаселенной (малая плотность соединений) до весьма плотной. Требование полносвязности (full-mesh) топологии IBGP было существенно

ослаблено за счет Route Reflection, AS Confederations и комбинаций этих расширений. Функция BGP Route Reflection была изначально определена в [RFC1966] и обновлена в [RFC2796]. Конфедерации AS для BGP были определены в [RFC1965] и потом обновлены в [RFC3065].
На момент создания этого документа BGP-4 используется как протокол маршрутизации между автономными системами во всех AS, подключенных к сети Internet. Общее число активных AS в глобальной таблице маршрутизации Internet составляет около 21000.
BGP используется как для обмена маршрутными данными между транзитными и оконечными (stub) AS, так и для обмена между транзитными AS. Протокол не различает сложившееся исторически разделение сетей на магистральные (backbones), региональные (regional) и краевые (edge).
Полный набор внешних маршрутов, передаваемых с помощью BGP включает более 170000 агрегированных записей, что в несколько раз превышает число подключенных сетей. Число активных путей в магистральных маршрутизаторах некоторых сервис-провайдеров превышает 2,5 миллиона. Естественная длина AS path для некоторых маршрутов составляет 10, но существуют AS, «дополняющие» длину пути до 25 или более.

6. Использование TCP

BGP использует TCP [RFC793] в качестве протокола транспортного уровня. В силу этого присущие TCP характеристики наследуются протоколом BGP.
Например, функции управления полосой не могут быть реализованы, поскольку принятый в TCP алгоритм замедленного старта приводит к разрыву соединения BGP.

7. Метрика

В этой главе обсуждаются различные варианты метрики, используемые в BGP. Протокол BGP имеет различные параметры метрики для IBGP и EBGP. Это позволяет дать метрике, основанной на правилах, более высокий приоритет по сравнению с метрикой, определяемой дистанцией. Благодаря этому каждая автономная система может определить независимую политику как внутри AS, так и для внешних маршрутов. Атрибут BGP MED используется в качестве метрики узлами EBGP (междоменная маршрутизация), а LOCAL_PREF — узлами IBGP (внутридоменная маршрутизация).

7.1. Атрибут MED

В BGP версии 4 старый атрибут метрики INTER-AS был переопределен как MULTI_EXIT_DISC (MED). Это значение может использоваться в процессе отбрасывания лишних маршрутов (tie-breaking), когда выбирается предпочтительный путь к данному пространству адресов, и может обеспечивать узлам BGP возможность выбора оптимальной точки входа в локальную AS со стороны AS партнера.
Хотя атрибут MED может использоваться только при сравнении путей, полученных от разных партнеров из одной AS, многие реализации предоставляют возможность сравнения MED для различных автономных систем.
Несмотря на то, что это может представляться эффективным в некоторых конфигурациях, нужно с осторожностью подходить к сравнению MED из различных автономных систем. Узлы BGP часто определяют значение MED на основе метрики IGP, связанной с достижимостью данного BGP NEXT_HOP в локальной AS. Это позволяет обоснованно отражать в атрибутах MED топологию IGP при анонсировании маршрутов партнерам. Очень удобно использовать MED для сравнения множества путей, полученных из одной смежной AS, но при сравнении MED из различных автономных систем могут приниматься некорректные решения. Типичным случаем является использование автономной системой различных механизмов для установки метрики IGP и значений BGP MED; возможно даже использование разных протоколов IGP с сильно различающимися пространствами метрики.
Другим вопросом, связанным с использованием атрибута MED, является влияние агрегирования маршрутной информации BGP на значения MED. Объединенные маршруты часто генерируются множества точек в AS для обеспечения стабильности, резервирования и т. п. Когда значения MED получаются на основе метрики IGP,связанной упомянутыми агрегированными маршрутами, анонсируемое партнеру значение MED может привести к выбору далекого от оптимума маршрута.
Атрибут MED был специально создан как «слабая» метрика, которая будет использоваться только в конце процесса выбора лучшего маршрута. Рабочая группа BGP была озабочена тем, что любая метрика, заданная удаленным оператором, будет воздействовать на маршрутизацию в локальной

AS лишь в тех случаях, когда не указано иных предпочтений. Основной целью создания MED было обеспечение гарантий того, что партнеры не смогут «потерять» или «поглотить» трафик для сетей, которые они анонсируют.

7.1.1. MED и картошка

Рассмотрим поток трафика между парой адресатов, каждый из которых соединен с двумя транзитными сетями. В этом случае каждая из транзитных сетей может выбирать между передачей трафика партнеру, ближайшему к второму транзитному провайдеру, или партнеру, анонсирующему «более дешевый» путь через другого провайдера. Первый метод называют "hot potato routing", поскольку он напоминает быстрое перебрасывание горячей картофелины, удерживаемой голыми руками. Маршрутизация этого типа осуществляется без передачи полученного от EBGP значения MED в IBGP. Это минимизирует транзитный трафик для провайдера, маршрутизирующего трафик. Значительно менее распространенным методом является "cold potato routing", когда транзитный провайдер использует свою транзитную емкость для получения трафика, направляемого смежному транзитному провайдеру, анонсируемому как ближайший к адресату. Этот тип маршрутизации выполняется путем передачи полученного от EBGP значения MED в IBGP.
Если один из транзитных провайдеров использует метод «hot potato», а другой — «cold potato», трафик между адресатами будет симметричным. В зависимости от конкретных отношений между провайдерами, если один из них имеет большую емкость или существенно менее загруженную транзитную сеть, он может использовать метод «cold potato». Созданная NSF магистральная сеть NSFNET и региональные сети NSF являлись в середине 1990 годов примером повсеместного использования маршрутизации по методу «cold potato».
В некоторых случаях провайдер может использовать метод «hot potato» для некоторых адресатов в данной партнерской AS и метод «cold potato» — для других. Разное отношение к коммерческому и исследовательскому трафику в сети NSFNET середины 1990 годов является примером такого решения. Однако этот вариант можно описать термином «mashed potato routing», отражающим сложность настройки конфигурации маршрутизаторов в то время.
По-видимому более понятными терминами, не относящимися к огородной сфере, будут "best exit routing" вместо «cold potato routing » и "closest exit routing" вместо «hot potato routing».

7.1.2. Передача MED партнерам BGP

[RFC4271] позволяет передавать атрибуты MED, полученные узлом BGP от любого из партнеров EBGP, своим партнерам IBGP. Хотя анонсирование атрибутов MED партнерам IBGP не является требуемым, оно обычно используется по умолчанию. Атрибуты MED, полученные узлом BGP от партнеров EBGP не следует передавать другим партнерам EBGP.
Отметим, что многие реализации поддерживают механизм получения значений MED из метрики IGP, что позволяет отражать в атрибутах BGP MED топологию и метрику IGP внутренней сети при распространении информации в смежные автономные системы.

7.1.3. Нулевое значение MED и отсутствие MED

[RFC4271] требует от реализации поддержки механизма удаления атрибутов MED. Предыдущие реализации не рассматривали отсутствие атрибута MED тождественным MED = 0. Спецификация [RFC4271] требует, чтобы отсутствие атрибута MED трактовалось как нулевое значение.
Отметим, что многие реализации поддерживают механизм явной трактовки отсутствующего атрибута как «плохого признака» или меньшего уровня предпочтения по сравнению с нулевым или положительным значением атрибута.

7.1.4. Атрибуты MED и выбор маршрута с учетом его «возраста»

В некоторых реализациях используются уловки для учета «возраста» при выборе маршрута на основе MED. Т. е., при прочих равных выбирается более старый маршрут без учета значения атрибута MED. Причина такого выбора заключается в том, что «старый» маршрут наверняка более стабилен, поэтому следует предпочесть его. Однако такие уловки приводят к недетерминированому поведению и, вследствие этого, могут быть нежелательны.

8. Локальные предпочтения

Атрибут LOCAL_PREF может добавляться для того, чтобы позволить оператору легко настроить правила, переписывающие стандартный механизм выбора лучшего пути без необходимости настройки политики локальных предпочтений на каждом маршрутизаторе AS.
Одним из недостатков спецификации BGP-4 было предложение использовать принятое по умолчанию значение LOCAL_PREF. Использование по умолчанию нулевого или максимального значения имеет свои ограничения, поэтому единое значение, используемое по умолчанию во всех маршрутизаторах AS будет упрощать использование маршрутизаторов разных фирм в одной AS.
Атрибут LOCAL_PREF управляется, следовательно проблем за пределами границы AS не возникает.
[RFC4271] требует, чтобы значение LOCAL_PREF передавалось партнерам IBGP и не передавалось партнерам EBGP. Хотя используемого по умолчанию значения LOCAL_PREF спецификация не определяет, чаще всего используется значение 100.
Другой областью, требующей исследований, является метод посредством которого исходная (originating) AS может оказывать влияние на процесс выбора наилучшего пути. Например, сайт с двумя подключениями может выбрать одну AS как основного транзитного провайдера, а вторую использовать как резервную.

```
                    /---- транзит B ----\
конечный пользователь                    транзит A----
                    /---- транзит C ----\
```

В топологии, где два транзитных провайдера соединены с третьим, реальное решение будет приниматься третьим провайдером. Не существует механизма индикации своих предпочтений этому провайдеру.
Возможным вариантом решения этой задачи будет передача дополнительного вектора, соответствующего AS_PATH, где каждая транзитная AS может указывать уровень предпочтения для данного маршрута. Взаимодействующие автономные системы в таком случае смогут выбирать трафик на основе сравнения интересующих частей этого вектора в соответствии со своей политикой маршрутизации.
Хотя защита политики маршрутизации данной AS является основной заботой, избавление от необходимости ручной настройки конфигурации правил маршрутизации потребует более тщательной проверки в будущих протоколах, подобных BGP.

9. Internal BGP в больших автономных системах

Хотя этот вопрос и не связан напрямую с протоколом, он возникает у операторов, которым требуется поддерживать автономные системы с большим числом партнеров. Каждый узел, имеющий партнерство с внешним маршрутизатором, отвечает за распространение информации о доступности и путях всем остальным транзитным и граничным маршрутизаторам данной AS. Это обычно осуществляется путем организации внутренних соединений BGP со всеми транзитными и граничными маршрутизаторами локальной AS.
Отметим, что число партнеров BGP, для которых нужно обеспечить полносвязность, зависит от множества факторов, включая число префиксов в системе маршрутизации, число уникальных путей, стабильность системы и (возможно более важно) эффективность реализации. В результате, хотя и сложно определить, что такое «большое число партнеров», всегда существуют некоторые практические границы этого числа.
В больших AS это ведет к полносвязности соединений TCP (n * (n-1)) и использованию того или иного метода настройки и поддержки этих соединений. Протокол BGP не задает способов распространения такой информации. Следовательно, предпринимаются различные альтернативные попытки (такие, как вставка маршрутной информации BGP в локальный протокол IGP), но многие из них мало применимы на практике.
Для смягчения необходимости организации полносвязного IBGP были разработаны несколько вариантов, включая BGP Route Reflection [RFC2796] и AS Confederations for BGP [RFC3065].

10. Динамика Internet

Как обсуждается в [RFC4274], рост нагрузки на CPU и использования полосы каналов обусловлены динамической природой маршрутизации в Internet. По мере расширения сети Internet возрастает частота изменения маршрутов.
Мы автоматически получаем некоторый уровень подавления этого роста, когда более специфичные NLRI агрегируются в более крупные блоки, однако этого недостаточно. В

Приложении F к документу [RFC4271] приводится описание методов подавления, которые могут применяться к передаче анонсов. В будущих спецификациях протоколов типа BGP методы такого подавления нужно рассматривать как обязательные для реализации.
Механизм BGP Route Flap Damping определен в документе [RFC2439]. BGP Route Flap Damping помогает снизить объем маршрутной информации, передаваемой между партнерами BGP, что ведет к снижению нагрузки на эти узлы без негативного влияния на время схождения для относительно стабильных маршрутов.
Ни одна из современных реализаций BGP Route Flap Damping не сохраняет маршрутную историю в уникальных NRLI или AS Path, хотя RFC 2439 требует это. Потенциальным результатом отказа от раздельного рассмотрения каждого AS Path является чересчур агрессивное подавление адресатов в сети с высокой плотностью соединений (densely meshed network). Наиболее важным следствием этого является подавление адресата после единственного отказа. Поскольку автономные системы верхних уровней в Internet имеют высокий уровень связности, описанные здесь негативные последствия уже наблюдаются.
Изменения маршрутов анонсируются с помощью сообщений UPDATE. Наибольший трафик, связанный с сообщениями UPDATE, возникает при неэффективной упаковке анонсов с сообщениями об изменениях маршрутов. Анонсирование маршрутных изменений с общими атрибутами в одном сообщении BGP UPDATE помогает снизить расход полосы и загрузку процессора при обработке анонсов, как описано в главе 12 «Упаковка сообщений UPDATE».
Постоянные ошибки BGP могут заставить партнеров BGP постоянно переключать (flap) маршруты, если не реализованы средства подавления. Такое переключение ведет к значительному росту нагрузки на CPU маршрутизатора. Разработчики могут счесть полезной для предотвращения таких ситуаций реализацию функций подавления [RFC4271].

11. Базы маршрутной информации BGP (RIB)

В [RFC4271] сказано: "Вопросы локальной политики, которая может приводить к включению маршрутов в базу Adj-RIB-Out без их добавления в таблицу пересылки локального узла BGP выходят за пределы данного документа."
Однако несколько широко распространенных реализаций не подтверждают, что записи Loc-RIB используются для заполнения таблицы пересылки до их установки в базу Adj-RIB-Out. Чаще всего это наблюдается в тех случаях, когда данный префикс представлен несколькими протоколами и уровень предпочтения для маршрута, полученного от BGP, ниже, чем для маршрута,полученного от другого протокола. В результате маршрут, полученный от другого протокола, включается в таблицу пересылки.
Для реализации может оказаться желательным обеспечение «кнопки», позволяющей анонсировать «неактивные» маршруты BGP.
Может также оказаться желательным для реализации обеспечение механизма, позволяющего узлу BGP анонсировать маршруты BGP, которые не были выбраны в процессе принятия решения.

12. Упаковка сообщений UPDATE

Множество недоступных маршрутов можно анонсировать в одном сообщении BGP UPDATE. Кроме того, в одном сообщении UPDATE может анонсироваться один или несколько доступных маршрутов, если префиксы этих маршрутов разделяют общий набор атрибутов пути.
Протокол BGP4 позволяет анонсировать множество префиксов с общим набором атрибутов пути в одном сообщении UPDATE.
Обычно такое анонсирование называют упаковкой обновлений (update packing). По возможности рекомендуется использовать такую упаковку, поскольку она обеспечивает механизм для более эффективного поведения в нескольких областях, включая:

- снижение нагрузки на систему, связанной с генерацией и приемом сообщений UPDATE;
- снижение уровня сетевого трафика за счет уменьшения числа пакетов и расхода полосы;
- снижение частоты обработки атрибутов пути и поиска соответствий в базе данных AS_PATH (если таковая имеется); упорядочивание атрибутов пути упрощает поиск соответствий в базе данных, поскольку в этом случае не может существовать нескольких представлений одного набора данных.

Протокол BGP предлагает помещать отзываемые маршруты в начале сообщения UPDATE, а за ними включать информацию о доступных маршрутах. Это позволяет избавиться от ненужных переключений маршрутов (route flapping) в BGP.

13. Ограничение частоты обновлений

Протокол BGP определяет различные механизмы ограничения частоты передачи сообщений UPDATE. Параметр MinRouteAdvertisementInterval задает минимальное время, которое должно пройти между двумя последовательными анонсами одному получателю от одного узла BGP. Это значение устанавливается независимо для каждого партнера BGP.
BGP использует транспортный протокол TCP, а последний может временно блокировать передачу данных при пустом окне. В результате множество обновлений может передаваться с меньшим интервалом, нежели они были помещены в очередь на передачу. Реализациям следует избегать таких ситуаций.

13.1. Учет характеристик TCP

Если получатель TCP обрабатывает данные более медленно, чем отправитель, или скорость соединения TCP является ограничивающим фактором, наблюдается обратное воздействие (backpressure) на передающее приложение TCP. Когда буфер TCP заполняется, передающее приложение будет блокировать запись или получать сообщение об ошибке при попытке записи. В старых или примитивных современных реализациях зачастую присутствует ошибочная установка опции для блокирования записи или опций запрета такой блокировки. Такие реализации трактуют ошибки, связанные с заполнением буфера, как критические.
Для реализаций, понимающих, что буфер записи может заполняться, существует иная западня. Приложению не следует пытаться сохранить поток TCP внутри самого приложения. Если принимающая сторона или соединение TCP постоянно не обеспечивают нужной скорости, буфер может возрасти до таких размеров, что поглотит всю доступную память. От реализации BGP требуется передача изменений всем партнерам, для которых соединения TCP не заблокированы, и передавать изменения остальным партнерам после того, как соответствующие соединения будут разблокированы.
Если NLRI многократно изменяется в течение периода блокировки записи для одного или нескольких партнеров, передавать следует только последний лучший маршрут. В этом случае BGP работает в соответствии с рекомендациями [RFC4274]. В случаях экстремально частых изменений маршрутов партнерам, способным обрабатывать информацию с высокой скоростью, передается больший объем маршрутных данных, нежели более медленным партнерам.
Для реализаций, способных дифференцировать партнеров по скорости восприятия маршрутных изменений, в тех случаях, когда основные изменения маршрутов вносятся подмножеством нестабильных NLRI, единственным влиянием изолированного изменения маршрутов на относительно стабильные маршруты является замедление схождения, время которого остается ограниченным, независимо от уровня нестабильности.

14. Упорядочивание атрибутов пути

Спецификация BGP рекомендует узлам BGP, передающим множество префиксов в сообщении UPDATE, сортировать и упорядочивать атрибуты пути в соответствии с кодами типа. Это будет помогать их партнерам быстрее идентифицировать семантически отличающиеся наборы атрибутов из разных сообщений.
Разработчики могут счесть полезной упорядочивание атрибутов пути по кодам типа, поскольку такие наборы атрибутов с идентичной семантикой можно идентифицировать быстрее.

15. Сортировка AS_SET

Наборы AS_SET обычно используются при агрегировании маршрутов BGP. Использование таких атрибутов снижает объем данных AS_PATH за счет однократного перечисления номеров AS независимо от количества экземпляров данного номера в объединяемых маршрутах. Наборы AS_SET обычно сортируются в порядке возрастания для обеспечения эффективного контроля за включенными в набор номерами AS. Такая оптимизация является необязательной.

16. Контроль согласования версий

Поскольку агрегирование маршрутов в предварительных вариантах BGP-4 не поддерживается более ранними версиями BGP, реализациям, поддерживающим другие варианты агрегирования (кроме BGP-4), следует обеспечивать поддержку версий независимо для каждого партнера.

Предполагается, что на момент подготовки этого документа все узлы BGP в сети Internet используют протокол BGP версии 4.

17. Вопросы безопасности

BGP обеспечивает гибкий и расширяемый механизм аутентификации и защиты. Этот механизм позволяет поддерживать схемы различной сложности. Сеансы BGP аутентифицируются по IP-адресам партнеров. В дополнение к этому все сессии BGP аутентифицируются по номерам автономных систем, анонсируемых партнерами.
Поскольку BGP работает на основе протоколов TCP и IP, схемы аутентификации BGP могут быть расширены путем добавления любых механизмов аутентификации и защиты, поддерживаемых протоколами TCP и IP.

17.1. Опция TCP MD5

[RFC2385] определяет способ использования сигнатур TCP MD5 для проверки информации, передаваемой между двумя партнерами. Этот метод позволяет предотвратить вставку третьими сторонами информации (например, TCP Reset) в поток данных или изменение маршрутной информации, передаваемой между двумя узлами BGP.
В настоящий момент сигнатуры TCP MD5 не используются достаточно широко (особенно в системах междоменной маршрутизации) главным образом в связи с проблемой распространения ключей. Многие механизмы распространения ключей представляются слишком «тяжелыми» для такой задачи.
Многие наивно предполагают, что атакующему нужно точно угадать порядковый номер TCP (вместе с адресами и номерами портов отправителя и получателя) для вставки сегмента данных или сброса транспортного соединения TCP между парой узлов BGP. Однако недавние наблюдения и обсуждения показывают, что злонамеренные данные достаточно «втолкнуть» в окно приема TCP, которое может быть достаточно большим, что существенно снижает сложность организации таких атак.
В связи с этим рекомендуется использовать опцию MD5 TCP для защиты BGP от сброса соединений и вставки данных.

17.2. Использование BGP с IPsec

BGP может работать на основе Ipsec как в туннельном, так и в транспортном режиме. В этом случае TCP-часть пакетов IP шифруется. Это не только предотвращает возможность вставки информации в поток данных между двумя узлами BGP, но и не позволяет атакующему изучать данные, передаваемые между партнерами.
IPsec предлагает несколько опций для обмена сеансовыми ключами, которые могут быть полезны в системах междоменной маршрутизации. Эти опции были исследованы на множестве реальных систем, но не было найдено определенного решения для распространения ключей при работе BGP на основе IPsec.
Поскольку BGP работает на базе протоколов TCP и IP, следует отметить, что протокол BGP уязвим к тем же атакам на службы и систему аутентификации, которым подвержены все протоколы, работающие на основе TCP.

17.3. Разное

Другой проблемой протоколов маршрутизации является подтверждение корректности и полномочности маршрутной информации, передаваемой в системе маршрутизации. В настоящее время на решении этой задачи сосредоточено много усилий, включая работы по определению угроз маршрутным данным BGP [BGPATTACK] и разработке мер проверки корректности и полномочности маршрутных данных, передаваемых BGP [SBGP] [soBGP].
В дополнение к этому IETF создана рабочая группа RPSEC для обсуждения и оказания помощи в сфере защиты протоколов маршрутизации. В рамках RPSEC данный документ является откликом на BGP4, направленным на дальнейшее совершенствование протокола.

18. Рабочие группы PTOMAINE и GROW

Рабочая группа GROW, недавно созданная взамен группы PTOMAINE, занимается вопросами роста размеров таблиц маршрутизации, взаимодействия протоколов внутренней и внешней маршрутизации, а политики и практики распределения адресов в контексте влияния на глобальную систему маршрутизации. Кроме того, GROW будет также документировать эксплуатационные проблемы измерения, политики, безопасности и VPN.
GROW в настоящее время изучает влияние агрегирования маршрутов и вопрос невозможности агрегирования в результате неадекватной координации провайдеров.
В рамках GROW данный документ является откликом на BGP4, направленным на дальнейшее совершенствование протокола.

19. Реестры маршрутизации Internet (IRR)

Многие организации регистрируют свою политику маршрутизации и происхождение префиксов в различных распределенных базах данных Реестра маршрутизации Internet (IRR). Эти базы данных обеспечивают доступ к информации, хранящейся в формате языка RPSL, определенного в [RFC2622]. Хотя регистрационная информация для некоторых провайдеров может поддерживаться и корректироваться, недостаточная корректность и своевременность (актуальность) данных в различных базах данных IRR препятствует широкому использованию этого ресурса.

20. Региональные реестры (RIR) и IRR, немного истории

Программа NSFNET использовала протокол EGP (а впоследствии BGP) для обеспечения внешней маршрутной информации. Политика NSF заключалась в дифференцировании цен и типов услуг для исследовательских или учебных (RE) и коммерческих (CO) сетей, что привело к созданию изначальных требований к политике BGP. Кроме большей платы сети CO не могли использовать магистраль NSFNET для доступа в другие сети CO. Причиной более высокой платы для коммерческих пользователей NSFNET было желание субсидировать сети RE. Признание того, что сеть Internet изменилась от иерархической к многосвязной одноуровневой привело к отказу от EGP и BGP-1, а также представлений об иерархии сетей.
Реализация политики NSF обеспечивалась за счет поддержки базы данных NSF PRDB. База PRDB не только содержала классификацию каждой сети как CO или RE, но также включала список предпочтительных точек входа в NSFNET для доступа в каждую сеть. Это послужило основой для современной системы оценки уровня предпочтений BGP LOCAL_PREF. Средства, обеспечиваемые PRDB позволяли полностью сгенерировать конфигурационные параметры маршрутизатора для NSFNET.
Использование PRDB обеспечивало существенное повышение надежности NSFNET по сравнению с одноранговыми системами того времени. PRDB обеспечивала более оптимальную маршрутизацию для тех сетей, которые были достаточно открыты для изучения и предоставляли достоверные сведения о себе.
После того, как NSFNET прекратила в 1995 году предоставление магистральных услуг, было принято решение о том, что PRDB следует отделить от конкретного провайдера, а унаследованное содержимое базы данных и ее последующее обновление сделать достоянием всех провайдеров, желающих пользоваться этой базой данных. Европейское сетевое сообщество в течение долгого времени считало, что PRDB слишком концентрируется на США. На базе RIPE был создан открытый формат RIPE-181 и поддерживается открытая база данных о регистрации адресов и номеров AS, а также политики маршрутизации. База PRDB была преобразована в формат RIPE-181 и были также переработаны инструменты для работы с этим форматом. Набор баз данных стал называться Реестром маршрутизации Internet (IRR), начальными компонентами которого были база данных RIPE и созданный с помощью NSF Routing Arbitrator (RA).
Через некоторое время стала ясна необходимость расширения RIPE-181 и было получено согласие RIPE на такое расширение, которое было создано рабочей группой IETF RPS, как язык RPSL.
Другим результатом работы RPS явилась модель аутентификации и способ широкого распространения базы данных с сохранением контроля и синхронизацией множества хранилищ.
Были разработаны свободно распространяемые инструменты, созданные прежде всего RIPE, Merit, и ISI (наиболее полным является набор ISI). Усилия участников IRR были несколько затруднены провайдерами, не желающими предоставлять в IRR актуальную информацию. Наиболее крупные из таких провайдеров заявляли устно, что поддержка записей в базе данных создает дополнительную нагрузку для администраторов, содержащаяся в базе информация может давать преимущества их конкурентам, использующим IRR. Результатом этого стала фактическая

бесполезность IRR и повышение уровня уязвимости Internet к атакам, направленным на систему маршрутизации, и эпизодическим вставкам ложной маршрутной информации.
Известно множество случаев, когда случайного искажения маршрутизации Internet удавалось избежать провайдерам, использующим IRR, но не удавалось избавиться от проблем тем провайдерам, которые не использовали эту базу. Были разработаны фильтры для уменьшения области покрытия, вызванного разрушением целостности IRR; такие повреждения время от времени продолжают происходить и масштабы их влияния возрастают.

21. Благодарности

Мы благодарим Paul Traina и Yakov Rekhter за разработку предыдущей версии этого документа и создание хорошей основы для данного обновления. Благодарим также Curtis Villamizar за написание текста и просмотр документа. Спасибо Russ White, Jeffrey Haas, Sean Mentzer, Mitchell Erblich и Jude Ballard за их проницательный взгляд на документ.
В заключение мы хотим поблагодарить участников рабочей группы IDR за их вклад в подготовку этого документа.

22. Литература

22.1. Нормативные документы

1. [RFC1966] Bates, T. and R. Chandra, "BGP Route Reflection An alternative to full mesh IBGP", RFC 1966, June 1996.
2. [RFC2385] Heffernan, A., "Protection of BGP Sessions via the TCP MD5 Signature Option", RFC 2385, August 1998.
3. [RFC2439] Villamizar, C., Chandra, R., and R. Govindan, "BGP Route Flap Damping", RFC 2439, November 1998.
4. [RFC2796] Bates, T., Chandra, R., and E. Chen, "BGP Route Reflection — An Alternative to Full Mesh IBGP", RFC 2796, April 2000.
5. [RFC3065] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 3065, February 2001.
6. [RFC4274] Meyer, D. and K. Patel, "BGP-4 Protocol Analysis", RFC 4274, January 2006.
7. [RFC4276] Hares, S. and A. Retana, "BGP 4 Implementation Report", RFC 4276, January 2006.
8. [RFC4271] Rekhter, Y., Li, T., and S. Hares, Eds., "A Border Gateway Protocol 4 (BGP-4)", RFC 4271, January 2006
9. [RFC1657] Willis, S., Burruss, J., Chu, J., "Definitions of Managed Objects for the Fourth Version of the Border Gateway Protocol (BGP-4) using SMIv2", RFC 1657, July 1994.
10. [RFC793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, September 1981

22.2. Дополнительная литература

11. [RFC1105] Lougheed, K. and Y. Rekhter, "Border Gateway Protocol (BGP)", RFC 1105, June 1989.
12. [RFC1163] Lougheed, K. and Y. Rekhter, "Border Gateway Protocol (BGP)", RFC 1163, June 1990.
13. [RFC1264] Hinden, R., "Internet Engineering Task Force Internet Routing Protocol Standardization Criteria", RFC 1264, October 1991.
14. [RFC1267] Lougheed, K. and Y. Rekhter, "Border Gateway Protocol 3 (BGP-3)", RFC 1267, October 1991.
15. [RFC1269] Willis, S. and J. Burruss, "Definitions of Managed Objects for the Border Gateway Protocol: Version 3", RFC 1269, October 1991.
16. [RFC1656] Traina, P., "BGP-4 Protocol Document Roadmap and Implementation Experience", RFC 1656, July 1994.
17. [RFC1771] Rekhter, Y. and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, March 1995.
18. [RFC1773] Traina, P., "Experience with the BGP-4 protocol", RFC 1773, March 1995.
19. [RFC1965] Traina, P., "Autonomous System Confederations for BGP", RFC 1965, June 1996.
20. [RFC2622] Alaettinoglu, C., Villamizar, C., Gerich, E., Kessens, D., Meyer, D., Bates, T., Karrenberg, D., and M. Terpstra, "Routing Policy Specification Language (RPSL)", RFC 2622, June 1999.

21. [BGPATTACK] Convery, C., "An Attack Tree for the Border Gateway Protocol", Work in Progress.
22. [SBGP] "Secure BGP", Work in Progress.
23. [soBGP] "Secure Origin BGP", Work in Progress.

Адреса авторов

Danny McPherson
Arbor Networks
EMail: ten.robra@ynnad
Keyur Patel
Cisco Systems
EMail: moc.ocsic@etapuyek

Перевод на русский язык

Николай Малых, moc.milib@hkylamn
(source)