Network Working Group
Request for Comments: 1265

Y. Rekhter, Editor
T.J. Watson Research Center, IBM Corp.

October 1991

Анализ протокола BGP

BGP Protocol Analysis

1. Статус документа

В этом документе содержится информация для сообщества Internet. Документ не задает каких-либо стандартов Internet. Допускается свободное распространение документа.

2. Введение

Целью этого отчета является документирование того, как требования для публикации протоколов маршрутизации в качестве Draft Standard были удовлетворены для протокола BGP[1]. Этот отчет резюмирует основные свойства BGP и анализирует протокол с точки зрения масштабирования и производительности. Это первый из 2 отчетов, посвященных протоколу BGP.

BGP представляет собой протокол маршрутизации между автономными системами, созданный для сетей TCP/IP. Версия 1 протокола BGP была опубликована в RFC 1105. После этого были разработаны версии BGP 2 и 3. Версия 2 была опубликована в 1163. Версия 3 опубликована в [1]. Различия между версиями 1, 2 и 3 описаны в Приложении 3 документа [1].

Возможности применения BGP в сети Internet описаны в [2].

Комментарии к документу направляйте по адресу iwg@rice.edu.

3. Благодарности

Протокол BGP был создан рабочей группой IETF[2] IWG/BGP. Мы рады выразить свою глубочайшую признательность Guy Almes (Rice University), который был предыдущим руководителем рабочей группы IWG. Мы также хотим отдельно поблагодарить Bob Braden (ISI) и Bob Hinden (BBN) за просмотр документа и полезные, конструктивные замечания.

4. Основные свойства и алгоритмы протокола BGP

В этой главе кратко рассматриваются ключевые функции и алгоритмы протокола BGP. BGP представляет собой протокол маршрутизации между автономными системами (AC); он разработан для использования в среде, включающей множество AC. BGP предполагает, что маршрутизация внутри автономных систем осуществляется с помощью протоколов внутридоменной маршрутизации. BGP не делает предположений о протоколах внутридоменной маршрутизации, используемых в различных AC. В частности, BGP не требует от всех автономных систем использования одного протокола внутридоменной маршрутизации.

BGP является реальным протоколом междоменной маршрутизации. Он не накладывает ограничений на топологию соединений Internet. Информации, передаваемой в рамках BGP, достаточно для построения графа связности автономных систем, из которого могут быть исключены маршрутные петли, а также могут быть выполнены некоторые правила маршрутизации на уровне автономной системы.

Ключевым свойством протокола является представление (нотация) атрибутов пути (Path Attribute). Это свойство обеспечивает гибкость и расширяемость BGP. Атрибуты пути делятся на общепринятые (well-known) и дополнительные (optional). Использование дополнительных атрибутов обеспечивает возможность проведения экспериментов, которые могут затрагивать группу маршрутизаторов BGP, не влияя на остальную часть Internet. Новые дополнительные атрибуты могут добавляться практически так же, как добавляются опции для протоколов (например, Telnet). Одним из наиболее важных атрибутов пути является AS-PATH[3]. По мере перемещения информации о доступности через Internet к этим данным добавляется список автономных систем, через которые проходит информация. В результате такого добавления формируется атрибут AS-PATH. Использование AS-PATH обеспечивает простой способ удаления петель в маршрутной информации. В дополнение к этому AS-PATH является мощным и универсальным механизмом маршрутизации на основе правил.

Алгоритм, используемый BGP, не является в чистом виде ни алгоритмом на базе «векторов удаления» (distance vector), ни алгоритмом на основе состояний каналов (link state). Передача полного пути через AS в атрибуте AS-PATH позволяет реконструировать крупные части полной топологии. Это похоже на работу алгоритмов на основе состояния каналов (link state). Обмен между партнерами лишь используемыми в данный момент маршрутами делает этот алгоритм похожим на алгоритмы distance vector.

В целях снижения расхода полосы и ресурсов процессора BGP использует нарастающие обновления, при которых после начального обмена полной маршрутной информацией пара маршрутизаторов BGP обменивается только данными об изменении этой информации. Метод нарастающих обновлений требует использования между партнерами BGP транспорта с гарантией доставки. Для этого BGP использует в качестве транспорта протокол TCP.

---

[1] Border Gateway Protocol - протокол граничного шлюза.
[2] Internet Engineering Task Force.
[3] Autonomous System Path - путь через AC. *Прим. перев.*

BGP является самодостаточным протоколом, т. е., BGP задает обмен маршрутной информации как между узлами BGP в разных автономных системах, так и между узлами BGP одной АС.

Для обеспечения возможности сосуществования с EGP протокол BGP поддерживает передачу полученных от EGP внешних маршрутов. BGP также позволяет передавать статически заданные внешние маршруты.

5. Производительность и масштабируемость BGP

В этой главе рассматриваются вопросы загрузки полосы каналов, памяти маршрутизатора и ресурсов процессора для работы BGP в нормальных условиях. Рассматриваются также вопросы масштабирования BGP и ограничения этого протокола.

BGP не требует, чтобы все маршрутизаторы автономной системы принимали участие в работе протокола BGP. Только граничные маршрутизаторы, которые обеспечивают соединения между локальной автономной системой и смежными автономными системами, принимают участие в работе BGP. Ограничение числа участников является лишь одним из способов решения задачи обеспечения масштабируемости.

5.1 Полоса канала и загрузка процессора

Непосредственно после организации соединения BGP партнеры обмениваются полными наборами маршрутной информации. Если обозначить общее число маршрутов в Internet N, среднюю дистанцию в Internet (среднее число AS на пути) - M, общее число автономных систем в Internet - A и предположить, что сети равномерно распределены между автономными системами, максимальная полоса, расходуемая на первоначальный обмен данными между парой партнеров BGP (P) составит

```
O(N + M * A)
```

(при условии, что реализация поддерживает множество сетей на одно сообщение, как указано в Приложении 5 к документу [1]). Эта информация по порядку величины совпадает с числом сетей, доступных через каждого партнера (см. также параграф 5.2).

Приведенная ниже таблица иллюстрирует типовой расход полосы[1] при начальном обмене между парой узлов BGP на основе приведенных выше оценок (полоса на передачу заголовков BGP не учитывалась).

Число сетей	*Среднее число AS на пути*	*Число AS*	*Полоса*
2100	5	59	9000 байт
4000	10	100	18000 байт
10000	15	300	49000 байт
100000	20	3000	520000 байт

Отметим, что основной расход полосы связан с обменом информацией о доступности сетей (Network Reachability Information).

После завершения начального обмена расход полосы и машинных тактов процессора для работы BGP зависит только от стабильности Internet. Если Internet сохраняет стабильность, то полоса и ресурсы процессора потребляются протоколом BGP лишь на обмен сообщениями BGP KEEPALIVE, которые передаются только между партнерами. Предлагаемая частота составляет 1 сообщение в течение 30 секунд. Сообщения KEEPALIVE достаточно малы (19 октетов) и практически не требуют обработки. В результате расход полосы на передачу сообщений KEEPALIVE составляет около 5 бит/сек. Практика показывает, что связанная с этими сообщениями дополнительная нагрузка (в терминах расхода полосы и процессорного времени) пренебрежимо мала. Если состояние Internet нестабильно, между маршрутизаторами передается только информация об изменениях состояния доступности (которые вызваны нестабильностью), путем обмена сообщениями UPDATE. Если обозначить число обновлений маршрутов в секунду C, то максимальный расход полосы для BGP можно выразить, как O(C*M). Наибольшие издержки при передаче сообщений UPDATE возникают в тех случаях, когда каждое сообщение UPDATE содержит информацию для единственной сети. Следует отметить, что на практике изменения маршрутов локализованы относительно атрибутов пути (т. е., изменяемые маршруты имеют общие атрибуты пути). В таких случаях информацию для множества сетей можно передать в одном сообщении UPDATE, что приведет к существенному снижению расхода полосы (см. Приложение 5 к документу [1]).

Поскольку в установившемся состоянии потребляемая протоколом BGP полоса и ресурсы процессора зависят только от стабильности Internet и совершенно не зависят от количества сетей, составляющих Internet, у протокола BGP не должно возникать проблем масштабирования, связанных с расходом полосы и ресурсов процессора, по мере роста Internet, поскольку общая стабильность связности между АС в Internet может контролироваться. Вопросы стабильности могут решаться введением той или иной формы подавления (например, удержания каналов в неактивном состоянии). По природе BGP такое подавление следует рассматривать, как локальный вопрос АС (см. Приложение 5 в [1]). Отметим, что независимо от BGP не следует недооценивать значимость стабильности в Internet. Рост сети Internet делает вопрос стабильности одним из наиболее критичных для сети в целом. Важно, что протокол BGP, сам по себе, не вносит дополнительной нестабильности в Internet. Текущие наблюдения в сети NSFNET показывают, что основной причиной нестабильности является некорректная маршрутизация внутри автономных систем, образующих Internet. Следовательно, несмотря на обеспечение протокола BGP механизмами стабилизации, следует обратить пристальное внимание на корень проблемы и обеспечить стабильность маршрутизации внутри АС.

Полезно также сравнить требования к полосе и процессорным ресурсам для протоколов BGP и EGP. Обмен полной информацией в BGP осуществляется только при организации соединения, а в EGP такой обмен выполняется периодически (обычно каждые 3 минуты). Отметим, что для обоих протоколов объем передаваемой при обмене информации определяется числом сетей, доступных через передающий информацию узел (см. также параграф 5.2). Следовательно, даже при максимальной нестабильности BGP его поведение в худшем случае будет аналогично поведению протокола EGP в установившемся состоянии.

Опыт работы с BGP показывает, что модель нарастающих обновлений, реализованная в BGP, обеспечивает значительное повышение эффективности в части расхода полосы и загрузки процессора по сравнению с обменом

[1] Корректней будет сказать «оценка объема передаваемых данных». *Прим. перев.*

полной информацией в EGP (см. также презентацию Dennis Ferguson на конференции Twentieth IETF, March 11-15, 1991, St.Louis).

5.2 Требования к памяти

Максимальный расход памяти для BGP, определяемый общим числом сетей в Internet (N), средней AC-дистанцией[1] в Internet (M), общим числом AC в Internet (A) и общим числом узлов BGP, с которыми данная система имеет партнерские отношения, (K[2]). Максимальные требования к памяти (MR) можно описать выражением

```
MR = O((N + M * A) * K)
```

В современной[3] опорной сети NSFNET (N = 2110, A = 59, and M = 5), если каждая сеть будет сохраняться в 4 октетах, а каждая AC - в 2 октетах, дополнительный расход памяти для хранения информации о путях AC (сверх расхода на хранение информации о внешних маршрутах) составит менее 7 от общего расхода памяти.

Интересно отметить, что до введения протокола BGP в опорной сети NSFNET требования к памяти на магистральных маршрутизаторах NSFNET, использующих протокол EGP, имели порядок O(N*K). Следовательно, дополнительный расход памяти на маршрутизаторах NSFNET после внедрения BGP составил менее 7 процентов.

Поскольку средняя AC-дистанция растет значительно медленней увеличения общего числа сетей (в опорной сети NSFNET имеется около 60 AC, а число известных магистральным маршрутизаторам сетей превышает 2000, следовательно, средняя AC-дистанция в современной сети Internet будет меньше 5), для практических целей максимальный расход памяти на маршрутизаторах будет иметь порядок значения общего числа сетей в Internet, умноженного на число партнеров локальной системы. Мы предполагаем, что общее число сетей в Internet будет расти значительно быстрей, нежели среднее число партнеров каждого маршрутизатора. Следовательно, масштабирование в плане расхода памяти будет определяться прежде всего факторами, пропорциональными общему числу сетей в Internet.

Приведенная ниже таблица иллюстрирует типовые требования к памяти для маршрутизаторов BGP. Предполагается, что для каждой сети расходуется 4 байта, для каждой AC - 2 байта и каждая сеть доступна через некоторую часть имеющихся партнеров (число партнеров BGP на сеть).

Число сетей	*Средняя AC-дистанция*	*Число AC*	*Число партнеров BGP на сеть*	*Расход памяти*
2100	5	59	3	27000 байтов
4000	10	100	6	108000 байтов
10000	15	300	10	490000 байтов
100000	20	3000	20	1040000 байтов

Для корректной оценки требований BGP к памяти в перспективе попробуем на минутку забыть об информации, используемой лоя построения таблиц пересылки в маршрутизаторах и сосредоточиться на самих таблицах. В этом случае резонно поинтересоваться предельным размером этих таблиц. С учетом того, что сейчас в таблицах пересылки магистральных маршрутизаторов NSFNET содержится более 2000 записей, возникает вопрос - сможет ли маршрутизатор работать с таблицей, включающей 20000 записей. Очевидно, что ответ на этот вопрос никак не зависит от BGP. С другой стороны, ответы на исходные вопросы (применительно к BGP) напрямую связаны с ответом на этот вопрос. Очень интересный комментарий дал Paul Tsuchiya (член группы по обзору BGP, назначенной Bob Hinden) в своем обзоре BGP в марте 1990. От сказал: «BGP не обеспечивает требуемого масштабирования. Реально это не является проблемой BGP. Проблема заключается в плоском пространстве адресов IP. В плоском адресном пространстве IP любой протокол маршрутизации должен передавать номера сетей в своих обновлениях.» Ограничения BGP в части требований к памяти напрямую связаны с лежащим в основе протоколом Internet (IP) и, в частности, с реализованной в IP схемой адресации. BGP может масштабироваться значительно лучше в средах с более гибкими схемами адресации. Следует отметить, что путем незначительных добавлений протокол BGP можно расширить для поддержки иерархий автономных систем. Такие иерархии в комбинации со схемой адресации, поддерживающий более гибкие возможности агрегирования, могут применяться в BGP, обеспечивая практически неограниченные возможности масштабирования этого протокола.

6. Применимость BGP

В этом разделе мы пытаемся ответить на вопрос, для каких сред BGP подходит хорошо, а для каких не подходит. Частично ответ на этот вопрос был дан в разделе 2 документа [1], где сказано следующее:

> «Для характеризации набора правил (политики), который можно реализовать за счет использования BGP, нужно сосредоточить внимание на правиле, в соответствии с которым AC анонсирует в соседние AC только те маршруты, которые она использует сама. Это правило отражает парадигму «поэтапной[4]» маршрутизации, повсеместно используемой в современной сети Internet. Отметим, что некоторые правила не могут поддерживаться парадигмой поэтапной маршрутизации и это требует использования таких методов, как задаваемая отправителем маршрутизация[5]. Например, BGP не позволяет AC передать в соседнюю AC трафик так, чтобы его маршрут отличался от маршрута трафика, происходящего из этой соседней AC. С другой стороны, BGP может поддерживать любую политику, соответствующую парадигме поэтапной маршрутизации. Поскольку в современной сети Internet используется только парадигма поэтапной маршрутизации, а BGP может поддерживать любые правила, соответствующие этой парадигме, протокол BGP хорошо подходит в качестве протокола маршрутизации между AC в современной сети Internet.»

Протокол BGP не только хорошо подходит для современной сети Internet, но и является практической необходимостью в Internet. Опыт эксплуатации EGP показывает, что этот протокол не отражает потребности современной сети Internet. Топологические ограничения EGP не оправданы технически и невыполнимы с практической точки зрения. Неспособность EGP к эффективному обслуживанию информационного обмена между партнёрами является причиной серьёзных маршрутных неустойчивостей в сети Internet. Кроме того, обеспечиваемая BGP информация хорошо подходит для реализации различных вариантов политики маршрутизации.

---

[1] Средняя дистанция на уровне AC определяется числом автономных систем на пути.
[2] Отметим, что большую часть K обычно будут составлять узлы BGP в своей AC.
[3] 1991 год. *Прим. перев.*
[4] Hop-by-hop.
[5] Source routing.

Вместо попыток предсказания будущего и перегрузки BGP множеством функций, которые могут (или не могут) потребоваться, разработчики BGP выбрали иной подход. Протокол включает только необходимый минимум функций, обеспечивая в то же время гибкие механизмы расширения функциональности. Поскольку BGP разрабатывался с учетом гибкости и масштабируемости, мы полагаем, что этот протокол сравнительно легко сможет соответствовать новым и меняющимся потребностям. Доказательством этого утверждения может служить способ, который уже используется для добавления в протокол новых функций (например, «ремонт» разделенных АС в BGP).

В заключение отметим, что BGP хорошо подходит в качестве протокола маршрутизации между АС для современной сети Internet, основанной на протоколе IP (RFC 791) и парадигме поэтапной маршрутизации. Трудно предположить, подойдет ли BGP для межсетевых сред, работающих на основе протоколов, отличных от IP, а также в случае использования иной парадигмы маршрутизации.

Литература

[1] Lougheed, K., and Y. Rekhter, "A Border Gateway Protocol 3 (BGP-3)", RFC 1267[1], cisco Systems, T.J. Watson Research Center, IBM Corp., October 1991.

[2] Rekhter, Y., and P. Gross, Editors, "Application of the Border Gateway Protocol in the Internet", RFC 1268, T.J. Watson Research Center, IBM Corp., ANS, October 1991.

Вопросы безопасности

Вопросы безопасности не рассматриваются в этом документе.

Адрес автора

Yakov Rekhter

T.J. Watson Research Center IBM Corporation

P.O. Box 218

Yorktown Heights, NY 10598

Phone: (914) 945-3896

EMail: yakov@watson.ibm.com

IETF BGP WG mailing list: iwg@rice.edu. Для включения в список отправьте письмо по адресу: iwg-request@rice.edu

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru

[1] Этот протокол в настоящее время уже стал «достоянием истории». *Прим. перев.*