Санкт-Петербург, ул. Седова, 80, телефон (812) 449-0770, факс (812) 449-0771, E-mail: info@bilim.com

Network Working Group
Request for Comments: 1772
Obsoletes: 1655
Category: Standards Track

Y. Rekhter
T.J. Watson Research Center, IBM Corp.
P. Gross
MCI
Editors

March 1995

Application of the Border Gateway Protocol in the Internet

Применение протокола BGP в Internet

Статус документа

Этот документ содержит спецификацию стандартного протокола, предложенного сообществу Internet, и является запросом для дальнейшего обсуждения в целях совершенствования протокола. Текущий статус протокола указан в "Internet Official Protocol Standards" (STD 1). Документ может распространяться свободно.

Тезисы

Этот документ вместе с "A Border Gateway Protocol 4 (BGP-4)"[1] определяет протокол маршрутизации между автономными системами в среде Internet. "A Border Gateway Protocol 4 (BGP-4)" содержит спецификацию протокола BGP, а данный документ описывает использование BGP в сети.
Информация об изменении протокола BGP и другие, связанные с этим протоколом сведения рассылаются по списку bgp@ans.net.

Благодарности

Первый вариант этого документа был выпущен как RFC 1164 в июне 1990 командой в составе Jeffrey C. Honig (Cornell University), Dave Katz (MERIT), Matt Mathis (PSC), Yakov Rekhter (IBM), Jessica Yu (MERIT).
В подготовке RFC 1164 значительную роль сыграли также Guy Almes (ANS, в то время Rice University), Kirk Lougheed (Cisco Systems), Hans-Werner Braun (SDSC, в то время MERIT), Sue Hares (MERIT).
Авторы выражают свою благодарность Bob Braden (ISI) за его обзор предыдущего варианта этого документа.
Обновленная версия документа является результатом работы группы IETF BGP, с Phill Gross (MCI) и Yakov Rekhter (IBM) в качестве редакторов.
John Moy (Proteon) внес свой вклад в подготовку главы 7.
Scott Brim (Cornell University) принимал участие в создании основы главы 8.
В создании большей части текста главы 9 принимал участие Gerry Meyer (Spider).
Фрагменты введения почти дословно перенесены из работы [3].
Авторы выражают свою благодарность Dan Long (NEARNET) и Tony Li (Cisco Systems) за их обзор и комментарии к текущей версии документа.
Работу Yakov Rekhter частично финансировал National Science Foundation в рамках Grant Number NCR-9219216.

1. Введение

В этом документе рассматривается использование протокола BGP (Border Gateway Protocol) [1] в среде Internet. BGP представляет собой протокол маршрутизации между автономными системами (AS). Информация о доступности сетей, передаваемая с помощью BGP, обеспечивает достаточно данных для обнаружения петель и принятия решений о маршрутизации на основе предпочтений и политики, согласованных в соответствии с RFC 1104 [2]. В частности, BGP обеспечивает обмен маршрутными данными, содержащими полные пути AS и обеспечивающими реализацию политики маршрутизации на основе конфигурационных параметров.
По мере роста сети Internet стали очевидными серьезные проблемы масштабирования, включающие:

- Нехватку адресного пространства для сетей класса B. Одной из фундаментальных причин такой нехватки является бедность выбора класса сети по числу поддерживаемых адресов. Сети класса C позволяют адресовать 254 хоста и слишком малы для множества организаций, тогда как сети класса B, позволяющие адресовать 65534 хоста, слишком велики.
- Рост размеров таблиц маршрутизации в сети Internet превышает возможности программ (и людей) по эффективному управлению этими таблицами.
- Возможность нехватки пространства 32-битовых адресов IP.

Очевидно, что первые две проблемы достигнут критического уровня в ближайшие три года[2]. Бесклассовая междоменная маршрутизация CIDR (Classless inter-domain routing) является попыткой решения этих проблем за счет замедления роста таблиц маршрутизации и выделения новых номеров для сетей IP. CIDR не включает попыток решения третьей проблемы, которая по своей природе является не такой критичной во времени, но в результате возникает ряд осложнений, связанных с обеспечением эффективной работы сети Internet до решения глобальной проблемы нехватки адресного пространства.

[1] RFC 1771 (на сайте www.protocols.ru имеется перевод этого документа на русский язык). *Прим. перев.*
[2] Т. е. в 1996-98 гг. *Прим. перев.*

Протокол BGP-4 является расширением BGP-3, обеспечивающим поддержку агрегирования маршрутной информации и снижения объема передаваемых данных за счет использования архитектуры бесклассовой междоменной маршрутизации CIDR [3]. Данный документ описывает использование BGP-4 в среде Internet.

Обсуждение в этом документе основано на рассмотрении сети Internet как набора произвольным образом соединенных автономных систем (AS). Таким образом, моделью Internet будет служить общий граф, узлами которого являются AS, а ребра соединяют пары AS.

По классическому определению автономная система представляет собой множество маршрутизаторов с единым техническим администрированием, использующих один протокол внутренней маршрутизации (IGP) и единую метрику для маршрутизации пакетов внутри AS, а для передачи пакетов в другие автономные системы применяющих протокол внешней маршрутизации (exterior gateway protocol или EGP). Со временем классическое определение было расширено и в современном понимании AS может использовать несколько протоколов внутренней маршрутизации, а в некоторых случаях даже несколько наборов метрик в рамках одной AS. Использование термина AS в таких случаях обусловлено тем, что даже при использовании множества метрик и протоколов IGP администрирование такой AS с точки зрения других автономных систем выглядит как единый план внутренней маршрутизации и показывает согласованную картину доступности адресатов с использованием данной AS.

2. Топологическая модель BGP

Когда мы говорим о соединении между двумя AS, следует различать два аспекта:

Физическое соединение – общая подсеть канального (Data Link) уровня между двумя AS, в которой каждая из AS имеет, по крайней мере, один граничный маршрутизатор, относящийся к данной AS. Таким образом, граничный маршрутизатор каждой AS может пересылать пакеты граничному маршрутизатору другой AS без использования внутридоменной или междоменной маршрутизации.

Соединение BGP – сеанс BGP между двумя узлами BGP в каждой AS, используемый для анонсирования маршрутов, которые могут быть использованы для тех или иных адресатов.

В этом документе к узлам BGP, формирующим BGP-соединение, предъявляется ряд дополнительных требований. Эти узлы должны находиться в той же подсети канального уровня (Data Link subnetwork), к которой относятся их граничные маршрутизаторы. Таким образом, сессия BGP между смежными AS не требует поддержки внутридоменной или междоменной маршрутизации. Ситуации, когда это требование не выполняется, выходит за пределы рассмотрения данного документа.

Таким образом, в любом соединении каждая AS имеет, по крайней мере, один узел BGP и, по крайней мере, один граничный маршрутизатор – эти узлы BGP и граничные маршрутизаторы должны находиться в одной подсети канального уровня. Отметим, что узлы BGP не обязаны быть граничными маршрутизаторами и наоборот. Пути, анонсируемые узлом BGP одной AS в данном соединении, пригодны для всех граничных маршрутизаторов другой AS той же подсети канального уровня, т. е., возможен обмен при отсутствии прямого соседства.

Большая часть трафика, передаваемого внутри AS, генерируется или завершается в этой AS (т. е., IP-адрес отправителя или получателя пакетов IP идентифицирует хост, входящий в данную AS). Трафик, соответствующий этому описанию, будем называть локальным, а весь остальной трафик – транзитным. Основной задачей протокола BGP является управление потоками транзитного трафика.

В зависимости от того, как конкретная AS поступает с транзитным трафиком, автономные системы можно разделить на три категории:

Тупиковая (stub) AS – автономная система, имеющая соединение только с одной AS. Очевидно, что тупиковая AS может поддерживать только локальный трафик.

Многодомная (multihomed) AS – автономная система, соединенная с несколькими AS, но не принимающая транзитный трафик.

Транзитная AS – автономная система, соединенная с множеством других AS и предназначенная (с некоторыми ограничениями на уровне политики) для поддержки как локального, так и транзитного трафика.

Поскольку полный путь AS обеспечивает простой и эффективный способ предотвращения маршрутных петель и избавляет от проблемы "count-to-infinity" (подсчет бесконечности), связанной с некоторыми алгоритмами на базе вектора расстояния, протокол BGP не вносит ограничений в топологию соединений между AS.

3. BGP в сети Internet

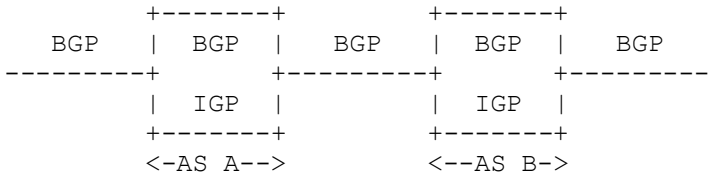
3.1 Топология

Топологию сети Internet в целом можно рассматривать как произвольное соединение транзитных, многодомных и тупиковых AS. Для минимизации влияния на текущую инфраструктуру Internet со стороны тупиковых и многодомных AS не требуется использования BGP. Такие AS могут использовать другие протоколы (например, EGP) для обмена данными о доступности с транзитными AS. Транзитные AS, использующие BGP, будут отмечать такие данные, как полученные способом, отличным от BGP. Факт необязательности использования BGP в многодомных и тупиковых AS не оказывает отрицательного влияния на качество междоменной маршрутизации для трафика, который адресован в такую AS или исходит из нее.

Рекомендуется, однако, использовать протокол BGP и для тупиковых/многодомных AS. В таких ситуациях протокол BGP обеспечивает более высокую производительность и меньший расход полосы, по сравнению с другими протоколами (типа EGP). Кроме того, этот протокол избавляет от необходимости задания принятых по умолчанию маршрутизаторов.

3.2 Глобальная природа BGP

С глобальной точки зрения протокол BGP используется для распространения маршрутной информации между множеством автономных систем. Поток информации можно представить следующим образом:



Из приведенной диаграммы видно, что BGP служит для передачи информации между AS, а внутри каждой автономной системы может использоваться как BGP, так и IGP. Обеспечение совместимости маршрутной информации между протоколами BGP и IGP внутри AS имеет важное значение и рассматривается в Приложении A.

3.3 Соседство в BGP

Мы рассматриваем Internet как множество произвольно соединенных между собой AS. Маршрутизаторы, непосредственно взаимодействующие один с другим про протоколу BGP, называют узлами BGP (BGP speaker). Узлы BGP могут располагаться в одной или разных AS. Узлы BGP в каждой AS обмениваются между собой информацией о доступности сетей на основе наборов правил, заданных в каждой AS. По отношению к данному узлу BGP другие узлы, с которыми данный узел взаимодействует, называются внешними партнерами (external peer), если они размещаются в другой AS, или внутренними партнерами (internal peer), если они находятся в той же AS.

В автономной системе может присутствовать множество узлов BGP (сколько требуется для данной AS). Обычно автономные системы, имеющие множество соединений с другими AS, должны использовать более одного узла BGP. Все узлы BGP, представляющие одну AS, должны предоставлять согласованные маршрутные данные. Это требует согласования маршрутной информации между узлами BGP в автономной системе. Маршрутизаторы могут обмениваться информацией по протоколу BGP или иным способом. Правила, применяемые ко всем узлам BGP одной AS, должны быть согласованы. Для обнаружения рассогласований могут использоваться методы типа tagged IGP (см. стр. 7).

В случае внешних партнеров узлы находятся в разных AS, но должны входить в общую подсеть канального уровня (Data Link subnetwork). Эта общая подсеть должна использоваться для передачи информации BGP между узлами. Использование BGP через AS, мешающие обмену, делает информацию о путях AS некорректной. Для указания автономной системы, к которой относится узел BGP, следует использовать номер AS.

4. Требования к агрегированию маршрутов

Соответствующие требованиям BGP-4 реализации должны обеспечивать возможность информирования о тех случаях, когда агрегированный маршрут может быть сгенерирован только для части маршрутной информации. Например, узел BGP на границе автономной системы (или группы AS) должен быть способен генерировать агрегированный маршрут для целого набора IP-адресов получателей (в терминах BGP-4 такой набор называется Network Layer Reachability Information или NLRI), по отношению к которым он имеет административный контроль (включая делегированные узлом адреса), даже если некоторые из этих адресов в данный момент недоступны.

Соответствующие спецификации BGP-4 реализации протокола могут обеспечивать возможность указать, когда может генерироваться агрегированный NLRI.

Соответствующие спецификации BGP-4 реализации протокола должны обеспечивать возможность указать, как можно деагрегировать NLRI.

Соответствующие спецификации BGP-4 реализации протокола должны поддерживать перечисленные ниже опции при работе с перекрывающимися маршрутами:

- установка как более специфичных, так и менее специфичных маршрутов
- установка только более специфичного маршрута
- установка только менее специфичного маршрута
- отказ от установки маршрута

Некоторые аспекты политики маршрутизации могут зависеть от NLRI (например, "research" в сравнении с "commercial"). Следовательно, узел BGP, выполняющий агрегирование маршрутов, должен быть (по возможности) осведомлен о потенциальной причастности правил маршрутизации при агрегировании NLRI.

5. Создание политики BGP

BGP обеспечивает возможность реализации маршрутной политики на основе различных предпочтений и условных соглашений. Политика не заложена в протокол непосредственно, она передается протоколу BGP в форме конфигурационных параметров.

BGP реализует политику путем выбора варианта пути при наличии множества возможностей и путем контроля за распространением маршрутной информации. Политика определяется администраторами AS.

Политика маршрутизации связана с вопросами безопасности, экономическими и политическими аспектами. Например, если AS не желает передавать трафик в какую-либо другую AS, она может реализовать политику, блокирующую передачу такого трафика. Ниже приведено несколько примеров политики маршрутизации, которая может быть реализована при использовании BGP:

- Многодомная AS может отказаться выполнять функции транзитной для других AS (это осуществляется путем анонсирования маршрутов только к внутренним адресатам данной AS).
- Многодомная AS может обеспечивать транзит для ограниченного набора смежных AS (только некоторые из смежных AS могут использовать данную AS в качестве транзитной). Такая политика реализуется путем анонсирования маршрутной информации только в ограниченное число AS.
- AS может предпочесть использование некоторых других AS в качестве транзитных или наоборот, отказаться от транзита через те или иные автономные системы.

С помощью протокола BGP можно контролировать множество аспектов производительности:

- AS может минимизировать число транзитных AS (короткий путь предпочтительней длинного).
- Качество транзитных AS. Если данная AS знает более одного пути, по которому можно достичь адресата, AS может использовать различные критерии выбора пути из числа возможных. Транзитные AS могут оцениваться по

диаметру, скорости канала, емкости, тенденции к насыщению, качеству работы. Информация об этих параметрах может определяться не только с помощью BGP, но и другими средствами.

◆Внутренние маршруты более предпочтительны, чем внешние.

Для обеспечения согласованности внутри AS пути с равной стоимостью, определяемые в результате применения политики и/или обычных процедур выбора, маршрута должны сравниваться согласованным способом.

Основой BGP является правило, согласно которому AS анонсирует в соседние AS только те маршруты, которые использует сама. Это правило отражает парадигму поэтапной (hop-by-hop) маршрутизации, широко используемую в современной сети Internet.

6. Выбор пути для BGP

Одной из основных задач узла BGP является оценка различных путей от этого узла к множеству адресатов, указанному префиксом адреса, выбор лучшего пути, применение подходящей политики и анонсирование маршрута всем своим BGP-соседям. Ключевым вопросом является оценка и сравнение путей. В традиционных протоколах на основе вектора расстояния (например, RIP) с путем связывается единственная метрика (например, счетчик интервалов). В таких случаях выбор маршрута определяется простым сравнением двух чисел. Сложность маршрутизации между AS обусловлена отсутствием универсальной согласованной между AS метрики для оценки внешних маршрутов. Чаще всего каждая AS использует свой способ оценки внешних путей.

Узел BGP создает базы данных о маршрутах, содержащие все возможные пути и список адресатов (указываются адресными префиксами), доступных с использованием каждого из путей. Для более точного описания полезно рассматривать набор возможных путей для набора адресатов, связанного с данным префиксом адреса. В большинстве случаев можно предполагать существование единственно возможного пути. Однако в тех случаях, когда возможно несколько путей, должны поддерживаться все эти пути и быстрый переход от одного пути к другому позволяет предотвратить потерю основного пути. Анонсироваться в каждый момент времени будет только основной путь.

Процесс выбора пути может быть формализован путем определения порядка для всего набора возможных путей к набору адресатов, указанных данным префиксом адреса. Одним из способов определения такого порядка является функция отображения каждого полного пути AS в неотрицательное целое число, определяющее уровень предпочтения для этого пути. После этого процесс выбора пути сводится к применению функции ко всем возможным путям и выбору пути с максимальным уровнем предпочтения.

В реализациях BGP критерии определения уровня предпочтения для каждого из путей задаются конфигурационными параметрами.

Процесс определения уровня предпочтения для каждого из путей может основываться на нескольких источниках информации:

◆сведения, явно указанные в полном пути AS.

◆комбинация сведений, которые могут быть получены из полного пути AS, и данных, полученных из других (не BGP) источников (например, условия политики маршрутизации, заданные конфигурационными параметрами).

При выборе уровня предпочтения для пути возможно использование следующих критериев:

Счетчик AS. Путь с меньшим числом AS обычно является более предпочтительным.

Учет политики. BGP поддерживает маршрутизацию на основе правил, базирующуюся на контролируемом распространении маршрутной информации. Узел BGP может принимать во внимание некоторые аспекты политики (как внутри своей AS, так и за ее пределами) при выборе пути. Путь, не соответствующий требованиям политики просто исключается из рассмотрения.

Присутствие или отсутствие заданных AS в пути. Используя информацию, полученную от других протоколов (не BGP), AS может знать некоторые параметры производительности (например, полосу, MTU, внутренний диаметр AS) некоторых AS и использовать это знание при выборе пути.

Источник пути. Путь, полученный только от BGP (т. е., конечная точка пути является внутренней по отношению к последней AS на пути), в общем случае лучше путей, информация о которых получена от EGP или иным способом.

Подмножество пути AS. Путь AS, являющийся подмножеством более длинного пути AS к тому же адресату, является более предпочтительным. Любые проблемы, которые могут возникнуть на коротком пути, будут присутствовать и на длинном.

Динамика каналов. Стабильные пути более предпочтительны, нежели нестабильные. Отметим, что этот критерий нужно использовать с осторожностью во избежание ненужных флуктуаций маршрутов. В общем случае любые критерии, зависящие от динамической информации, следует использовать с осторожностью.

7. Требуемый набор поддерживаемых правил маршрутизации

Политика BGP задается в форме конфигурационных параметров. Эта информация не содержится непосредственно в протоколе, следовательно, BGP может поддерживать очень сложные схемы политики маршрутизации. Однако поддержка такой политики не требуется от любой реализации BGP.

Мы не пытаемся стандартизировать политику маршрутизации, которая должна поддерживаться каждой реализацией BGP, но настоятельно рекомендуем всем разработчикам поддерживать перечисленные ниже возможности:

Реализация BGP должно позволять AS контроль за анонсированием полученных BGP маршрутов в смежные AS. Реализация должна также поддерживать такой контроль с гранулярностью по крайней мере одного адресного префикса. Для такого контроля должна поддерживаться также гранулярность на уровне автономных систем (исходных AS для пути или AS, анонсирующих маршрут в локальную систему – смежную AS). Следует с осторожностью подходить к выбору узлом BGP новых маршрутов, которые не могут быть анонсированы тем или иным внешним узлам, если ранее эти маршруты анонсировались тем же узлам. В частности, локальная система должна явно показывать партнеру, что анонсированный ранее маршрут сейчас недоступен.

Реализациям BGP следует давать AS возможность предпочтения конкретного пути к адресату (когда доступно несколько путей). Реализация должна, по крайней мере, поддерживать такую функциональность для административного выбора уровня предпочтения маршрутов на основе IP-адреса соседа, от которого получен маршрут. Допустимый диапазон значений уровня предпочтения составляет 0 - $(2^{31} – 1)$.

Реализации BGP следует предоставлять AS возможность игнорировать маршруты с некоторыми AS в атрибуте AS_PATH. Такая функция может быть реализована с использованием метода, описанного в работе [2], или путем присваивания таким AS «бесконечного» веса. Процесс выбора маршрута должен игнорировать пути с бесконечным весом.

8. Взаимодействие с другими протоколами внешней маршрутизации

Приведенные здесь рекомендации согласованы с рекомендациями работы [3].

AS следует анонсировать минимальный блок своих внутренних адресатов в соответствии с реальным использованием адресного пространства. Эта информация может использоваться автономными системами, не применяющими BGP 4 для определения количества маршрутов, которые могут быть выделены из такого блока.

Маршрут, содержащий атрибут пути ATOMIC_AGGREGATE, не следует экспортировать в системы BGP-3 или EGP2, если такой экспорт не может быть организован без использования NLRI для маршрута.

8.1 Обмен информацией с протоколом EGP2

Предлагается осуществлять обмен маршрутной информацией между BGP-4 и EGP2 в соответствии с приведенными здесь рекомендациями.

Для обеспечения элегантного перехода узел BGP может принимать участие в работе EGP2, используя одновременно BGP-4. Таким образом, узел BGP может получать данные о доступности IP-адресов как от EGP2, так и от BGP-4. Информация, полученная от EGP2, может быть помещена в BGP-4 с установкой для атрибута пути ORIGIN значения 1. Подобно этому, сведения от BGP-4 могут быть переданы EGP2. В последнем случае следует принимать во внимание сложности, которые могут возникнуть в тех случаях, когда префикс IP, полученный от BGP-4, обозначает набор последовательных сетей класса A/B/C. Вставка полученных от BGP-4 значений NLRI, соответствующих подсетям IP, требует от узла BGP вставки соответствующей сети в EGP2. Локальная система должна обеспечивать механизм контроля за обменом информацией о доступности между EGP2 и BGP-4. В частности, соответствующие стандарту реализации должны поддерживать при вставке полученных от BGP-4 сведений о доступности в EGP2 следующие опции:

- по умолчанию используется только (0.0.0.0) без экспорта каких либо NLRI
- обеспечивается контролируемое деагрегирование, но только для указанных маршрутов; обеспечивается экспорт неагрегированных NLRI
- обеспечивается экспорт только неагрегированных NLRI

Обмен маршрутной информацией на основе EGP2 между узлами BGP, участвующими в работе BGP-4, и чистыми узлами EGP2 может происходить только на границах доменов (автономных систем).

8.2 Обмен информацией с протоколом BGP-3

Предлагаемая схема обмена маршрутной информацией между BGP-4 и BGP-3, описана ниже.

Для обеспечения элегантного перехода узел BGP может принимать участие в работе BGP-3, используя в то же время BGP-4. Таким образом, узел BGP может поучать данные о доступности адресов IP как от BGP-3, так и от BGP-4.

Узел BGP может помещать полученные от BGP-4 сведения в BGP-3 описанным здесь способом.

Если атрибут AS_PATH маршрута BGP-4 содержит сегменты пути AS_SET, атрибут AS_PATH маршрута BGP-3 следует строить, трактуя сегменты AS_SET, как сегменты AS_SEQUENCE, чтобы результирующее значение AS_PATH являлось одним AS_SEQUENCE. Хотя такая процедура приводит к потере информации set/sequence, она не влияет на процесс подавления маршрутных петель, но может влиять на политику, если последняя основана на содержимом или порядке атрибута AS_PATH.

При вставке полученных от BGP-4 значений NLRI в BGP-3 следует принимать во внимание сложности, которые могут возникнуть в тех случаях, когда префикс IP, полученный от BGP-4, обозначает набор последовательных сетей класса A/B/C. Вставка полученных от BGP-4 значений NLRI, которые обозначают подсети IP, требует от узла BGP вставки соответствующей сети в BGP-3. Локальная система должна обеспечивать механизм контроля за обменом информацией о доступности между BGP-3 и BGP-4. В частности, соответствующие стандарту реализации должны поддерживать при вставке полученных от BGP-4 сведений о доступности в BGP-3 следующие опции:

- по умолчанию используется только (0.0.0.0) без экспорта каких либо NLRI
- обеспечивается контролируемое деагрегирование, но только для указанных маршрутов; обеспечивается экспорт неагрегированных NLRI
- обеспечивается экспорт только неагрегированных NLRI

Обмен маршрутной информацией с использованием BGP-3 между узлами BGP, участвующими в работе BGP-4, и чистыми узлами BGP-3 может происходить только на границах автономных систем. В одной автономной системе BGP-обмен между узлами BGP осуществляется всегда на базе BGP-3 или BGP-4, но не в смешанном варианте.

9. Работа в системах с SVC

При использовании BGP в подсетях с SVC (коммутируемые виртуальные устройства) может оказаться желательной минимизация трафика, генерируемого BGP. В частности, может оказаться желательным избавление от трафика, связанного с периодическими сообщениями KEEPALIVE. Протокол BGP включает специальный механизм для работы с SVC, позволяющий избежать постоянного использования коммутируемых каналов и не передавать периодических сообщений KEEPALIVE.

В этом разделе описана организация работы без периодических сообщений KEEPALIVE для минимизации использования SVC при работе с интеллектуальными менеджерами устройств SVC. Предложенная схема может использоваться и для постоянных соединений, которые поддерживают функции типа мониторинга качества или эхо-запросов для контроля за состоянием соединений.

Описанный здесь механизм может использоваться только при непосредственном соединении между узлами BGP через общий виртуальный канал.

9.1 Организация соединения BGP

Для активизации этого режима нужно задать нулевое значение параметра Hold Time в сообщении OPEN.

9.2 Свойства менеджера устройств

Возможности менеджера устройств должны обеспечивать компенсацию потери периодических сообщений KEEPALIVE:

- Менеджер устройств должен обеспечивать возможность определения недоступности канала за предсказуемое конечное время после сбоя на канале.
- После определения недоступности канала менеджер должен:
  - запустить настраиваемый dead-таймер (значение таймера сравнимо с типичным значением Hold timer).
  - попытаться восстановить соединение на канальном уровне.
- По завершении отсчета dead-таймера менеджер должен:
  - передать информацию о разрыве соединения внутренними средствами (индикация DEAD) протоколу TCP.
- После того, как соединение будет восстановлено, менеджер должен:
  - сбросить dead-таймер.
  - передать информацию о восстановлении (внутренний сигнал UP) протоколу TCP.

9.3 Свойства TCP

Для обработки сигналов от менеджера устройств требуется внести небольшие изменения в работу протокола TCP:

DEAD – сброс очередей передачи и разрыв соединения TCP.

UP – передача всех данных из очереди и возможность обработки входящих вызовов TCP.

9.4 Координация работы

Некоторые реализации не могут гарантировать, что процесс BGP и менеджер устройств будут работать как единое целое – т. е., один процесс может существовать сам по себе в результате остановки или краха другого процесса.

В таких случаях должен быть реализован двухсторонний опрос между процессом BGP и менеджером устройств. Если процесс BGP обнаруживает, что менеджер устройств не работает, он должен закрыть все связанные с ним соединения TCP. Если менеджер устройств определяет неработоспособность процесса BGP, он должен закрыть все связанные с этим процессом соединения и отвергать все входящие вызовы.

10. Заключение

Протокол BGP обеспечивает высокий уровень контроля и гибкости для междоменной маршрутизации, позволяя устанавливать политику и условия работы, а также предотвращая возникновение маршрутных петель. Представленное в этом документе руководство может служить отправной точкой при использовании BGP для организации мощной и управляемой системы маршрутизации в растущей сети Internet.

Приложение A. Взаимодействие BGP и IGP

В этом разделе схематически рассмотрены методы обмена маршрутной информацией между протоколами BGP и IGP. Описанная схема не предлагается как часть стандарта на использование BGP и приведена только для информации. Разработчики могут использовать описанные здесь методы при импортировании данных IGP.

Приведенная здесь информация имеет общий смысл и может применяться к любому протоколу IGP.

Взаимодействие между BGP и каким-либо конкретным протоколом IGP не рассматривается в данном документе и может быть включено в будущие стандарты.

Обзор

По определению все транзитные AS должны обеспечивать передачу трафика, отправители и/или получатели которого находятся за пределами данной AS. Это требует некоторого взаимодействия и координации между BGP и протоколами внутренней маршрутизации (IGP), используемыми в каждой AS. В общем случае трафик, сгенерированный за пределами данной AS передается с использованием как внутренних (поддерживают только IGP), так и граничных (поддержка IGP и BGP) маршрутизаторов. Все внутренние маршрутизаторы получают сведения о внешних маршрутах от одного или нескольких граничных маршрутизаторов AS по протоколам IGP.

В зависимости от механизма, используемого для передачи информации BGP в данной AS, могут потребоваться специальные средства согласования между протоколами BGP и IGP, поскольку сведения об изменении состояний распространяются в AS с различными скоростями. В качестве такого средства может использоваться временное окно между моментом, когда тот или иной граничный маршрутизатор (A) получает новые маршрутные данные BGP, происходящие от другого граничного маршрутизатора (B) в той же AS, и моментом, когда IGP в данной AS сможет передавать транзитный трафик граничному маршрутизатору (B). Пока длится это «временное окно», возможна некорректная маршрутизация или возникновение «черных дыр».

Для минимизации влияния таких проблем граничному маршрутизатору (A) не следует анонсировать кому-либо из своих внешних партнеров маршрут к некому набору внешних адресатов, связанных с данным префиксом X через граничный маршрутизатор (B) до тех пор, пока внутренние маршрутизаторы в AS не будут готовы передавать трафик этим адресатам через корректный выходной граничный шлюз (B). Иными словами, внутренняя маршрутизация должна сойтись на подходящем выходном граничном маршрутизаторе до анонсирования маршрутов через этот шлюз внешним партнерам.

A.2 Методы стабилизации

Ниже схематически рассмотрены некоторые методы стабилизации взаимодействия между протоколами BGP и IGP внутри автономной системы.

A.2.1 Распространение информации BGP по протоколу IGP

Хотя протокол BGP может обеспечивать перенос информации BGP внутри AS за счет собственных средств, можно также использовать протокол IGP для транспортировки такой информации, если IGP поддерживает лавинную рассылку (flooding) маршрутных данных (обеспечивает механизм распространения информации BGP) и обеспечивает достаточную сходимость. Если протокол IGP используется для передачи информации BGP, тогда описанный выше период рассинхронизации не будет возникать совсем, поскольку информация BGP распространяется внутри AS синхронно с

IGP и схождение IGP происходит более или менее синхронно с прибытием новых маршрутных данных. Отметим, что IGP только переносит информацию BGP и не должен интерпретировать или обрабатывать эти данные.

A.2.2 Tagged IGP

Некоторые протоколы IGP могут помечать с идентификацией точек выхода маршруты, являющиеся внешними по отношению к AS, при их распространении внутри AS. Каждый граничный маршрутизатор должен использовать идентичные метки (теги) для анонсирования внешней маршрутной информации (полученной от BGP) как для IGP, так и при распространении этой информации другим внутренним партнерам (узлы той же AS). Метки, генерируемые граничным маршрутизатором, должны уникально идентифицировать конкретный граничный маршрутизатор (другой маршрутизатор должен использовать отличающиеся метки).

Все граничные маршрутизаторы одной AS должны придерживаться следующих правил:

Информация, полученная от внутреннего партнера граничным маршрутизатором A и декларирующая недоступность набора адресатов, связанных с данным префиксом, должна незамедлительно рассылаться всем внешним партнерам A.

Информация, полученная от внутреннего партнера граничным маршрутизатором A и связанная с доступностью адресатов, которые указаны префиксом X, не может распространяться внешним партнерам A, пока A не имеет внутреннего маршрута IGP к набору адресатов, связанному с префиксом X, а также маршрутной информации IGP и BGP, имеющей идентичные теги.

Соблюдение этих правил дает гарантию того, что маршрутная информация не будет распространяться за пределы AS, пока IGP не обеспечит корректную поддержку для этих маршрутов. Эти правила позволяют также предотвратить возникновение «черных дыр».

Одним из методов обозначения маршрутов BGP и IGP внутри AS является использование в качестве метки IP-адреса выходного граничного маршрутизатора, анонсировавшего внешний маршрут в AS. В этом случае поле "gateway" сообщения BGP UPDATE служит в качестве метки.

Другой метод установки меток для маршрутов BGP и IGP заключается в использовании протоколами BGP и IGP согласованных идентификаторов для маршрутизаторов (Router ID). В этом случае значение Router ID доступно всем узлам BGP (начиная с версии 3). Поскольку этот идентификатор является уникальным, его можно применять в качестве тега IGP.

A.2.3 Инкапсуляция

Инкапсуляция обеспечивает простейший (с точки зрения взаимодействия IGP и BGP) механизм передачи транзитного трафика через AS. В этом случае транзитный трафик просто инкапсулируется в дейтаграммы IP, адресованные выходному маршрутизатору. От протокола IGP в этом случае требуется лишь поддержка маршрутизации между граничными маршрутизаторами одной AS.

Адрес выходного маршрутизатора A для некого внешнего адресата X указывается в поле идентификатора BGP сообщения BGP OPEN, принятого от маршрутизатора A (по протоколу BGP) всеми прочими граничными маршрутизаторами этой AS. Для маршрутизации трафика адресату X каждый граничный маршрутизатор в AS инкапсулирует этот трафик в дейтаграммы, адресованные маршрутизатору A. Получив такие дейтаграммы, маршрутизатор A извлекает из них пакеты и пересылает подходящему маршрутизатору другой AS.

Поскольку инкапсуляция не требует от IGP переноса внешних маршрутных данных, не нужна и синхронизация между протоколами BGP и IGP.

При использовании такого метода должен быть определен способ идентификации инкапсулированных пакетов IP (например, тип протокола IP).

Отметим, что при инкапсуляции пакетов, размер которых близок к значению MTU, пакет будет фрагментироваться выполняющим инкапсуляцию маршрутизатором.

A.2.4 Повсеместное распространение BGP

Если все маршрутизаторы в AS являются узлами BGP, не возникает необходимости организации взаимодействия между BGP и IGP. В таких случаях все маршрутизаторы AS уже имеют полную информацию о маршрутах BGP и IGP используется только для маршрутизации внутри AS. Маршруты BGP не импортируются в IGP.

Маршрутизаторы, работающие в такой манере, должны обеспечивать возможность рекурсивного просмотра таблиц маршрутизации. Первый просмотр будет использовать маршрут BGP для определения выходного маршрутизатора, а при втором просмотре будет определяться путь IGP к этому маршрутизатору.

Поскольку в этом варианте IGP не переносит внешних маршрутных данных, все маршрутизаторы AS будут обеспечивать схождение, как только все узлы BGP получат информацию о новом маршруте. Задержки для IGP не требуется и маршруты можно анонсировать сразу же.

A.2.5 Другие случаи

Возможно существование AS, в которых протокол IGP не может ни переносить информацию BGP, ни помечать внешние маршруты (например, RIP). Кроме того, инкапсуляция может оказаться нежелательной или невозможной. В таких случаях нужно следовать приведенным ниже правилам:

Полученная от внутреннего партнера граничным маршрутизатором A информация о недоступности адресатов должна незамедлительно рассылаться всем внешним партнерам A.

Полученная от внутреннего партнера граничным маршрутизатором A информация о доступности адресата X, не может передаваться кому-либо из внешних партнеров A, пока маршрутизатор A не имеет маршрута IGP к адресату X и не пройдет время, достаточное для схождения маршрутов IGP.

Выполнение приведенных выше правил является необходимым (но не достаточным) условием для распространения маршрутной информации BGP в другие AS. В отличие от tagged IGP, эти правила не обеспечивают гарантии того, что внутренние маршруты к подходящим выходным шлюзам будут включены в таблицу до того, как начнется рассылка маршрутной информации в другие AS.

Если время схождения IGP меньше некого малого значения X, временное окно, в течение которого отсутствует синхронизация между IGP и BGP, также будет меньше X и проблему можно игнорировать за исключением кратких

(меньше X) периодов маршрутной нестабильности. Определение значений X является предметом для дальнейшего изучения, но (по возможности) это значение должно быть меньше секунды.
Если время схождения IGP нельзя игнорировать, требуется использовать другие механизмы. В настоящее время ведется работа по изучению таких механизмов.

Литература

[1] Rekhter Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4), RFC 1771[3], T.J. Watson Research Center, IBM Corp., Cisco Systems, March 1995.
[2] Braun, H-W., "Models of Policy Based Routing", RFC 1104[4], Merit/NSFNET, June 1989.
[3] Fuller, V., Li, T., Yu, J., and K. Varadhan, "Supernetting: an Address Assignment and Aggregation Strategy", RFC1519, BARRNet, Cisco, MERIT, OARnet, September 1993.

Вопросы безопасности

Вопросы безопасности не рассматриваются в данном документе.

Адреса авторов

Yakov Rekhter
T.J. Watson Research Center IBM Corporation
P.O. Box 704, Office H3-D40
Yorktown Heights, NY 10598
Phone: +1 914 784 7361
EMail: yakov@watson.ibm.com

Phill Gross
MCI Data Services Division
2100 Reston Parkway, Room 6001,
Reston, VA 22091
Phone: +1 703 715 7432
EMail: 0006423401@mcimail.com

Список рассылки IETF IDR WG: bgp@ans.net
Адрес для добавления в список: bgp-request@ans.net

Перевод на русский язык

Николай Малых
BiLiM Systems
nmalykh@acm.org
телефон: (812) 449-0770
факс: (812) 449-0771

[3] На сайте www.protocols.ru имеется перевод этого документа на русский язык. *Прим. перев.*
[4] На сайте www.protocols.ru имеется перевод этого документа на русский язык. *Прим. перев*