BGP

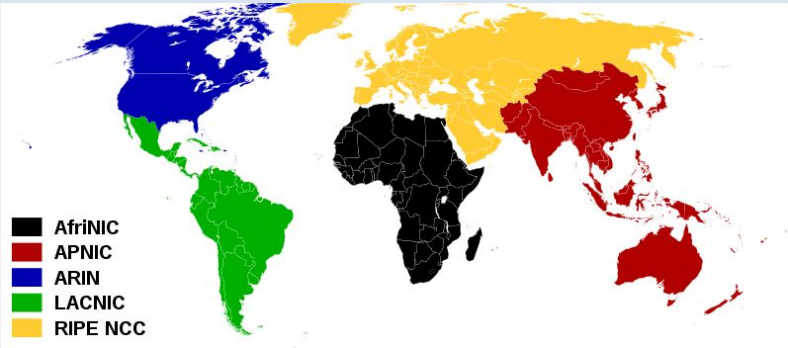
Елохин Р.Ю., Климанов М.М.
МФТИ
2013

Основные термины

- BGP – Border Gateway Protocol
- IGP – Interior Gateway Protocol
- EGP – Exterior Gateway Protocol
- AS – Autonomous System
- ASN – Autonomous System Number
- PA – Path Attributes
- AS_Path – набор ASN, через которые проходит маршрут
- NLRI – Network Layer Reachability Information
- iBGP – internal BGP
- eBGP – external BGP
- PI и PA IP-адреса
- Transit AS – транзитная автономная система (через неё передаётся трафик других AS)
- BGP speaker – маршрутизатор, на котором работает BGP
- Соседи (neighbor, peer) – два маршрутизатора, между которыми открыто TCP-соединение для обмена маршрутной информацией

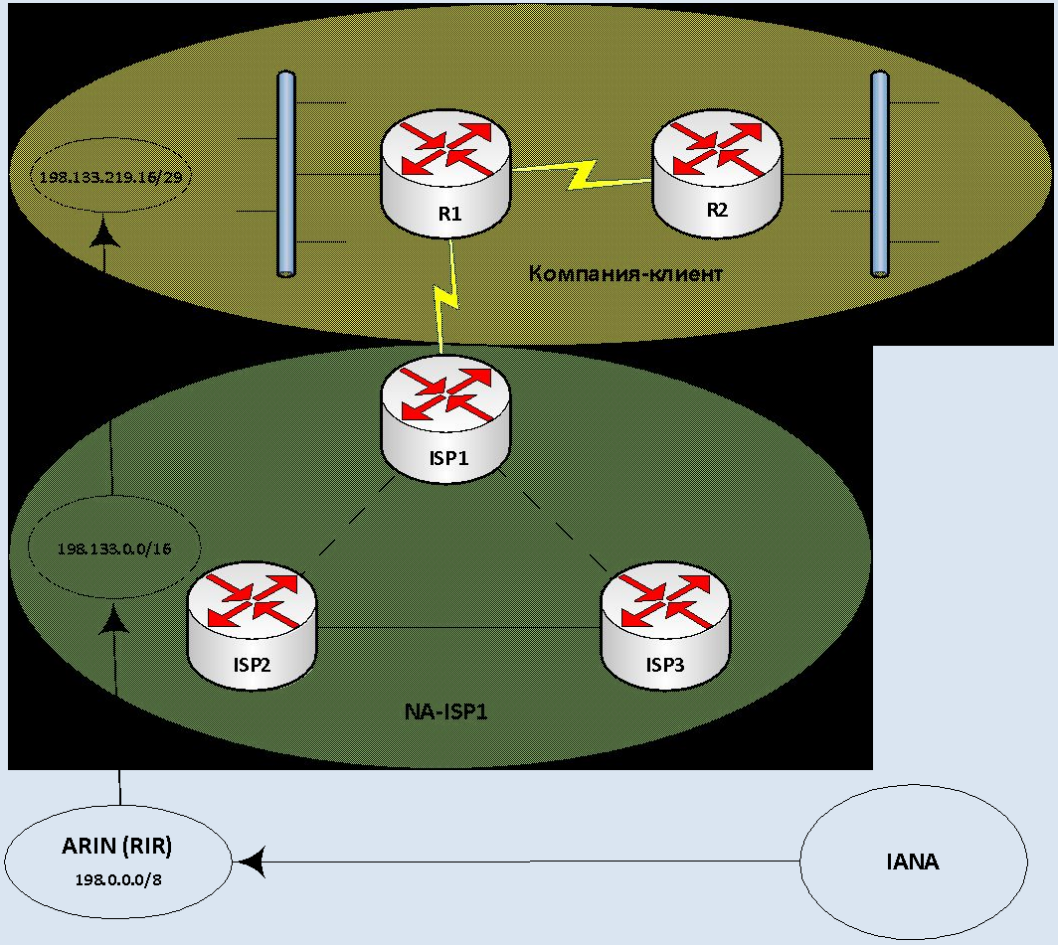
Распределение IP-адресов

- ICANN/IANA
- RIR - Regional Internet Registries
- NIR - National Internet Registries
- LIR - Local Internet Registries (обычно ISP)
- Конечный пользователь



Типы валидных IP-адресов

- PI – Provider Independent
- PA – Provider Aggregatable



Видео

Не маршрутизируемые глобально IP-адреса

Локальные адреса

Класс	Диапазон	Префикс
A	10.0.0.0	10.0.0.0/8
B	172.16.0.0-172.31.0.0	172.16.0.0/12
C	192.168.0.0-192.168.255.0	192.168.0.0/16

Зарезервированные адреса

Диапазон	Назначение
0.0.0.0/8	Self-identification в LAN
127.0.0.0/8	Loopback
169.254.0.0/16	Link-local
192.0.2.0/24	Документация и примеры
192.88.99.0/24	6to4 relay
198.18.0.0/15	Тесты производительности

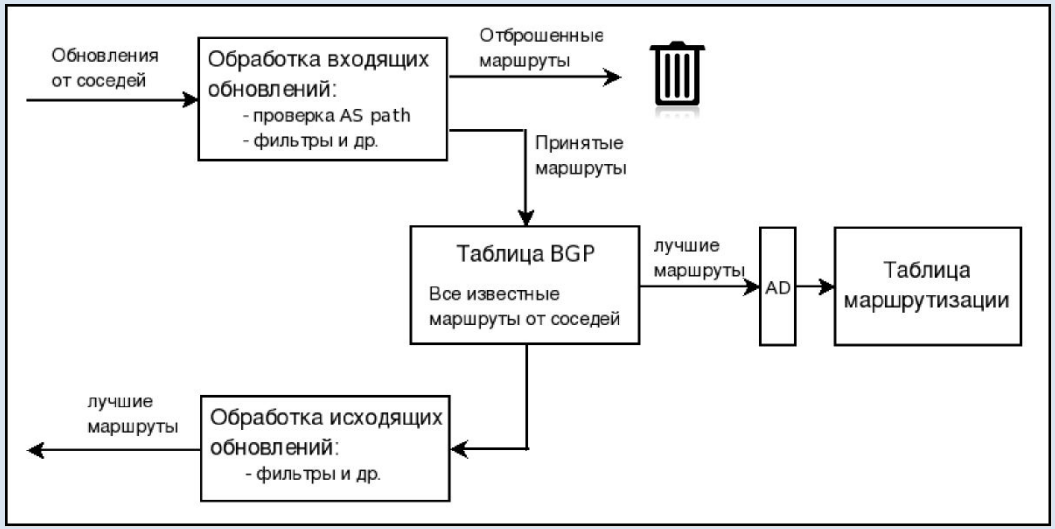
Видео

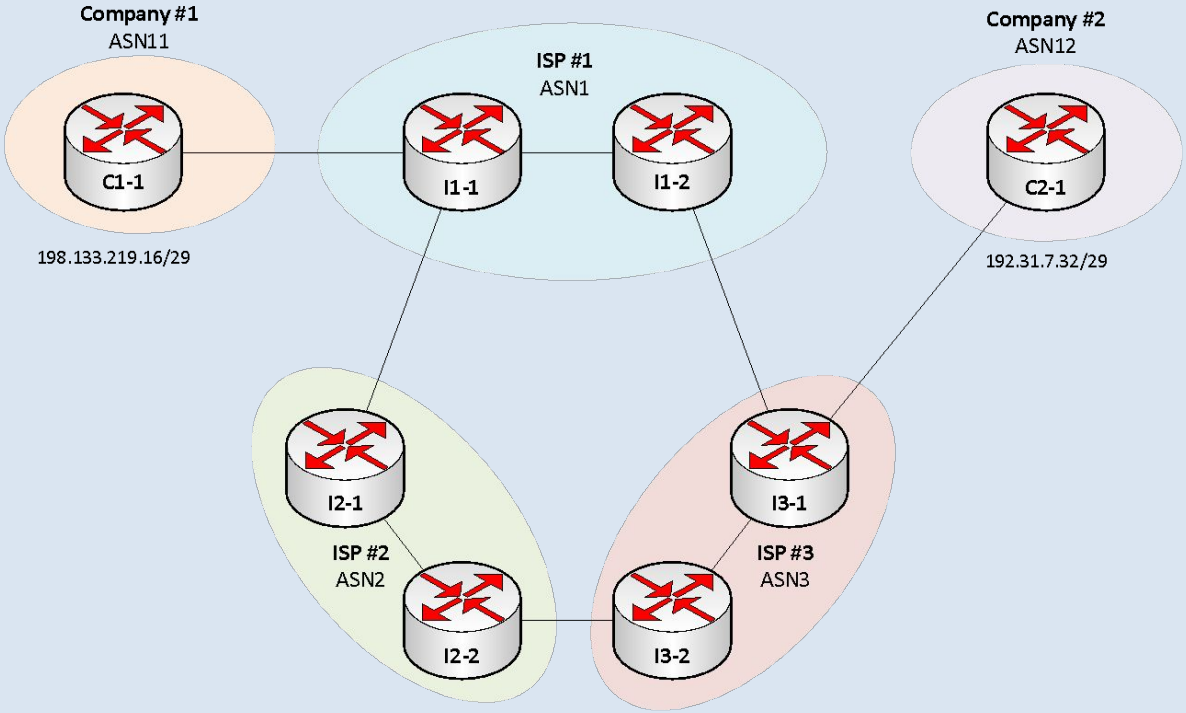
Сравнение IGP и EGP

OSPF/EIGRP	BGP
Устанавливаются соседские отношения перед отправкой маршрутной информации	Устанавливаются соседские отношения перед отправкой маршрутной информации
Автоматическое определение соседей (multicast)	IP-адреса соседей конфигурируются вручную
Не используют TCP	Использует TCP (порт 179)
Объявляют префикс/длину	Объявляют префикс/длину (NLRI)
Объявляют метрику	Объявляют целый набор атрибутов (PA), используемый для выбора лучшего маршрута вместо метрики
Ориентированы на быструю сходимость для выбора наиболее эффективного маршрута	Ориентированы на масштабируемость, выбираемый маршрут может быть не всегда самым эффективным
Link state (OSPF) Distance vector (EIGRP)	Path vector (похоже на distance vector)

Видео

Выбор лучшего пути, защита от петель





Видео

Типы ASN

Значение или диапазон	Тип
0	Зарезервировано
1-64495	Назначаются IANA для публичного использования
64496-64511	Зарезервировано для использования в документации
64512-65534	Для приватного использования
65535	Зарезервировано

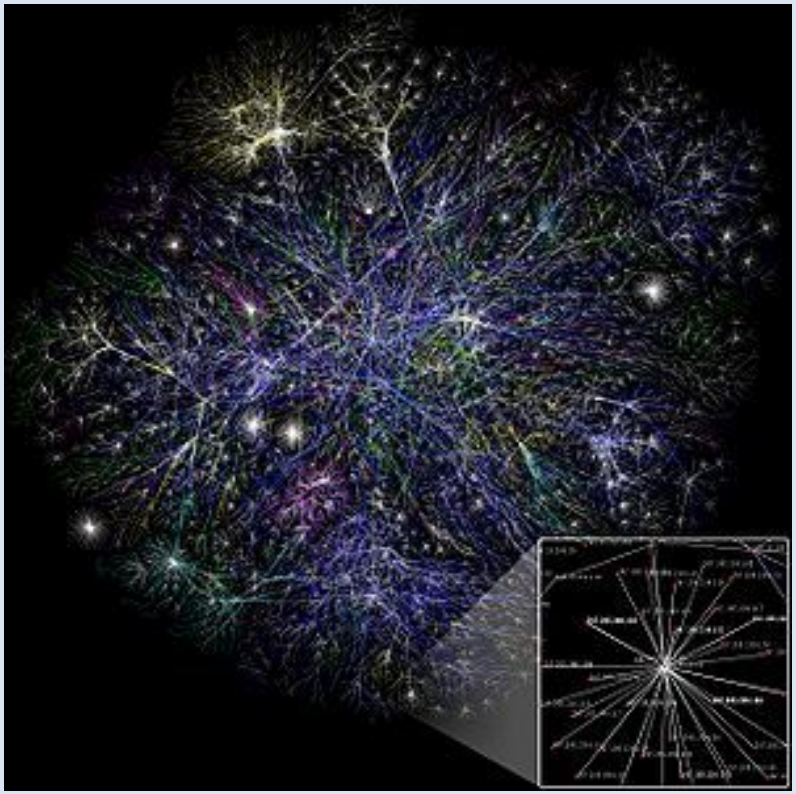
До 2007 года использовались только ASN длиной 16 бит, сейчас на ASN отводится 32 бита.

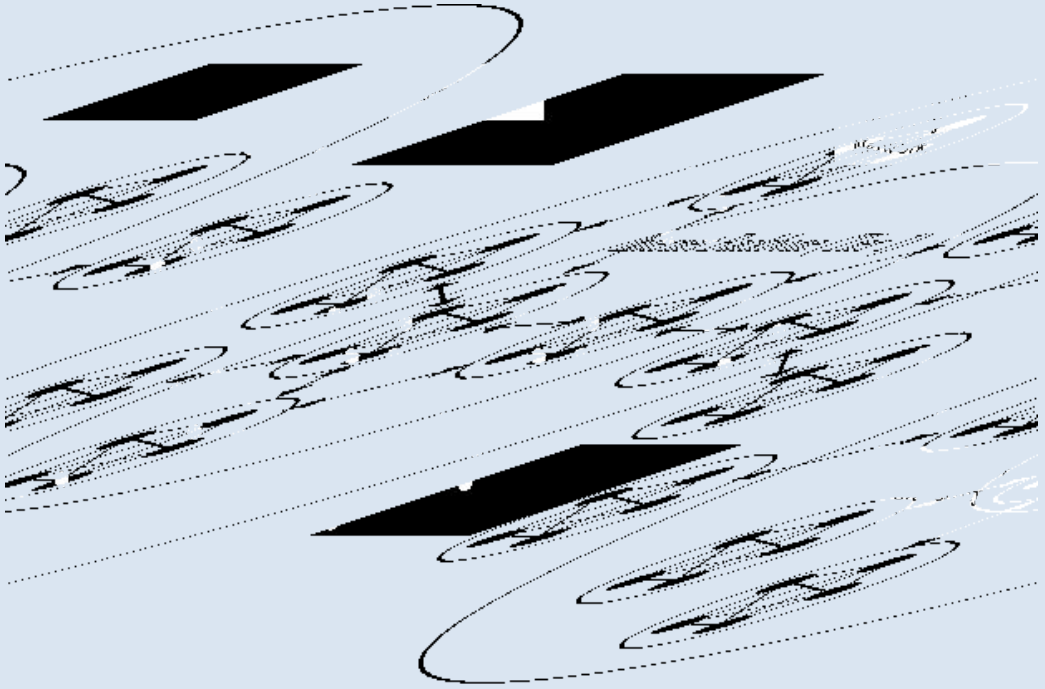
Типы подключений к провайдерам

- Single homed – один канал к одному оператору
- Dual homed – два канала к одному оператору
- Single multihomed – по одному каналу к нескольким операторам
- Dual multihomed – два и более каналов к нескольким операторам

Видео

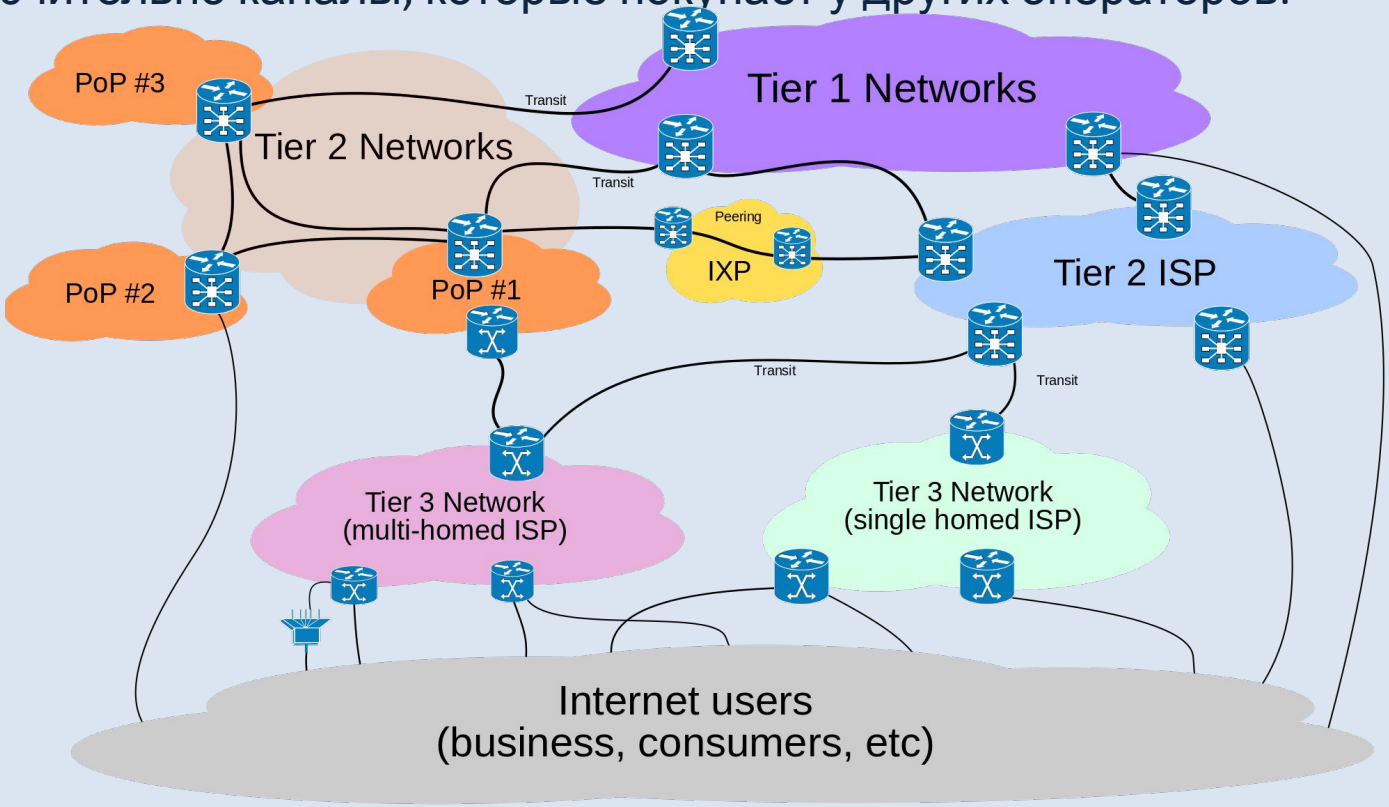
Карта глобальной сети





Типы провайдеров

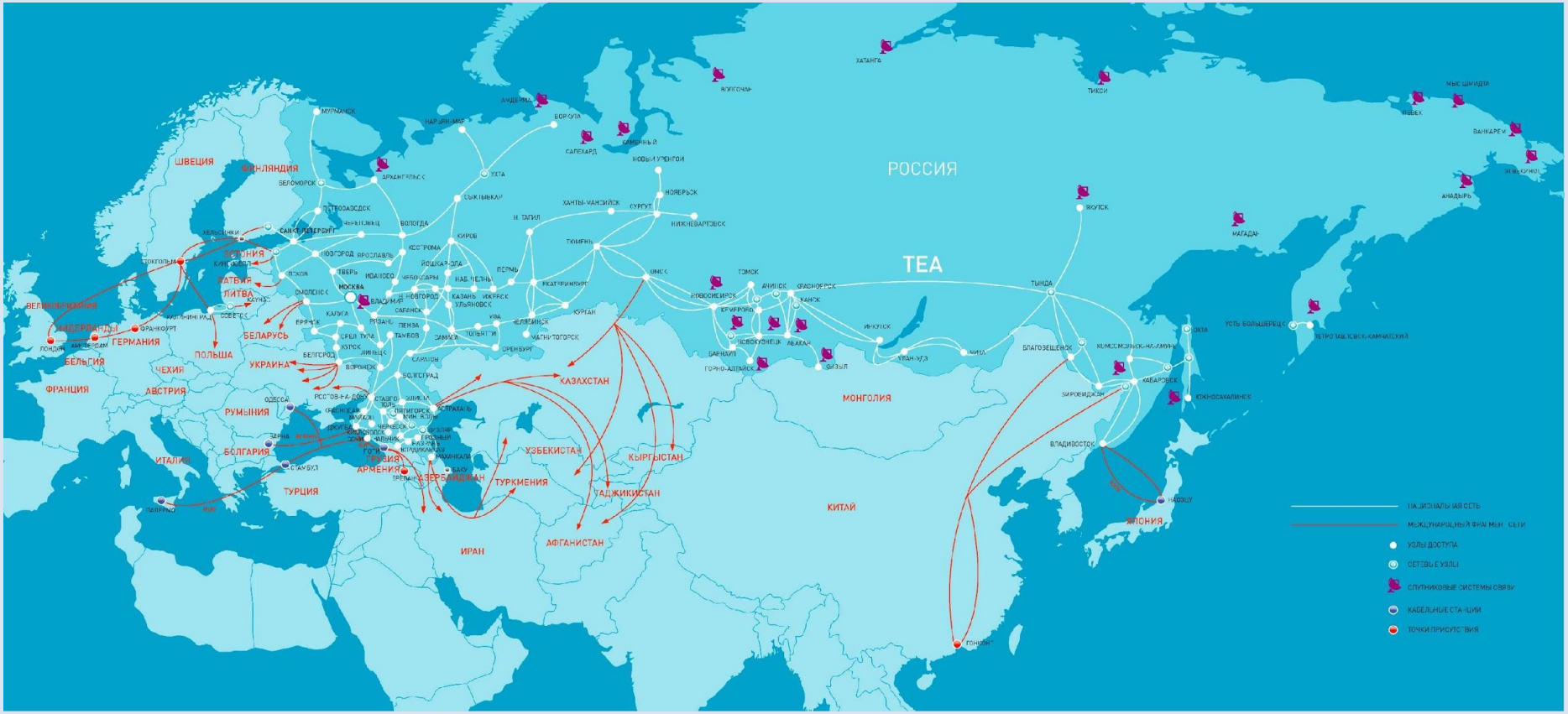
- Tier-1 — оператор, который имеет доступ к Интернету исключительно через пиринг-соединения.
- Tier-2 — оператор, который имеет доступ к части Интернета через пиринг-соединения, но покупает IP-транзит для доступа к остальной части Интернета.
- Tier-3 — оператор, который для доступа к Интернету использует исключительно каналы, которые покупает у других операторов.



Видео №1

Видео №2

Схема каналов Ростелекома



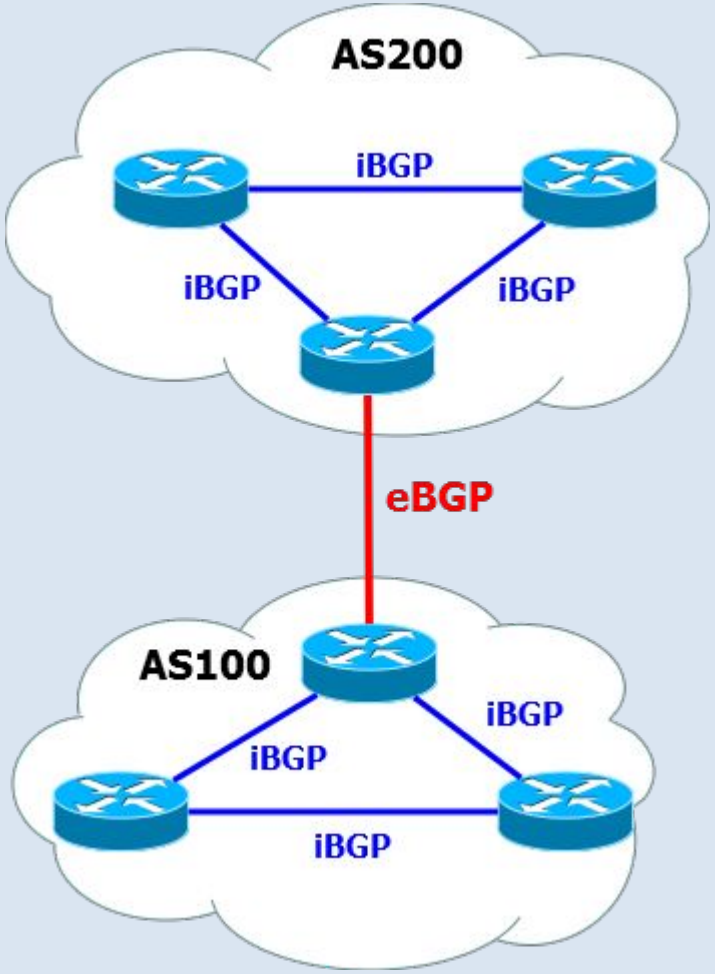
Типы обновлений маршрутной информации

- Только маршрут по умолчанию
- Полная таблица маршрутизации
- Частичная таблица маршрутизации (обычно вместе с маршрутом по умолчанию)

Видео

Типы соседей (связей)

- eBGP (external BGP) – BGP между разными автономными система
- iBGP (internal BGP) – BGP внутри одной автономной системы



Видео

Типы сообщений BGP

- **Open** — используется для установки отношений соседства и обмена базовыми параметрами. Отправляется сразу после установки TCP-соединения.
- **Update** — используется для обмена маршрутной информацией.
- **Notification** — используется когда возникают ошибки BGP. После отправки сообщения сессия с соседом разрывается.
- **Keepalive** — используется для поддерживания отношений соседства, для обнаружения неактивных соседей.

Видео

Выбор идентификатора маршрутизатора

1. Назначение вручную.
2. Наибольший IP-адрес присвоенный loopback-интерфейсу (в состоянии up/up).
3. Наибольший IP-адрес из всех других интерфейсов (в состоянии up/up).

Видео

Установление BGP-сессии

1. Маршрутизатор должен получить запрос на TCP-соединение с адресом отправителя, который маршрутизатор найдет указанным в списке соседей (команда neighbor).
2. 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).
3. Идентификаторы маршрутизаторов (Router ID) не должны совпадать.
4. Если настроена аутентификация, то соседи должны пройти её.

BGP выполняет проверку таймеров keepalive и hold, однако несовпадение этих параметров не влияет на установку отношений соседства. Если таймеры не совпадают, то каждый маршрутизатор будет использовать меньшее значение таймера hold.

Видео

Состояние связи с соседями

Состояние	Ожидание TCP	Инициация TCP	Установлено TCP	Отправлено Open	Получено Open	Сосед Up
Idle	Нет					
Connect	Да					
Active	Да	Да				
Open sent	Да	Да	Да	Да		
Open confirm	Да	Да	Да	Да	Да	
Established	Да	Да	Да	Да	Да	Да

Видео

Атрибуты пути (PA)

- **Well-known mandatory** — все маршрутизаторы, работающие по протоколу BGP, должны распознавать эти атрибуты. Должны присутствовать во всех обновлениях.
- **Well-known discretionary** — все маршрутизаторы, работающие по протоколу BGP, должны распознавать эти атрибуты. Могут присутствовать в обновлениях, но их присутствие не обязательно.
- **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.
- **Optional non-transitive** — могут не распознаваться всеми реализациями BGP. Если маршрутизатор не распознал атрибут, то атрибут игнорируется и при передаче соседям отбрасывается.

Видео

Примеры атрибутов

- **Well-known mandatory**
  - *Autonomous system path*
  - *Next-hop*
  - *Origin*
- **Well-known discretionary**
  - *Local preference*
  - *Atomic aggregate*
- **Optional transitive**
  - *Aggregator*
  - *Communities*
- **Optional non-transitive**
  - *Multi-exit discriminator (MED)*

Видео

Атрибут Autonomous system path

Принцип работы

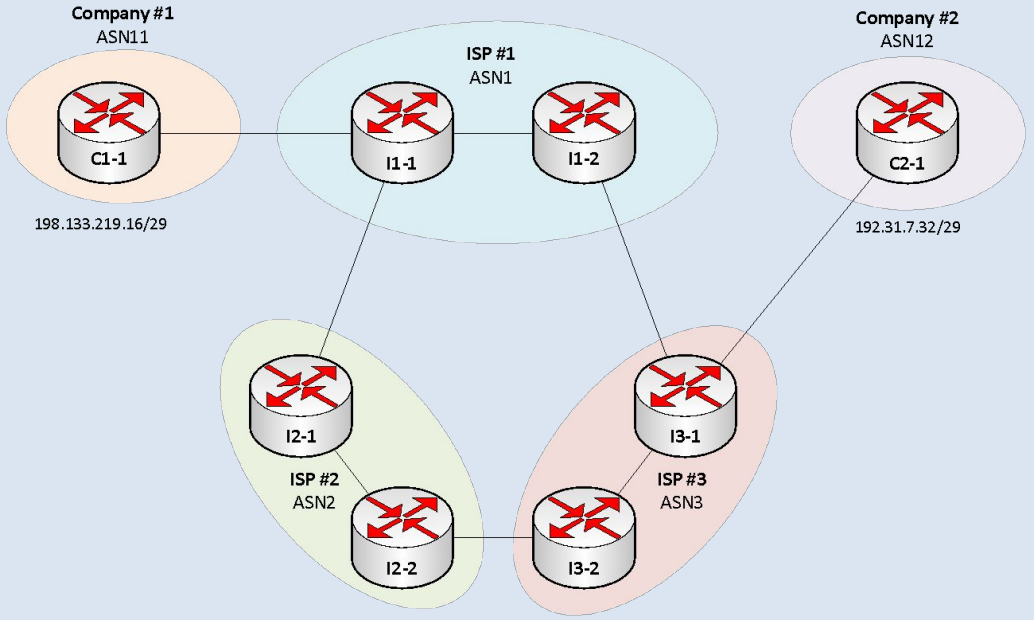
- Описывает через какие автономные системы надо пройти, чтобы дойти до сети назначения.
- Номер AS добавляется при передаче обновления из одной AS eBGP-соседу в другой AS.

Цели использования

- Обнаружение петель
- Применение политик

Каждая запись атрибута AS Path передаётся в виде поля TLV

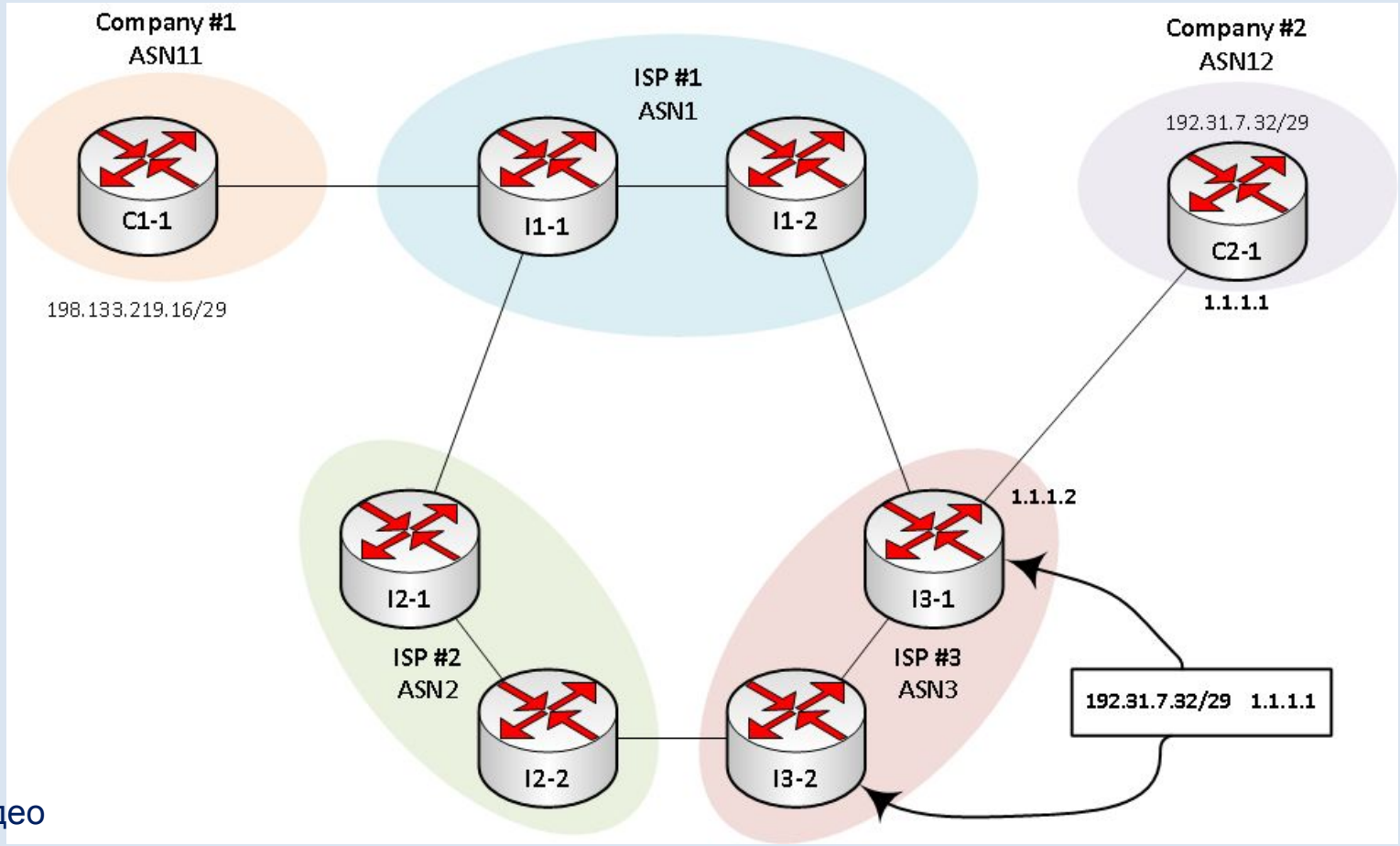
path segment type — поле (1 байт) для которого определены такие значения:
1 — *AS_SET*: неупорядоченное множество автономных систем, через которые прошел маршрут.
2 — *AS_SEQUENCE*: упорядоченное множество автономных систем, через которые прошел маршрут.
path segment length — поле (1 байт) указывает сколько автономных систем указано в поле path segment value.
path segment value — номера автономных систем, каждая представлена полем размером 2 байта.
Ручное изменение данного параметра обычно используется для управления входящим трафиком, но может применяться и для управления исходящим. Добавление номеров автономных систем в AS_Path называется AS_Path prepend.



Видео

Атрибут 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.



Видео

Атрибут Origin

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

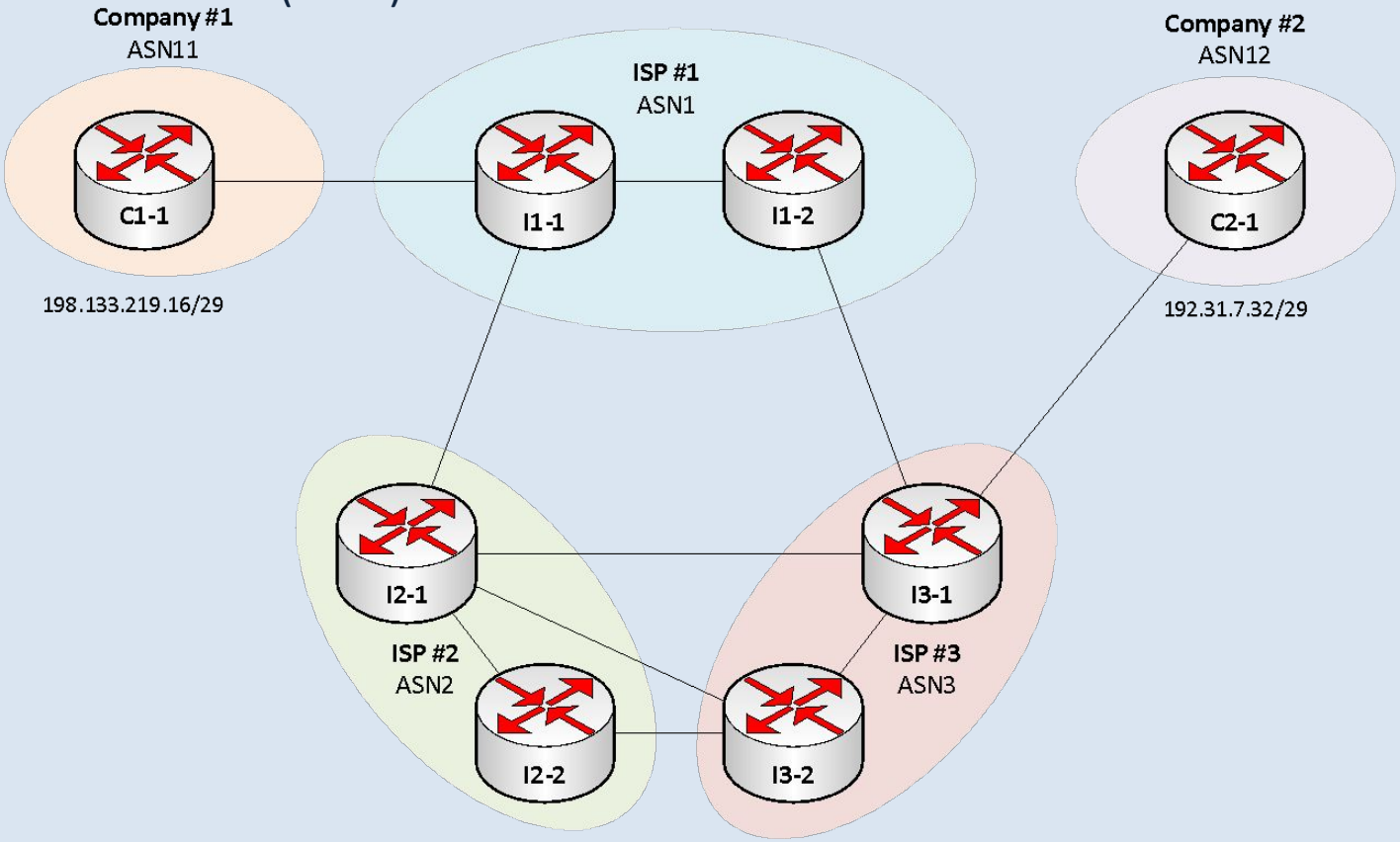
Возможные значения атрибута:

- **0 — *IGP***: NLRI получена внутри исходной автономной системы;
- **1 — *EGP***: NLRI выучена по протоколу Exterior Gateway Protocol (EGP). Предшественник BGP, сегодня не используется.
- **2 — *Incomplete***: NLRI была выучена каким-то другим образом

Видео

Атрибут 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.
- Диапазон от 0 до 4294967295 (2^{32}-1).

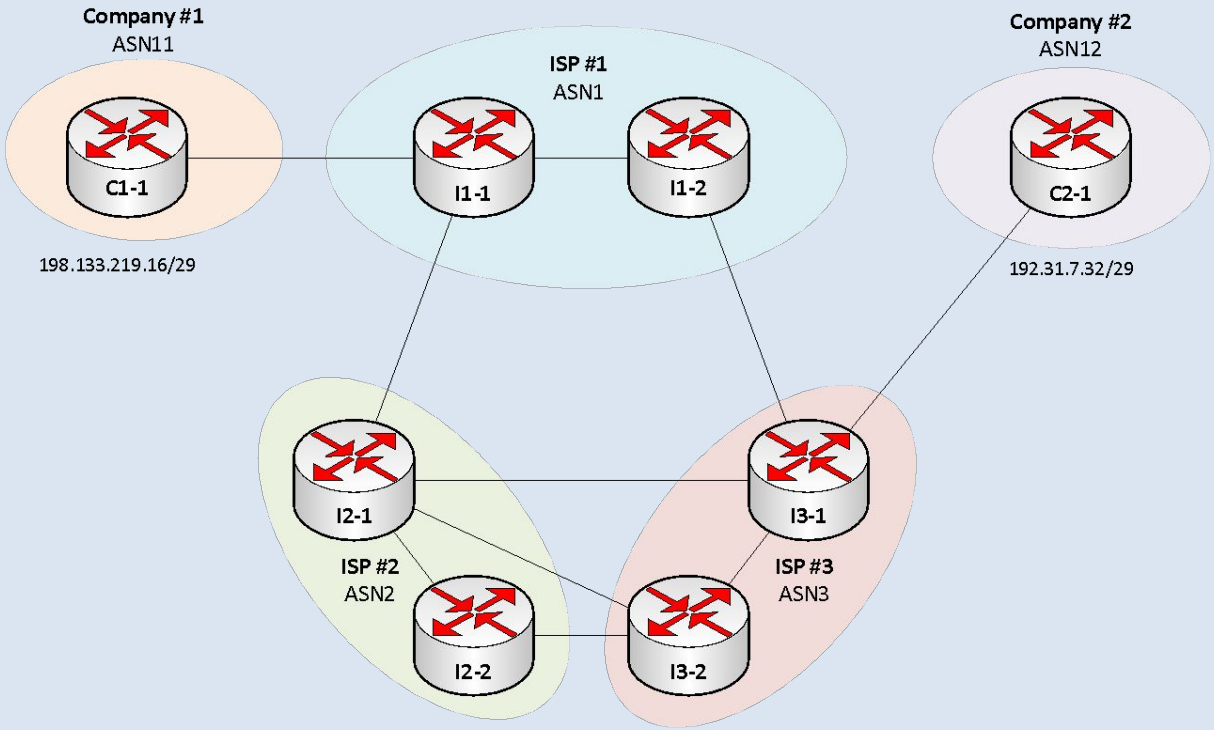


Видео

Атрибут Communities

Используется для управления анонсированием сетей.
Атрибут optional transitive, размерность - 32 бита.
В новом формате (ip bgp-community new-format) число представляется в виде ASN:NN, где ASN – номер автономной системы.
Существуют некоторые предопределённые сообщества:

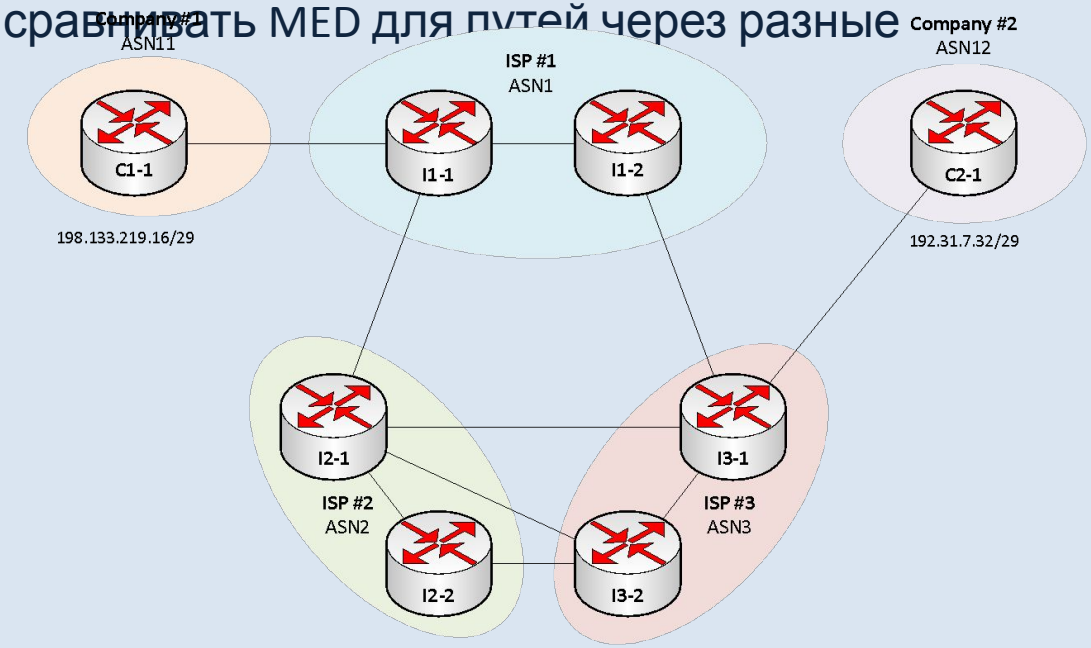
- NO_EXPORT (0xFFFFFF01) — не отдавать префиксы полноценными eBGP пирам, но отдавать eBGP пирам внутри конфедерации;
- NO_ADVERTISE (0xFFFFFF02) — не отдавать префиксы никому;
- NO_EXPORT_SUBCONFED (0xFFFFFF03) — не отдавать префиксы никаким eBGP пирам, включая конфедерацию;
- Internet (0x0) – отдавать всем пирам (отсутствует в RFC, введено Cisco).
- Blackhole (XX:666) – маршрут до /32 узла, который должен быть заблокирован вышестоящим провайдером.



Видео

Атрибут Multi exit discriminator (MED)

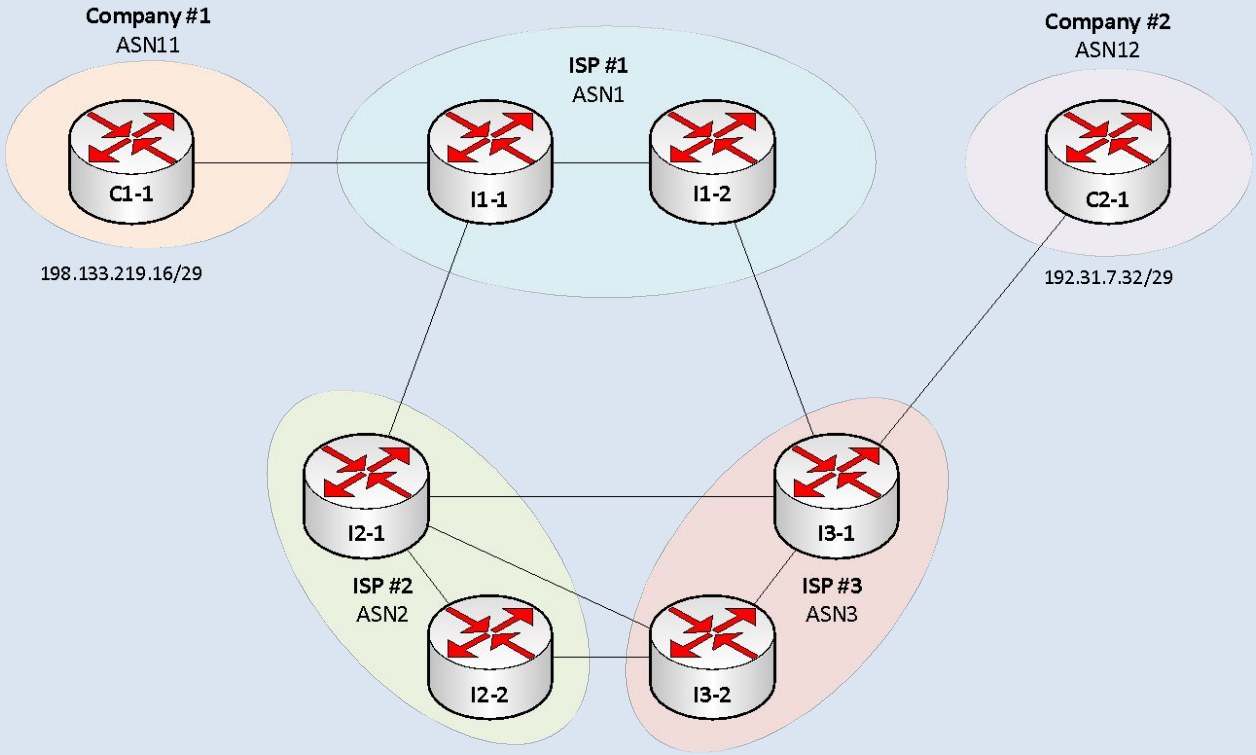
- Используется для информирования eBGP-соседей о том, какой путь в автономную систему более предпочтительный.
- Атрибут передается между автономными системами.
- Маршрутизаторы внутри соседней автономной системы используют этот атрибут, но, как только обновление выходит за пределы AS, атрибут MED отбрасывается.
- Чем меньше значение атрибута, тем более предпочтительна точка входа в автономную систему.
- Диапазон от 0 до 4294967295 (2^{32}-1), по умолчанию 0 (если не была введена команда *bgp bestpath med missing-as-worst*).
- Проверяется только для путей в через одну AS. Команда *bgp always-compare-med* позволяет сравнивать MED для путей через разные автономные системы.



Видео

Атрибут Weight (только в маршрутизаторах Cisco)

- Используется для выбора наиболее предпочтительного пути из автономной системы.
- Атрибут не передается ни между автономными системами, ни внутри собственной автономной системы.
- Чем больше значение атрибута, тем более предпочтительна точка выхода из автономной системы.
- Диапазон от 0 до 65535 (2^{16}-1).
- По умолчанию вес равен 0 для изученных маршрутов по BGP и 32768 для локальных маршрутов.



Видео

Выбор пути (Best Path Selection) для маршрутизаторов Cisco

0. Наличие маршрута до Next-hop.
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 (команды network/aggregate (network/redistribute более предпочтительные, чем aggregate))).
4. Кратчайший путь через автономные системы (самый короткий AS_PATH).
5. Минимальное значение origin type (IGP < EGP < incomplete).
6. Минимальное значение MED (распространяется между автономными системами).
7. Путь eBGP лучше чем путь iBGP.
8. Маршрут полученный не от RR (без атрибута ORIGINATOR_ID) предпочтительнее, чем маршрут, который был получен от RR.
9. Reflected-маршруты с более коротким cluster-list предпочтительнее.
10. Выбрать путь через ближайшего IGP-соседа.
11. Требуется ли внести несколько путей в таблицу маршрутизации.
12. Выбрать самый старый маршрут для eBGP-пути.
13. Выбрать путь через соседа с наименьшим BGP router ID.
14. Выбрать путь с наименьшим списком кластеров (при использовании BGP Route Reflector).

Видео

15. Выбрать путь через соседа с наименьшим IP-адресом.

Дополнительные возможности
конфедерации

Создание конфедерации позволяет:

- избежать необходимости создания полносвязной топологии между всеми iBGP-соседями,
- всем iBGP-соседям выучить все iBGP-маршруты в AS,
- предотвратить образование петель.

При использовании конфедераций BGP, автономная система разбивается на подавтономные системы (sub-AS). Маршрутизаторы, которые находятся в одной sub-AS называются confederation iBGP-соседи, а маршрутизаторы в разных sub-AS называются confederation eBGP-соседи.

Правила работы маршрутизаторов в конфедерации:

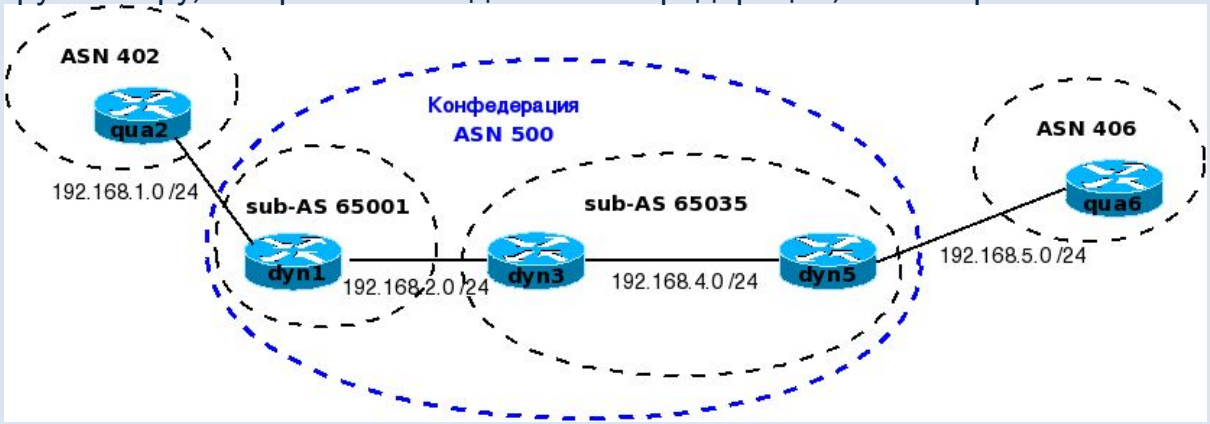
- iBGP-соседи в конфедерации должны быть соединены в полносвязную топологию (full mesh). Они, как и обычные iBGP-соседи, не передают iBGP-маршруты друг другу.
- eBGP-соседи в конфедерации:

✔ как и eBGP-соседи анонсируют iBGP-маршруты выученные внутри sub-AS конфедерации в другую sub-AS,

✔ как и eBGP-соседи по умолчанию используют для пакетов TTL 1 (изменяется neighbor ebgp-multihop),

✔ во всех остальных случаях работают как обычные iBGP-соседи (например, next-hop по умолчанию не изменяется).

- Внутри конфедераций для предотвращения петель используется атрибут AS Path. Маршрутизаторы, которые находятся в конфедерации добавляют в атрибут сегменты AS_CONFED_SEQ и AS_CONFED_SET.
- Когда маршрутизатор выбирает лучший маршрут на основании атрибута AS Path, номера автономных систем конфедераций не учитываются.
- Когда обновление отправляется маршрутизатору, который не находится в конфедерации, то номера конфедераций удаляются.



Видео

Дополнительные возможности

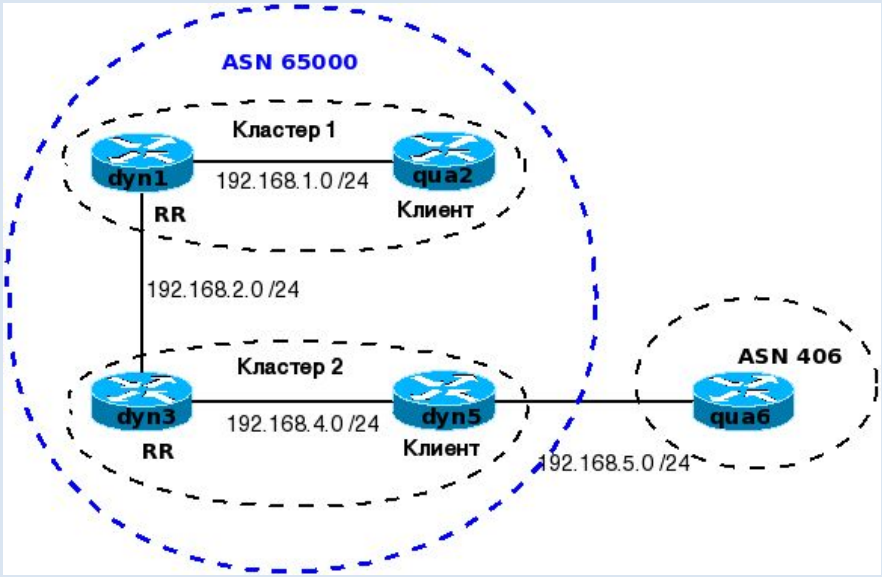
route reflector

- Предоставляет возможность избежать необходимости создания полносвязной топологии между всеми iBGP-соседями.
- Разрешает всем iBGP-соседям выучить все iBGP-маршруты в AS.
- Позволяет предотвратить образование петель.

При использовании RR для маршрутизаторов в AS определяются такие три роли:

- RR сервер (RR Server, RR)
- Клиент (Client)
- Не клиент (Nonclient)

Источник от которого выучен префикс	Анонсируется ли маршрут клиентам?	Анонсируется ли маршрут не клиентам?
Клиент	Да	Да
Не клиент	Да	Нет
eBGP	Да	Да



Видео

Вопросы к теоретической части

Краткое содержание

- IP-адреса и номера автономных систем.
- Типы провайдеров и их взаимоотношения.
- Особенности работы протокола BGP.
- Выбор пути в BGP.
- Дополнительные возможности

Видео

Ресурсы

- CCNP Route 642-902 Official Certification Guide, Wendell Odom, 2010, Cisco Press.
- http://www.cisco.com/en/US/tech/tk365/technologies_tech_note09186a0080094431.shtml
- http://xgu.ru/wiki/BGP
- http://xgu.ru/wiki/BGP_%D0%B2_Cisco