

BGP Route Reflection – альтернатива полносвязности IBGP

BGP Route Reflection

An alternative to full mesh IBGP

Статус документа

Этот документ определяет экспериментальный протокол (Experimental Protocol) для сообщества Internet. Документ не задаёт каких-либо стандартов Internet и служит приглашением к дискуссии в целях дальнейшего совершенствования. Допускается свободное распространение документа.

Аннотация

BGP¹ [1] представляет собой протокол междоменной маршрутизации, разработанный для сетей TCP/IP. В настоящее время в сети Internet протокол BGP настроен так, что все узлы BGP в одной AS должны образовывать полносвязный набор соединений (fully meshed) и любая внешняя маршрутная информация должна передаваться всем остальным маршрутизаторам внутри данной AS. Это порождает серьезные проблемы масштабирования, которые подробно описаны вместе с альтернативными предложениями в документах [2, 3].

В данном документе описан метод «отражения маршрутов» (Route Reflection) и его использование, ослабляющее требование полносвязности для IBGP.

1. Введение

В настоящее время в сети Internet протокол BGP настроен так, что все узлы BGP в одной AS должны образовывать полносвязный набор соединений и любая внешняя маршрутная информация должна передаваться всем остальным маршрутизаторам внутри данной AS. Очевидно, что требование полносвязности становится невыполнимым в системах, где большое число узлов IBGP обменивается значительными объемами маршрутной информации (такая ситуация наблюдается в большинстве современных сетей).

Для n узлов BGP в данной AS требуется организовать $n*(n-1)/2$ уникальных сессий IBGP. При конечных размерах полосы каналов и производительности процессоров в маршрутизаторах это требование очевидно невыполнимо.

Эта проблема масштабирования и многочисленные предложения по снижению ее остроты подробно описаны в документах [2, 3]. Данный документ представляет еще один вариант избавления от полносвязности, известный как Route Reflection. Этот метод позволяет узлу BGP (называемому Route Reflector) анонсировать полученные от IBGP маршруты некоторым партнерам IBGP. Он изменяет общепринятую концепцию работы и добавляет два новых необязательных непереходных² атрибута BGP для предотвращения петель при обновлении маршрутов.

2. Базовые требования

Метод Route Reflection удовлетворяет перечисленным ниже критериям.

- Простота

Любое дополнение должно быть понятным и простым в настройке.

- Простота перехода

Должна обеспечиваться возможность перехода от полносвязной конфигурации без необходимости изменения топологии или AS. Метод, предложенный в [3], вносит слишком высокие издержки в части управления.

- Совместимость

Должна обеспечиваться возможность сохранения не поддерживающих данный метод узлов IBGP, как части исходной AS или домена, без потери какой-либо маршрутной информации BGP.

Эти критерии основаны на опыте использования метода в очень больших сетях со сложной топологией и множеством внешних соединений.

3. Отражение маршрутов

Основная идея метода отражения очень проста. Рассмотрим пример, показанный на рисунке 1.

В автономной системе ASX имеется три узла IBGP (маршрутизаторы RTR-A, RTR-B, RTR-C). В рамках существующей модели BGP, если RTR-A получает внешний маршрут и выбирает этот маршрут в качестве лучшего, он должен

¹Border Gateway Protocol – протокол граничного шлюза.

²В оригинале ошибочно сказано про 2 переходных атрибута, что не соответствует определениями главы 7. Прим. перев.

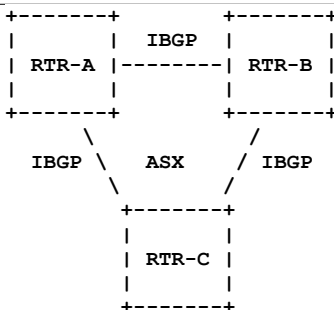


Рисунок 1. Полносвязная система IBGP.

анонсировать этот внешний маршрут обоим узлам RTR-B и RTR-C. Узлы RTR-B и RTR-C (как узлы IBGP) не будут заново анонсировать этот полученный от IBGP маршрут другим партнерам IBGP.

Если это правило ослабить и позволить узлу RTR-C анонсировать полученные от IBGP маршруты другим партнерам IBGP, тогда он будет реанонсировать (или отражать) маршруты IBGP, полученные от RTR-A, узлу RTR-B и наоборот. Это позволит отказаться от организации сессии IBGP между узлами RTR-A и RTR-B, как показано на рисунке 2.

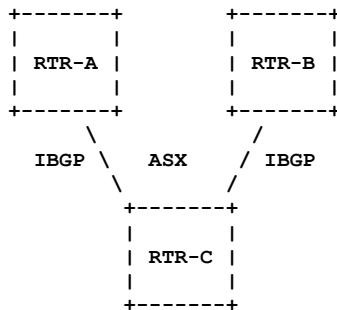


Рисунок 2. IBGP с отражением маршрутов.

Схема метода Route Reflection основана именно на этом принципе.

4. Терминология и концепции

Мы используем термин «рефлектор маршрутов» (Route Reflector или RR) для обозначения узла BGP, принимающего участие в отражении. Внутренние партнеры узла RR делятся на две группы:

- 1) Партнеры-клиенты.
- 2) Партнеры, не являющиеся клиентами.

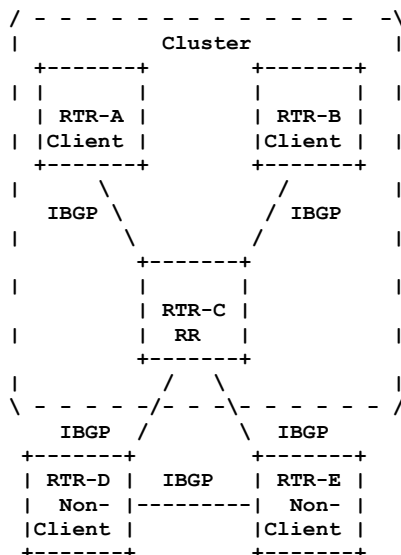


Рисунок 3. Компоненты RR.

Узел RR отражает маршруты между этими группами. Рефлектор RR вместе со своими клиентами образует кластер. Партнеры, не являющиеся клиентами, должны сохранять полносвязность, но для клиентов это требование снимается. Клиентам не следует организовывать партнерских отношений с внутренними узлами, не входящими в их кластер. На рисунке 3 показан пример сети с базовыми компонентами RR, иллюстрирующий терминологию.

5. Работа метода

Когда RR получает маршрут от партнера IBGP, он выбирает лучший путь на основе своих критериев. После выбора лучшего пути узел должен выполнить перечисленные ниже операции в зависимости от типа партнера, передавшего информацию о лучшем пути:

- 1) Маршрут получен от партнера, не являющегося клиентом.

Отразить маршрут всем клиентам.

2) Маршрут получен от клиента.

Отразить маршрут всем партнерам, не являющимся клиентами, а также всем партнерам-клиентам, кроме источника маршрута (поскольку клиенты могут не быть полносвязными).

3) Маршрут получен от партнера EBGP.

Передать всем партнерам, независимо от того, являются ли они клиентами.

Автономная система может включать множество RR. Узел RR трактует остальные рефлекторы RR, как обычные внутренние узлы BGP. Рефлектор RR может быть настроен на присутствие других RR как в числе клиентов, так и среди партнеров, не являющихся клиентами.

В простой конфигурации опорная сеть может быть поделена на множество кластеров. Каждый рефлектор RR настраивается на то, что другие RR не относятся к группе клиентов (таким образом, все RR будут образовывать полносвязную систему). Клиенты будут настраиваться на поддержку сессий IBGP только с RR в своем кластере. Благодаря отражению маршрутов, все узлы IBGP будут получать отраженную маршрутную информацию.

В автономной системе могут присутствовать узлы BGP, не понимающие концепцию отражения маршрутов (будем называть их обычными узлами BGP). Схема отражения маршрутов допускает сосуществование с обычными узлами BGP. Эти узлы могут относиться к группе клиентов или не являться клиентами RR. Это обеспечивает возможность простого и постепенного перехода от существующей модели работы IBGP к модели с отражением маршрутов. Можно начать с создания кластера путем настройки одного маршрутизатора в качестве означенного RR и настройки остальных RR и их клиентов как нормальных партнеров IBGP. Постепенно могут создаваться дополнительные кластеры.

6. Резервирование RR

Обычно кластер клиентов будет включать один рефлектор RR. В этом случае кластер будет идентифицироваться значением ROUTER_ID рефлектора RR. Однако такой вариант может не обеспечивать достаточной надежности и для резервирования в одном кластере может создаваться множество RR. Все рефлекторы RR одного кластера могут быть настроены на использование общего 4-байтового идентификатора CLUSTER_ID, который позволяет любому рефлектору RR отбрасывать маршруты, получаемые от других RR того же кластера.

7. Предотвращение петель

При использовании отражения маршрутов возможно возникновение петель при реанонсировании в результате некорректной конфигурации. Метод Route Reflection определяет два новых атрибута для детектирования и предотвращения таких петель.

ORIGINATOR_ID

ORIGINATOR_ID – необязательный, непереключаемый атрибут BGP с кодом типа 9. Этот атрибут имеет размер 4 байта и создается рефлектором RR в отраженном маршруте. Атрибут будет включать значение ROUTER_ID источника маршрута в локальной AS. Узлу BGP не следует создавать атрибут ORIGINATOR_ID, если последний уже присутствует. Маршрутизатору, распознающему атрибут ORIGINATOR_ID, следует игнорировать маршрут, содержащие значение его ROUTER_ID в качестве ORIGINATOR_ID.

CLUSTER_LIST

CLUSTER_LIST – необязательный, непереключаемый атрибут BGP с кодом типа 10. Этот атрибут представляет собой список значений CLUSTER_ID, представляющий путь отражения, по которому передавался маршрут. Формат атрибута показан ниже.

```

      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Attr. Flags |Attr. Type Code| Length | value ...
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Поле Length¹ указывает число октетов.

Когда рефлектор RR отражает маршрут, он должен добавить локальное значение CLUSTER_ID в конец списка CLUSTER_LIST. Если список CLUSTER_LIST отсутствует, узел должен создать новый список. Используя этот атрибут, RR может определять возникновение петель при передаче маршрутной информации в результате конфигурационных ошибок. Если локальное значение CLUSTER_ID присутствует в списке кластеров, полученный анонс следует игнорировать.

8. Вопросы реализации метода

Следует принять меры по предотвращению изменения описанных выше атрибутов пути (средствами конфигурации) в процессе обмена маршрутной информацией между RR и клиентами или партнерами, не являющимися клиентами. Такое изменение атрибутов может приводить к возникновению маршрутных петель.

В некоторых реализациях возможно изменение атрибута пути BGP NEXT_HOP. Например, рефлектору RR может потребоваться изменение NEXT_HOP для полученных от EBGP маршрутов при их передаче внутренним партнерам. Однако для рефлектора RR это должно быть невозможно на отражаемых маршрутах IBGP, поскольку такое изменение будет приводить к нарушению основного принципа отражения и может приводить к возникновению «черных дыр».

RR не следует изменять какие-либо атрибуты AS-PATH, которые могут влиять на согласованность выбора маршрута (т. е., LOCAL_PREF, MED, DPA). Такое изменение может приводить к возникновению маршрутных петель.

Протокол BGP не обеспечивает клиентам способа динамической идентификации себя в качестве клиентов RR. Простейшим способом такой идентификации является настройка конфигурации вручную.

¹Поле Length ошибочно указано на рисунке, как 1-октетное. Размер этого поля может составлять 1 или 2 октета в зависимости от значения флага Extended Length (см. параграф 4.3 [RFC 4271](#)). *Прим. перев.*

9. Вопросы безопасности

Вопросы безопасности не обсуждаются в этом документе.

10. Благодарности

Авторы благодарят Dennis Ferguson, John Scudder, Paul Traina и Tony Li за дискуссии, которые привели к созданию этого документа. Идея метода основана на давней дискуссии между Tony Li и Dimitri Haskin.

11. Литература

[1] Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", [RFC 1771](#), March 1995.

[2] Haskin, D., "A BGP/IDRP Route Server alternative to a full mesh routing", RFC 1863, October 1995.

[3] Traina, P., "Autonomous System Confederations for BGP"¹, [RFC 1965](#), June 1996.

12. Адреса авторов

Tony Bates

cisco Systems

170 West Tasman Drive

San Jose, CA 95134

Phone: +1 408 527 2470

E-Mail: tbates@cisco.com

Ravishanker Chandrasekeran

(Ravi Chandra)

cisco Systems

170 West Tasman Drive

San Jose, CA 95134

E-Mail: rchandra@cisco.com

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru

¹В оригинале этот документ ошибочно назван "Limited Autonomous System Confederations for BGP". Прим. перев.