Network Working Group
Request for Comments: 1773
Obsoletes: 1656
Category: Informational

P. Traina
cisco Systems

March 1995

Опыт использования протокола BGP-4

Experience with the BGP-4 protocol

Статус документа

Этот документ содержит информацию для сообщества Internet. Документ не задает каких-либо стандартов Internet и может распространяться свободно.

Введение

Целью настоящего документа является демонстрация того, как требования к повышению эффективности протокола маршрутизации, предложенного в проекте стандарта (Draft Standard) могут быть выполнены при использовании протокола BGP-4 (Border Gateway Protocol version 4). Документ основан на реальном опыте использования протокола BGP и является вторым и пары документов, посвященных опыту использования BGP. В соответствии с требованиями IAB (Internet Architecure Board) и IESG (Internet Engineering Steering Group) первый документ посвещен анализу работы протокола BGP.

В последующих разделах документа описано, как BGP выполняет общие требования (General Requirements), заданные в разделе 3 (Section 3.0), а также требования к проектам стандартов, указанные в разделе 5 (Section 5.0) документа Internet Routing Protocol Standardization Criteria [1].

Отчет основан на работах Peter Lothberg (Ebone), Andrew Partan (Alternet) и др. Результаты работы были представлены на 25-й конференции IETF и доступны в трудах IETF.

Комментарии к этому документу шлите по адресу iwg@ans.net.

Благодарности

Протокол BGP был разработан группой IDR (прежнее название BGP) в составе IETF. Автор документа выражает свою глубокую признательность Якову Рехтеру (Yakov Rekhter) и Сю Харес (Sue Hares) – сопредседателям рабочей группы IDR. Автор также хочет выразить свою благодарность Якову Рехтеру и Тони Ли (Tony Li) за прочтение документа и конструктивные комментарии.

Документация

BGP представляет собой протокол маршрутизации между автономными системами (AS) в сетях TCP/IP. Первая версия протокола BGP была опубликована в RFC 1105. После этого были разработаны версии BGP с номерами 2, 3 и 4. Вторая версия протокола была описана в RFC 1163, третья – в RFC 1267. Отличия между версиями 1, 2 и 3 рассмотрены в приложении Appendix 2 работы [2]. Версия 4 сохраняет все функции, поддерживаемые младшими версиями протокола.

В BGP версии 2 была удалена концепция отношений up, down, horizontal между автономными системами, которая использовалась в версии 1, и введена концепция атрибутов пути. Кроме того, в BGP версии 2 были определены части протокола, которые в первой версии не были специфицированы ("under-specified").

В BGP версии 3 были изменены ограничения на использование атрибута пути NEXT_HOP и добавлено поле BGP Identifier в сообщения BGP OPEN. Кроме того, были более четко сформулированы условия распространения маршрутов BGP между узлами BGP одной AS.

В BGP версии 4 была переопределена (до этого основанная на классах) часть, связанная с доступностью обновлений на сетевом уровне для использования префиксов произвольной длины, позволяющих анонсировать множество бесклассовых сетей в одном элементе, как было рассмотрено в работе [5]. В BGP-4 также был изменен атрибут AS-PATH для обеспечения возможности описывать с помощью этого атрибута как наборы автономных систем, так и отдельные AS. Кроме того, в четвертой версии был заново описан атрибут INTER-AS METRIC как MULTI-EXIT DISCRIMINATOR и добавлены атрибуты LOCAL-PREFERENCE и AGGREGATOR.

Возможности использования BGP в сети Internet рассмотрены в работе [3].

Протокол BGP был разработан группой IDR Working Group в составе IETF. Эта группа поддерживает список рассылок iwg@ans.net для обсуждения возможностей протокола и особенностей его использования. Рабочая группа IDR регулярно собирается на квартальных конференциях IETF. Отчеты о результатах этих встреч публикуются в IETF Proceedings.

MIB

Информационная база MIB (Management Information Base) для протокола BGP-4 была опубликована в работе [4]. Авторами MIB являются Стив Вилс (Steve Willis - Wellfleet), Джон Берруз (John Burruss – Wellfleet) и Джон Чу (John Chu - IBM).

За исключением нескольких системных переменных BGP MIB делится на две таблицы: BGP Peer (партнеры) и BGP Received Path Attribute (атрибуты пути).

Таблица Peer содержит сведения о соединениях с партнерами BGP (состояние, активность и т.п.). В таблице Received Path Attribute содержатся атрибуты путей, полученные от всех партнеров без применения локальной политика маршрутизации. Реальные атрибуты, используемые для определения маршрутов, являются подмножеством таблицы полученных атрибутов.

Вопросы безопасности

BGP обеспечивает гибкий и расширяемый механизм поддержки аутентификации и обеспечения безопасности, позволяющий использовать схемы различной сложности. Во всех сеансах BGP выполняется аутентификация на основе значения BGP Identifier партнера. Кроме того, для всех сеансов BGP используется аутентификация на основе номеров AS, анонсируемых партнерами. Как часть механизма аутентификации протокол BGP позволяет передавать зашифрованные цифровые сигнатуры в каждом сообщении BGP. При отказе в аутентификации генерируется сообщение NOTIFICATION и соединение BGP немедленно разрывается.

Поскольку BGP работает с использованием протоколов TCP и IP, схема аутентификации BGP может использовать любые механизмы аутентификации, обеспечиваемые этими протоколами. Однако, использование протоколов TCP и IP ведет к тому, что все уязвимости этих протоколов (например, возможность атак на службы или систему аутентификации) проявляются и в BGP.

Реализации

В настоящее время существует множество независимых интероперабельных реализаций протокола BGP. В этом разделе дан краткий обзор реализаций, используемых в сети Internet. К ним относятся реализации протокола:

- cisco Systems
- консорциум gated
- 3COM
- Bay Networks (Wellfleet)
- Proteon

Для упрощения разработки реализаций BGP и предотвращения типовых ошибок опыт реализации BGP-4 компании Cisco описан в RFC 1656 [4]. При реализации протокола настоятельно рекомендуется следовать рекомендациям этого документа и приложений к работе [2].

Опыт реализации BGP-4 показывает, что протокол достаточно прост и для реализации основных возможностей протокола BGP-4 достаточно 2 человеко-месяцев (без учета затрат времени на реализацию поддержки CIDR).

Отметим, что в соответствии с требованиями IAB/IESG для проектов стандарта (Draft Standard) существует множество независимых и полностью интероперабельных реализаций протокола.

Опыт использования протокола

В этом разделе рассматривается опыт использования BGP и BGP-4 в реальных сетях.

Протокол BGP используется в реальной сетевой среде с 1989 г., а BGP-4 – с 1993. В такое использование вовлечены по крайней мере две из перечисленных выше реализаций. Использование BGP включает применение всех значимых возможностей протокола. Сетевая среда, в которой BGP используется как протокол маршрутизации между автономными системами, является гетерогенной. В терминах полосы каналов диапазон возможных значений простирается от 28 кбит/с до 150 мбит/с. Системы, на которых реализован протокол BGP также существенно различаются по своей производительности – от небольших компьютеров PC/RT до высокопроизводительных RISC-систем. Протокол используется как на специализированных маршрутизаторах, так и на рабочих станциях общего назначения, использующих ОС UNIX.

Существенно различаются и топологии систем, где применяется протокол – от малочисленных соединений (остовное дерево ICM) до систем с высокой плотностью соединений (опорная сеть NSFNET).

Во время подготовки этого документа протокол BGP-4 использовался для маршрутизации между всеми наиболее значимыми AS, включая сети Alternet, ANS, Ebone, ICM, IIJ, MCI, NSFNET, Sprint. Меньшая из известных магистралей включает единственный маршрутизатор, а наиболее крупная – почти 90 узлов BGP. В общей сложности известно несколько сот узлов BGP[1].

BGP используется для обмена маршрутной информацией между транзитными и тупиковыми AS, а также для передачи маршрутных данных между множеством транзитных AS.

В большинстве транзитных сетей BGP используется как основной (и единственный) протокол передачи информации о внешних маршрутах. Во время подготовки этого документа протокол BGP на некоторых сайтах использоваляся вместе с протоколами внутренней маршрутизации для передачи информации о внешних маршрутах.

Полный набор внешних путей, передаваемый BGP превышает 20 000 агрегированных маршрутов[2], которые описывают пути в подключенные сети.

Опыт использования протокола, описанный выше основан, прежде всего, на реализациях cisco и gated.

Конкретные детали использования BGP в сетях Alternet, ICM и Ebone были представлены на 25-й конференции IETF (Торонто, Канада) Питером Лотбергом (Peter Lothberg - Ebone), Эндрю Партаном (Andrew Partan - Alternet) и Полом Трэйна (Paul Traina - cisco).

Опыт использования BGP охватывает все основные функции протокола, включая аутентификацию, подавление маршрутных петель, а также новые функции BGP-4 (расширенная метрика и агрегирование маршрутов).

[1] С момента написания этого документа число узлов BGP в сети Internet многократно возросло. *Прим. перев.*
[2] В начале 2003 года размер полной таблицы BGP превысил 120 000 записей. *Прим. перев.*

Полоса, расходуемая на передачу трафика BGP измерялась в точке соединения между магистралями CA*Net и T1 NSFNET. Результаты этих измерений были представлены Деннисом Фергусоном (Dennis Ferguson) на 25-й конференции IETF и опубликованы в трудах IETF. Эти результаты явно показывают превосходство BGP по сравнению с протоколом EGP в части расхода полосы. Измерения на магистралях CA*Net (Dennis Ferguson) и T1 NSFNET (Susan Hares) также подтверждают преимущества BGP по сравнению с EGP по уровню загрузки CPU .

Переход к BGP версии 4

На многочисленных встречах ряда членов IETF неоднократно обсуждался вопрос о переходе к протоколам, поддерживающим бесклассовые сети (типа BGP-4).

Протокол BGP-4 вызывает большой интерес в части использования его в сети Internet, поскольку экспоненциальный рост размеров таблиц маршрутизации влечет за собой соответствующий рост запросов на память и производительность процессоров в устройствах, использующих протокол BGP. Благодаря этому, проблем, которые обычно сопровождают переход на новые протоколы, можно избежать при разумной политике развертывания систем BGP-4 сетевыми операторами.

Обсуждался вопрос о создании “маршрутных детонаторов” (route exploder), которые смогут регистрировать отдельные сети того или иного класса (class-based network) из блоков CIDR для узлов BGP-3, однако даже при беглом рассмотрении этого вопроса становится ясно, что такое решение потребует много памяти и сильно загрузит процессоры устройств BGP-3. Протокол BGP сам по себе не может отличить известные используемые сети от неиспользуемой части блоков CIDR.

Принятый большинством операторов путь перехода известен как “CIDR или смерть” ("CIDR, default, or die!").

Для проверки работы протокола BGP-4 была создана виртуальная “теневая” сеть Internet путем соединения сетей Alternet, Ebone, ICM и Cisco с помощью туннелей GRE. Эксперимент проводился с передачей реальной маршрутной информации через соединения BGP-3, обеспечивавшие нормальное функционирование участвующих в эксперименте сайтов. Такой подход позволил протестировать протокол BGP-4 до его развертывания в реальной сети.

После тестирования в теневой сети реализации BGP-4 были развернуты на реально используемом оборудовании этих сетей. Поддерживающие BGP-4 маршрутизаторы организовывали соединения BGP-4 между собой и взаимодействовали с другими маршрутизаторами по протоколу BGP-3. В сети включались некоторые тестовые агрегированные маршруты (в дополнение к сетям class-based) для обеспечения совместимости с узлами BGP-3.

После этого теневая сеть была переведена в разряд рабочих с организацией туннельных соединений с большинством основных транзитных операторов, чтобы операторы могли получить некоторое представление о воздействии агрегирования префиксов на картину маршрутизации.

После успешного развертывания BGP-4 множество сайтов отказалось от анонсирования на основе классов и перешло на использование CIDR-анонсов. Это подтолкнуло транзитных операторов, которые еще не перешли к новой форме анонсов, поскольку приводило к потере маршрутов в некоторые популярные сети.

Метрика

BGP-4 переопределяет старую метрику INTER-AS как MULTI-EXIT-DISCRIMINATOR. Это значение может использоваться в процессах разрыва соединений (tie breaking) при выборе предпочтительного пути к заданному блоку адресов. Значение MED используется только в тех случаях, когда сравниваются пути, полученные от разных внешних партнеров в одной AS для индикации предпочтений анонсирующей маршруты AS.

Назначением MED является определение наилучшего пути на заключительном этапе. Рабочая группа BGP считала, что любая метрика, заданная удаленным оператором будет воздействовать на маршрутизацию в локальной AS только в тех случаях, когда отсутствуют иные предпочтения. Первоочередной задачей MED было обеспечение гарантий того, что партнеры не будут «сливать» или «поглощать» трафик сетей, которые они анонсируют .

Атрибут LOCAL-PREFERENCE был добавлен для того, чтобы локальный оператор мог легко настроить правила, переопределяющие стандартный механизм выбора лучшего маршрута без настройки локальных предпочтений на других маршрутизаторах.

Одним из недостатков спецификации BGP4 было предложение использовать принятое по умолчанию значение LOCAL-PREF, если этот параметр не был задан. Использование 0 или максимального значения имеют свои недостатки, поскольку основной смысл принятого по умолчанию значения заключается в обеспечение взаимодействия маршрутизаторов разных производителей в рамках одной AS (поскольку LOCAL-PREF задается локальным администратором, при пересечении границы AS проблем совместимости не возникает).

Другрй требующей исследования областью является метод влияния исходной (originating) AS на процесс выбора лучшего пути. Например, сайт с двойным подключением может выбрать одну AS в качестве основного транзита, а другую использовать в качестве резерва.

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

В топологии, где два транзитных провайдера обеспечивают подключение к третьему провайдеру, реальное решение принимает этот провайдер и не существует механизма индикации предпочтений.

Предложение общего вида заключалось в передаче дополнительного вектора, соответствующего AS-PATH, где кадая транзитная AS могла указывать уровень предпочтения для данного маршрута. Взаимодействующие AS могли бы управлять трафиком на базе сравнения представляющих для них интерес фрагментов вектора в соответствии со своей политикой маршрутизации.

Хотя защита политики маршрутизации AS имеет важное значение, исключение многочисленных «ручных» настроек при выборе маршрутов представляется более важным для протоколов типа BGP.

Использование BGP внутри больших AS

Хотя это и не относится в полной мере к протоколу, у многих операторов возникает вопрос поддержки автономных систем с большим числом узлов (peer). Каждый узел, имеющий партнерские отношения с внешним маршрутизатором, отвечает за распространение информации о доступности и путях всем другим транзитным и граничным маршрутизаторам данной AS. Обычно это реализуется путем организации внутренних соединений BGP со всеми транзитными и граничными маршрутизаторами локальной AS.

По мере роста AS число таких соединений TCP будет расти по закону n^2 и потребуется тот или иной метод настройки и поддержки таких соединений.. Протокол BGP не задает способ распространения такой информации, поэтому были предложены дополнения типа вставки атрибутов BGP в локальные протоколы IGP. Прилагались также усилия по разработке маршрутных рефлекторов BGP (route reflectors) или надежной транспортировки информации IBGP с использованием групповой адресации, что позволило бы снизить издержки по настройке, а также требования к памяти и ресурсам CPU на передачу данных всем внутренним партнерам BGP.

Динамика Internet

Как обсуждалось в работе [7], движущей силой порта производительности CPU и расширения полосы пропускания служит динамическая природа маршрутизации в Internet. По мере расширения сети растет и число изменений маршрутизации в единицу времени. Некоторый уровень снижения числа изменений возникает автоматически за счет агрегирования NLRI в более крупные блоки, однако этого не достаточно. В Приложении 6 к работе [2] описаны методы подавления, которые могут быть применены для анонсов. В будущих спецификациях протоколов, подобных BGP, методы подавления осцилляций следует рассматривать, как обязательные для соответствующих протоколу реализаций.

Подтверждения

Протокол BGP-4 был разработан рабочей группой IETF IDR/BGP. Автор выражает свою признательность Yakov Rekhter за подготовку документа RFC 1266, а также хочет явно отметить Yakov Rekhter и Tony Li за рецензирование данного документа, а также конструктивную критику и ценные замечания.

Адрес автора

Paul Traina

Cisco Systems, Inc.

170 W. Tasman Dr.

San Jose, CA 95134

EMail: pst@cisco.com

Литература

[1] Hinden, R., "Internet Routing Protocol Standardization Criteria", RFC 1264, BBN, October 1991.

[2] Rekhter, Y., and T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771[1], T.J. Watson Research Center, IBM Corp., cisco Systems, March 1995.

[3] Rekhter, Y., and P. Gross, Editors, "Application of the Border Gateway Protocol in the Internet", RFC 1772[2], T.J. Watson Research Center, IBM Corp., MCI, March 1995.

[4] Willis, S., Burruss, J., and J. Chu, "Definitions of Managed Objects for the Fourth Version of the Border Gateway Protocol (BGP-4) using SMIv2", RFC 1657, Wellfleet Communications Inc., IBM Corp., July 1994.

[5] Fuller V., Li. T., Yu J., and K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, BARRNet, cisco, MERIT, OARnet, September 1993.

[6] Traina P., "BGP-4 Protocol Document Roadmap and Implementation Experience", RFC 1656, cisco Systems, July 1994.

[7] Traina P., "BGP Version 4 Protocol Analysis", RFC 1774, cisco Systems, March 1995.

Перевод на русский язык

Николай Малых

nmalykh@gmail.com

[1] Этот вариант спецификации устарел и заменен RFC 4271. Перевод имеется на сайте http://www.protocols.ru. *Прим. перев.*

[2] На сайте www.protocols.ru можно найти перевод этого документа на русский язык. *Прим. перев.*