**BGP**

**BGP (Border Gateway Protocol)** — это основной протокол динамической маршрутизации, который используется в Интернете.

Маршрутизаторы, использующие протокол BGP, обмениваются информацией о доступности сетей. Вместе с информацией о сетях передаются различные атрибуты этих сетей, с помощью которых BGP выбирает лучший маршрут и настраиваются политики маршрутизации.

Один из основных атрибутов, который передается с информацией о маршруте — это список автономных систем, через которые прошла эта информация. Эта информация позволяет BGP определять где находится сеть относительно автономных систем, исключать петли маршрутизации, а также может быть использована при настройке политик.

Маршрутизация осуществляется пошагово от одной автономной системы к другой. Все политики BGP настраиваются, в основном, по отношению к внешним/соседним автономным системам. То есть, описываются правила взаимодействия с ними.

Так как BGP оперирует большими объемами данных, то принципы его настройки и работы отличаются от внутренних протоколов динамической маршрутизации (IGP).

**Терминология протокола**

**Внутренний протокол маршрутизации (interior gateway protocol)** – протокол, который используется для передачи информации о маршрутах внутри автономной системы.

**Внешний протокол маршрутизации (exterior gateway protocol)** – протокол, который используется для передачи информации о маршрутах между автономными системами.

**Автономная система (autonomous system, AS)** — набор маршрутизаторов, имеющих единые правила маршрутизации, управляемых одной технической администрацией и работающих на одном из протоколов IGP (для внутренней маршрутизации AS может использовать и несколько IGP).

**Транзитная автономная система (transit AS)** — автономная система, через которую передается трафик других автономных систем.

**Путь (path)** — последовательность, состоящая из номеров автономных систем через которые нужно пройти для достижения сети назначения.

**Атрибуты пути (path attributes, PA)** — характеристики пути, которые позволяют выбрать лучший путь.

**BGP speaker** — маршрутизатор, на котором работает протокол BGP.

**Соседи (neighbor, peer)** — любые два маршрутизатора, между которыми открыто TCP-соединение для обмена информацией о маршрутизации.

Информация сетевого уровня о доступности сети (Network Layer Reachability Information, **NLRI**) — IP-префикс и длина префикса.

**Описание протокола**

BGP выбирает лучшие маршруты не на основании технических характеристик пути (пропускной способности, задержки и т.п.), а на основании политик. При выборе между каналами двух провайдеров, зачастую имеет значение не то, у какого канала лучше технические характеристики, а какие-то внутренние правила компании. Например, использование какого канала обходится компании дешевле. Поэтому в BGP выбор лучшего маршрута осуществляется на основании политик, которые настраиваются с использованием фильтров, анонсирования маршрутов, и изменения атрибутов.

Как и другие протоколы динамической маршрутизации, BGP может передавать трафик только на основании IP-адреса получателя. Это значит, что с помощью BGP нет возможности настроить правила маршрутизации, в которых будет учитываться, например, то, из какой сети был отправлен пакет или данные какого приложения передаются. Если принимать решение о том, как должен маршрутизироваться пакет, необходимо по каким-то дополнительным критериям, кроме адреса получателя, необходимо использовать механизм **policy-based routing (PBR)**.

**Основные характеристики протокола**

BGP это path-vector протокол с такими общими характеристиками:

* Использует TCP для передачи данных, это обеспечивает надежную доставку обновлений протокола (порт 179)
* Отправляет обновления только после изменений в сети (нет периодических обновлений)
* Периодически отправляет keepalive-сообщения для проверки TCP-соединения
* Метрика протокола называется path vector или атрибуты (attributes)

**Автономная система (autonomous system, AS)** — это система IP-сетей и маршрутизаторов, управляемых одним или несколькими операторами, имеющими единую, четко определенную политику маршрутизации с Интернетом.

Использование:

* 0 и 65535 (зарезервированы)
* 1-64495 (публичные номера)
* 65552-4294967295 (публичные номера)
* 64496-64511 – для использования в примерах и документации
* 64512-65534 (приватные номера)
* 23456 (представляет 32-битный диапазон на устройствах, которые работают с 16-битным диапазоном)

**PI и PA**

Когда вы подключаетесь к провайдеру, он выдаёт вам диапазон публичных адресов – так называемые **PA-адреса (Provider Aggregatable**).

PI – **Provider Independent–**

**Внутренний BGP (Internal BGP) и Внешний BGP (External BGP)**

**Внутренний BGP (Internal BGP, iBGP)** — BGP работающий внутри автономной системы. iBGP-соседи не обязательно должны быть непосредственно соединены.

**Внешний BGP (External BGP, eBGP)** — BGP работающий между автономными системами. По умолчанию, eBGP-соседи должны быть непосредственно соединены.

Если iBGP-маршрутизаторы работают в нетранзитной AS, то соединение между ними должно быть full mesh. Это следствие принципов работы протокола — если маршрутизатор, находящийся на границе AS, получил обновление, то он передает его всем соседям; соседи, которые находятся внутри автономной системы, больше это обновление не распространяют, так как считают, что все соседи внутри AS уже его получили.

**Таймеры протокола**

**Keepalive Interval** — Интервал времени в секундах, между отправкой сообщений keepalive. По умолчанию 60 секунд.

**Hold Time** — Интервал времени в секундах, по истечении которого сосед будет считаться недоступным. По умолчанию 180 секунд.

**Типы сообщений BGP**

**Open** — используется для установки отношений соседства и обмена базовыми параметрами. Отправляется сразу после установки TCP-соединения.

Кроме стандартного заголовка пакета BGP, в сообщении Open такие поля:

* **Version** — версия протокола BGP
* **My Autonomous System** — номер автономной системы отправителя
* **Hold Time** — максимальное время в секундах, которое может пройти между получением Keepalive и сообщением Update. Время выбирается минимальным
* **BGP Identifier** — играет роль в выборе пути пересылки BGP-сообщений при наличии более одного канала связи между BGP-соседями
* **Optional Parameters Length** — если равен 0, то в маркер записываются единицы, а Optional Parameters имеет нулевую длину; если не равен 0, то в Optional Parameters записываются данные для определения кода, который указывается в маркере.
* **Optional Parameters** — играет роль в формировании и последующем определении кода в поле маркер.

**Update** — используется для обмена информацией маршрутизации.

**Notification** — используется когда возникают ошибки BGP. После отправки сообщения сессия с соседом разрывается.

**Keepalive** — используется для поддерживания отношений соседства, для обнаружения неактивных соседей. Сообщения Keepalive состоят только из заголовка пакета (длина 19 октетов). Если периодичность отправки keepalive-сообщений выставлена в 0, то сообщения не отправляются.

**Отношения соседства**

Для того чтобы установить отношения соседства, в BGP надо настроить вручную каждого соседа. Когда указывается сосед локального маршрутизатора, обязательно указывается автономная система соседа. По этой информации BGP определяет тип соседа:

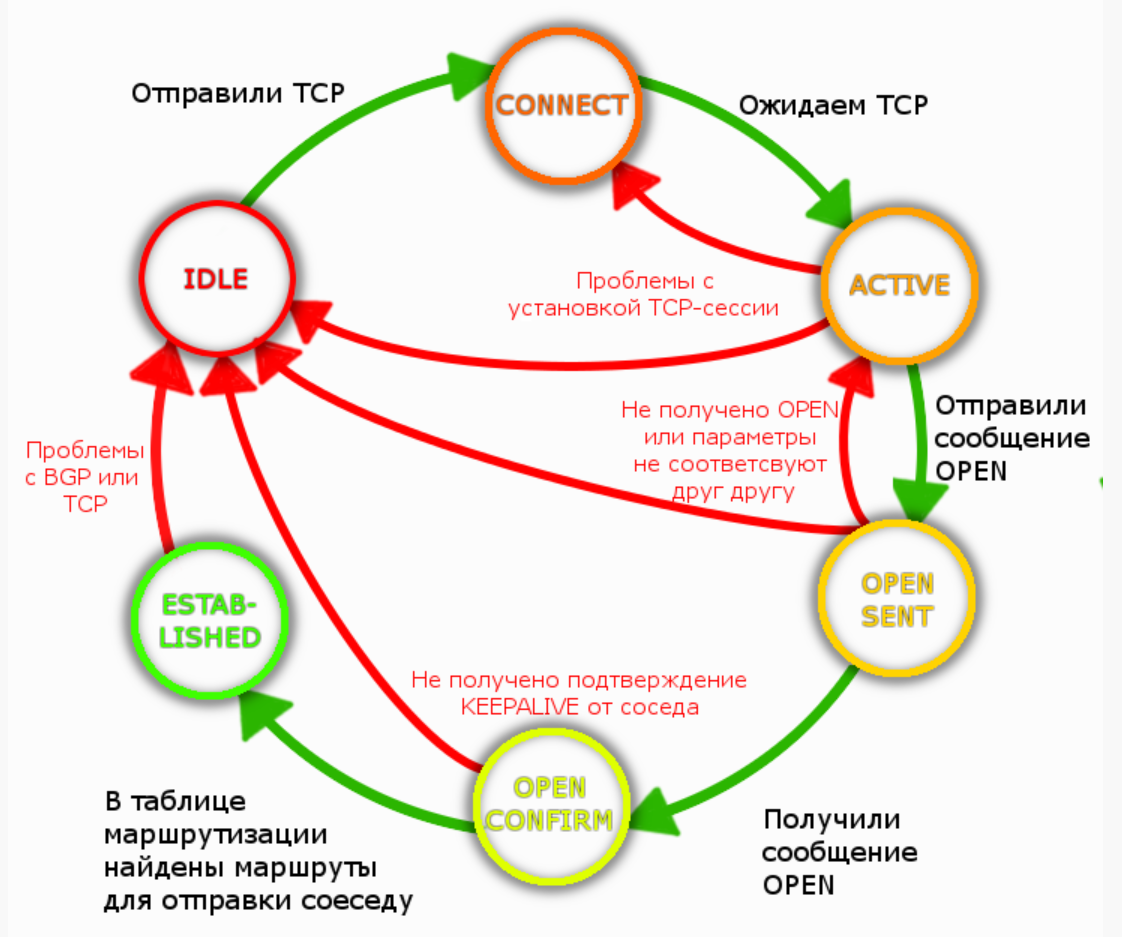
* Внутренний BGP сосед (**iBGP-сосед**) — сосед, который находится в той же автономной системе, что и локальный маршрутизатор. iBGP-соседи не обязательно должны быть непосредственно соединены.
* Внешний BGP сосед (**eBGP-сосед**) — сосед, который находится в автономной системе отличной от локального маршрутизатора. По умолчанию, eBGP-соседи должны быть непосредственно соединены.

Тип соседа мало влияет на установку отношений соседства. Более существенные отличия между различными типами соседей проявляются в процессе отправки обновлений BGP и добавлении маршрутов в таблицу маршрутизации.

BGP выполняет такие проверки, когда формирует отношения соседства:

* Маршрутизатор должен получить запрос на TCP-соединение с адресом отправителя, который маршрутизатор найдет указанным в списке соседей (команда neighbor).
* Номер автономной системы локального маршрутизатора должен совпадать с номером автономной системы, который указан на соседнем маршрутизаторе командой neighbor remote-as (это требование не соблюдается при настройках конфедераций).
* Идентификаторы маршрутизаторов (Router ID) не должны совпадать.
* Если настроена аутентификация, то соседи должны пройти её.

**Состояния связи с соседями**

**Full View и Default Route**

**Full View** – маршрутизатор изучает абсолютно все маршруты Интернета

**Default Route** - случае от каждого провайдера приходит только один маршрут по умолчанию

**Атрибуты пути (path attributes)**

Атрибуты пути разделены на 4 категории:

* **Well-known mandatory** — все маршрутизаторы, работающие по протоколу BGP, должны распознавать эти атрибуты. Должны присутствовать во всех обновлениях (update).
* Autonomous system path
* Next-hop
* Origin
* **Well-known discretionary** — все маршрутизаторы, работающие по протоколу BGP, должны распознавать эти атрибуты. Могут присутствовать в обновлениях (update), но их присутствие не обязательно.
  + Local preference
  + Atomic aggregate
* **Optional transitive** — могут не распознаваться всеми реализациями BGP. Если маршрутизатор не распознал атрибут, он помечает обновление как частичное (partial) и отправляет его дальше соседям, сохраняя не распознанный атрибут.
  + Aggregator
  + Communities
* **Optional non-transitive** — могут не распознаваться всеми реализациями BGP. Если маршрутизатор не распознал атрибут, то атрибут игнорируется и при передаче соседям отбрасывается.
  + Multi-exit discriminator (MED)
  + Originator ID
  + Cluster list

**Autonomous system path**

* Описывает через какие автономные системы надо пройти, чтобы дойти до сети назначения.
* Номер AS добавляется при передаче обновления из одной AS eBGP-соседу в другой AS.

Используется для:

1. Предотвращение петель маршрутизации. В AS-Path не должно быть повторяющихся номеров
2. Выбор наилучшего маршрута. Чем короче AS-Path, тем предпочтительнее маршрут

**Next-hop**

* IP-адрес следующей AS для достижения сети назначения.
* Это IP-адрес eBGP-маршрутизатора, через который идет путь к сети назначения.
* Атрибут меняется при передаче префикса в другую AS

**Origin**

Атрибут Origin — указывает на то, каким образом был получен маршрут в обновлении.

Возможные значения атрибута:

**0 — IGP**: NLRI получена внутри исходной автономной системы;

**1 — EGP**: NLRI выучена по протоколу Exterior Gateway Protocol (EGP). Предшественник BGP, не используется

**2 — Incomplete**: NLRI была выучена каким-то другим образом

**Local preference**

* Указывает маршрутизаторам внутри автономной системы как выйти за её пределы.
* Этот атрибут передается только в пределах одной автономной системы.
* На маршрутизаторах Cisco по умолчанию значение атрибута — 100.
* Выбирается та точка выхода у которой значение атрибута больше.
* Если eBGP-сосед получает обновление с выставленным значением local preference, он игнорирует этот атрибут.

**Atomic aggregate**

Метка, указывающая, что NLRI является summary.

**Aggregator**

Список RID и ASN маршрутизаторов, создавших summary NLRI.

**Communities**

* Тегирование маршрутов
* Существуют предопределенные значения
* По умолчанию не пересылаются соседям
* Один из вариантов применения: передается соседней AS для управления входящим трафиком

Как правило community отображаются в формате ASN:VALUE. В таком формате, доступны для использования community от 1:0 до 65534:65535. В первой части указывается номер автономной системы, а во второй значение community, которое определяет политику маршрутизации трафика.

Некоторые значения communities предопределены. RFC1997 определяет три значения таких community. Эти значения должны одинаково распознаваться и обрабатываться всеми реализациями BGP, которые распознают атрибут community.

Если маршрутизатор получает маршрут в котором указано предопределенное значение communities, то он выполняет специфическое, предопределенное действие основанное на значении атрибута.

Предопределенные значения communities (Well-known Communities):

* **no-export (0xFFFFFF01)** — Все маршруты которые передаются с таким значением атрибута community не должны анонсироваться за пределы конфедерации (автономная система, которая не является частью конфедерации считается конфедерацией). То есть, маршруты не анонсируются EBGP-соседям, но анонсируются внешним соседям в конфедерации,
* **no-advertise (0xFFFFFF02)** — Все маршруты которые передаются с таким значением атрибута community не должны анонсироваться другим BGP-соседям,
* **no-export-subconfed (0xFFFFFF03)** — Все маршруты которые передаются с таким значением атрибута community не должны анонсироваться внешним BGP-соседям (ни внешним в конфедерации, ни настоящим внешним соседям). В Cisco это значение встречается и под названием local-as.

**Multi exit discriminator (MED)**

* Используется для информирования eBGP-соседей о том, какой путь в автономную систему более предпочтительный.
* Атрибут передается между автономными системами.
* Маршрутизаторы внутри соседней автономной системы используют этот атрибут, но, как только обновление выходит за пределы AS, атрибут MED отбрасывается.
* Чем меньше значение атрибута, тем более предпочтительна точка входа в автономную систему.

**Weight (проприетарный атрибут Cisco)**

* Позволяет назначить "вес" различным путям локально на маршрутизаторе.
* Используется в тех случаях, когда у одного маршрутизатора есть несколько выходов из автономной системы (сам маршрутизатор является точкой выхода).
* Имеет значение только локально, в пределах маршрутизатора.
* Не передается в обновлениях.
* Чем больше значение атрибута, тем более предпочтителен путь выхода.

**Выбор пути**

Характеристики процедуры выбора пути протоколом BGP:

* В таблице BGP хранятся все известные пути, а в таблице маршрутизации — лучшие.
* Пути выбираются на основании политик.
* Пути не выбираются на основании пропускной способности.

Сначала проверяется доступен ли next-hop (Route Resolvability Condition).

Для того чтобы next-hop считался доступным (accessible), необходимо чтобы в таблице маршрутизации был IGP-маршрут, который ведет к нему.

Cisco

На маршрутизаторе Cisco, если не настроены никакие политики выбора пути, выбор пути происходит таким образом (на каждый следующий шаг маршрутизатор переходит только при совпадении значений на предыдущем):

1. Максимальное значение weight (локально для маршрутизатора).
2. Максимальное значение local preference (для всей AS).
3. Предпочесть локальный маршрут маршрутизатора (next hop = 0.0.0.0).
4. Кратчайший путь через автономные системы. (самый короткий AS\_PATH)
5. Минимальное значение origin code (IGP < EGP < incomplete).
6. Минимальное значение MED (распространяется между автономными системами).
7. Путь eBGP лучше чем путь iBGP.
8. Выбрать путь через ближайшего IGP-соседа.
9. Выбрать самый старый маршрут для eBGP-пути.
10. Выбрать путь через соседа с наименьшим BGP router ID.
11. Выбрать путь через соседа с наименьшим IP-адресом.

**Looking Glass**

Одним из очень мощных инструментов работы с BGP – Looking Glass. Это сервера, расположенные в Интернете, которые позволяют взглянуть на сеть извне: проверить доступность, просмотреть через какие AS лежит путь в вашу автономную систему, запустить трассировку до своих внутренних адресов.