Network Working Group
Request for Comments: 4456
Obsoletes: 2796, 1966
Category: Standards Track

T. Bates
E. Chen
Cisco Systems
R. Chandra
Sonoa Systems
April 2006

BGP Route Reflection - альтернатива полносвязности IBGP

BGP Route Reflection:

An Alternative to Full Mesh Internal BGP (IBGP)

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

BGP[1] [1] представляет собой протокол междоменной маршрутизации, разработанный для сетей TCP/IP. В типовой ситуации все узлы BGP в одной AS должны образовывать полносвязный набор соединений (fully meshed) потому, что 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 данном документе описан метод «отражения маршрутов» (Route Reflection) и его использование, ослабляющее требование полносвязности для IBGP.

Этот документ отменяет действие RFC 2796 и RFC 1966.

Оглавление

1. Введение

В типовой ситуации все узлы BGP в одной AS должны образовывать полносвязный набор соединений потому, что любая внешняя маршрутная информация должна передаваться всем остальным маршрутизаторам внутри данной AS. Для n узлов BGP в данной AS требуется организовать n*(n-1)/2 уникальных сессий IBGP. 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).

Эта проблема масштабирования и многочисленные предложения по снижению ее остроты подробно описаны в документах [2,3]. Данный документ представляет еще один вариант избавления от проблемы полносвязности, известный как Route Reflection. Этот метод позволяет узлу BGP (называемому Route Reflector) анонсировать полученные от IBGP маршруты некоторым партнерам IBGP. Предложенный метод изменяет общепринятую концепцию

[1]Border Gateway Protocol - протокол граничного шлюза.

работы протокола и добавляет два новых необязательных непереходных[2] атрибута BGP для предотвращения петель при обновлении маршрутов.

Этот документ отменяет действие 2796 и RFC 1966.

2. Уровни требований

Ключевые слова **необходимо** (MUST), **недопустимо** (MUST NOT), **требуется** (REQUIRED), **нужно** (SHALL), **не следует** (SHALL NOT), **следует** (SHOULD), **не нужно** (SHOULD NOT), **рекомендуется** (RECOMMENDED), **возможно** (MAY), **необязательно** (OPTIONAL) в данном документе интерпретируются в соответствии с RFC 2119 [7].

3. 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], порождает 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.

4. Отражение маршрутов

Основная идея метода отражения (Route Reflection) очень проста. Рассмотрим пример, показанный на рисунке 1.

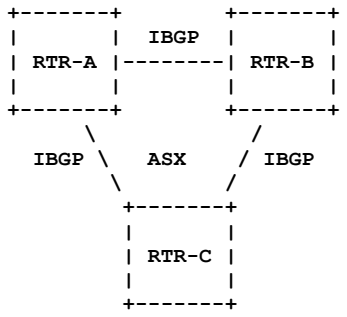

Рисунок 1. Полносвязная система IBGP.

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 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.

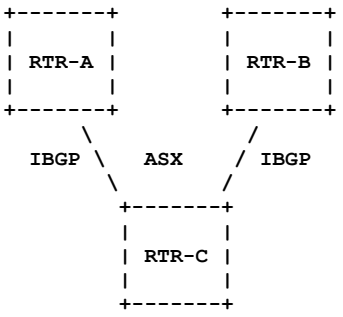

Рисунок 2. IBGP с отражением. маршрутов

Схема метода Route Reflection основана именно на этом принципе.

5. Терминология и концепции

Мы используем термин «отражение маршрутов» для описания действий узла BGP, анонсирующего полученный от IBGP маршрут другому партнеру IBGP. Если об узле BGP говорят как об «отражателе маршрутов» (Route Reflector или RR), это означает, что данный узел «отражает» полученные маршруты своим партнерам.

Внутренние партнеры узла RR делятся на две группы:

1) Партнеры-клиенты.

[2] В оригинале ошибочно сказано про 2 переходных атрибута, что не соответствует определениям главы 7. *Прим. перев.*

2) Партнеры, не являющиеся клиентами.

Узел RR отражает маршруты между этими группами и может отражать маршруты между клиентами. Узел RR вместе со своими клиентами образует кластер (Cluster). Партнеры, не являющиеся клиентами, должны сохранять полносвязность, но для клиентов это требование снимается. На рисунке 3 показан пример сети с базовыми компонентами RR, иллюстрирующий принятую здесь терминологию.

6. 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.

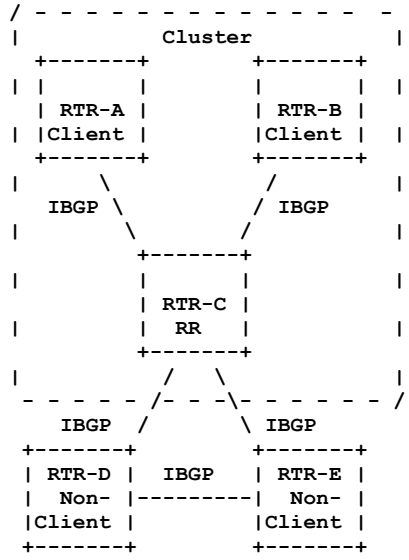

Рисунок 3. Компоненты RR.

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

   Отразить маршрут всем клиентам.

2) Маршрут получен от клиента.

   Отразить маршрут всем партнерам, не являющимся клиентами, а также партнерам-клиентам (поскольку клиенты могут не иметь полной связности).

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. Такие 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 (Route Reflection). 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.

7. Избыточные RR

Обычно кластер клиентов будет включать один рефлектор RR. В этом случае кластер будет идентифицироваться значением BGP Identifier 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.

8. Предотвращение петель

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

ORIGINATOR_ID

ORIGINATOR_ID - необязательный, непереходный атрибут BGP с кодом типа 9. Этот атрибут имеет размер 4 байта и создается рефлектором RR в отраженном маршруте. Атрибут будет включать значение BGP Identifier источника маршрута (originator) 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 BGP Identifier в качестве ORIGINATOR_ID.

CLUSTER_LIST

CLUSTER_LIST - необязательный, непереходный атрибут BGP с кодом типа 10. Этот атрибут представляет собой последовательность значений CLUSTER_ID, представляющих путь отражения, по которому передавался маршрут. Когда рефлектор RR отражает маршрут, он должен добавить локальное значение CLUSTER_ID в начало (prepend) CLUSTER_LIST. Если список CLUSTER_LIST пуст, узел должен создать новый список. Используя этот атрибут, RR может определять возникновение петель (возврат в тот же кластер) при передаче маршрутной информации в результате конфигурационных ошибок. Если локальное значение CLUSTER_ID присутствует в списке кластеров, полученный анонс следует игнорировать.

9. Влияние процесса выбора маршрута

Правила удаления лишнего в процессе выбора маршрута BGP[1] (параграф 9.1.2.2 документа [1]) изменяются следующим образом:

Если маршрут содержит атрибут ORIGINATOR_ID, тогда на этапе f) значение ORIGINATOR_ID **следует** трактовать как значение BGP Identifier узла BGP, анонсирующего маршрут.

Кроме того, между этапами f) и g) **следует** включить дополнительное правило:

Узлу BGP **следует** отдавать предпочтение маршруту с более коротким списком CLUSTER_LIST. Размер CLUSTER_LIST считается нулевым, если маршрут не включает атрибут CLUSTER_LIST.

10. 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.

Кроме того, когда RR отражает маршрут, ему **не следует** изменять в маршруте значения атрибутов NEXT_HOP, AS_PATH, LOCAL_PREF и MED, поскольку это может приводить к возникновению маршрутных петель.

11. Вопросы настройки и развертывания

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

Одним из ключевых моментов метода отражения маршрутов в контексте проблемы масштабирования является то, что RR обрабатывает полученную информацию и отражает только лучший путь.

На выбор маршрута BGP могут оказывать влияние обе метрики MED и IGP. Поскольку атрибуты MED не всегда совместимы, а метрика IGP может отличаться для каждого маршрутизатора, в некоторых вариантах топологии отражения метод отражения может давать при выборе маршрута результат, отличающийся от случая полносвязной системы IBGP. Для получения совпадающих результатов в случаях использования отражения и полносвязной системы IBGP следует сделать так, чтобы рефлекторы маршрутов никогда не выбирали лучший маршрут BGP на основе метрики IGP, которая существенно отличается от IGP-метрики их клиентов, или на основе несовместимых атрибутов MED. Первый вариант может быть достигнут путем настройки конфигурации таким образом, чтобы внутрикластерная метрика IGP всегда давала преимущество перед межкластерной метрикой IGP, и поддержки полной связности (full mesh) внутри кластера. Для реализации второго варианта можно использовать любой из перечисленных ниже способов:

- устанавливать на граничном маршрутизаторе уровень локального предпочтения маршрутов в соответствии с MED;
- обеспечить, чтобы длины AS_PATH для разных AS различались при использовании длины пути в качестве критерия выбора;
- настроить основанную на группах (community) политику, использование которой позволит рефлектору выбрать лучший путь.

Можно утверждать, что второй вариант вносит чрезмерные ограничения и в некоторых случаях будет непрактичным. Можно также утверждать, что при отсутствии маршрутных петель не существует жесткой необходимости обеспечивать совпадение результатов выбора маршрута с использованием отражения и полносвязной системы IBGP.

Для предотвращения маршрутных петель и поддержки согласованной картины маршрутизации важно аккуратно рассмотреть топологию сети при выборе топологии отражения маршрутов. В общем случае топологию отражения следует делать конгруэнтной топологии сети, когда существует множество путей для данного префикса. Общепринятым является использование отражения на базе POP[2], при котором каждая точка POP поддерживает свои рефлекторы маршрутов, обслуживающие клиентов POP, и все рефлекторы образуют между собой полносвязную систему. В дополнение к этому клиенты рефлекторов в каждой POP зачастую также образуют полносвязную систему в целях оптимальной маршрутизации внутри POP, а внутренняя (для POP) метрика IGP является предпочтительной по сравнению с метрикой inter-POP IGP.

12. Вопросы безопасности

Это расширение протокола BGP не изменяет состояния безопасности, присущего IBGP [1, 5].

13. Благодарности

Авторы благодарят Dennis Ferguson, John Scudder, Paul Traina и Tony Li за многочисленные дискуссии, результатом которых стал этот документ. Идея метода создана на основе давней дискуссии между Tony Li и Dimitri Haskin.

[1] BGP Decision Process Tie Breaking.
[2] Point of Presence - точка присутствия.

Кроме того, авторы хотят поблагодарить Yakov Rekhter за просмотр документа и полезные предложения, а также отметить полезные комментарии Tony Li, Rohit Dube и John Scudder и Bruce Cole.

14. Литература

14.1. Нормативные документы

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

14.2. Дополнительная литература

[2] Savola, P., "Reclassification of RFC 1863 to Historic", RFC 4223, October 2005.

[3] Traina, P., McPherson, D., and J. Scudder, "Autonomous System Confederations for BGP", RFC 3065, February 2001.

[4] Bates, T. and R. Chandra, "BGP Route Reflection An alternative to full mesh IBGP", RFC 1966, June 1996.

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

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

[7] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

Приложение A: Сравнение с RFC 2796

Добавлен параграф, посвящённый влиянию процесса выбора маршрутов BGP.

Удалён рисунок, иллюстрирующий формат атрибута CLUSTER_LIST, поскольку он дублировал описание в спецификации протокола BGP и поле размера атрибута по невнимательности было указано как одноктетное.[1]

Приложение B: Сравнение с RFC 1966

Изменения, перечисленные в Приложении A, а также изменения, перечисленные ниже.

Разъяснены некоторые термины, связанные с отражением маршрутов и исключены упоминания маршрутов и узлов EBGP.

Разъяснена и сделана более согласованной обработка получателем маршрутных петель (в результате отражения).

Способ добавления атрибута CLUSTER_ID в список CLUSTER_LIST был заменен с «append» на «prepend» в соответствии с реализованным кодом.

В главу «Вопросы настройки и развертывания» добавлено рассмотрение некоторых вопросов развертывания метода.

Адреса авторов

Tony Bates
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
EMail: tbates@cisco.com

Ravi Chandra
Sonoa Systems, Inc.
3255-7 Scott Blvd.
Santa Clara, CA 95054
EMail: rchandra@sonoasystems.com

Enke Chen
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
EMail: enkechen@cisco.com

Перевод на русский язык
Николай Малых
nmalykh@protokols.ru

[1] В определении атрибута ORIGINATOR_ID была произведена замена ROUTER_ID на BGP Identifier. *Прим. перев.*

Интеллектуальная собственность

IETF не принимает какой-либо позиции в отношении действительности или объема каких-либо прав интеллектуальной собственности (Intellectual Property Rights или IPR) или иных прав, которые, как может быть заявлено, относятся к реализации или использованию описанной в этом документе технологии, или степени, в которой любая лицензия, по которой права могут или не могут быть доступны, не заявляется также применение каких-либо усилий для определения таких прав. Сведения о процедурах IETF в отношении прав в документах RFC можно найти в BCP 78 и BCP 79.

Копии раскрытия IPR, предоставленные секретариату IETF, и любые гарантии доступности лицензий, а также результаты попыток получить общую лицензию или право на использование таких прав собственности разработчиками или пользователями этой спецификации, можно получить из сетевого репозитория IETF IPR по ссылке http://www.ietf.org/ipr.

IETF предлагает любой заинтересованной стороне обратить внимание на авторские права, патенты или использование патентов, а также иные права собственности, которые могут потребоваться для реализации этого стандарта. Информацию следует направлять в IETF по адресу ietf-ipr@ietf.org.

Подтверждение

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