# 1. Технология CIDR, маски переменной длины, понятие «префикс».

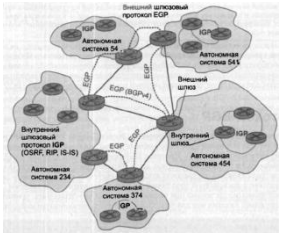
Бесклассовая адресация (англ. Classless Inter-Domain Routing, англ. CIDR) — метод IP-адресации, позволяющий на основе маски переменной длины гибко управлять пространством IP-адресов (не используя жёсткие рамки классовой адресации).

Во-первых, CIDR позволяет экономно использовать ограниченный ресурс IP-адресов. Пример разделения сети на подсети с использованием масок переменной длины и соответствующие правила маршрутизации см. [Олифер, стр. 538 – 541].

Во вторых, CIDR позволяет агрегировать маршруты на магистральных маршрутизаторах. Каждому поставщику услуг Интернета назначается непрерывный диапазон IP-адресов. Все адреса каждого поставщика услуг имеют общую старшую часть — префикс, поэтому маршрутизация на магистралях Интернета может осуществляться на основе префиксов, а не полных адресов сетей. А это значит, что вместо множества записей по числу сетей будет достаточно поместить одну запись сразу для всех сетей, имеющих общий префикс. Такое агрегирование адресов позволит уменьшить объем таблиц в маршрутизаторах всех уровней, а следовательно, ускорить работу маршрутизаторов и повысить пропускную способность Интернета. [Олифер, стр. 544 – 546]
Агрегирование маршрутов применяется, в частности, в протоколе BGP-4.

# 2. Понятие AS. Понятие маршрутизации между AS, протокол BGP.

Применение нескольких протоколов маршрутизации даже в пределах небольшой составной сети — дело не простое, от администратора требуется провести определенную работу по конфигурированию каждого маршрутизатора. Очевидно, что для крупных составных сетей нужно качественно иное решение. [Олифер, стр. 586]. Такое решение было найдено для самой крупной на сегодня составной сети — Интернета. Это решение базируется на понятии автономной системы.
**2.1. Понятие AS.** Автономная система (Autonomous System, AS), RFC 1930 (March 1996), — это совокупность сетей под единым административным управлением, обеспечивающим общую для всех входящих в автономную систему маршрутизаторов политику маршрутизации.
Обычно автономной системой управляет один поставщик услуг Интернета, самостоятельно выбирая, какие протоколы маршрутизации должны использоваться в некоторой автономной системе и каким образом между ними должно выполняться перераспределение маршрутной информации. Крупные поставщики услуг и корпорации могут представить свою составную сеть как набор нескольких автономных систем. Регистрация автономных систем происходит централизованно, как и регистрация IP-адресов и DNS-имен. Номер автономной системы состоит из 16 разрядов и никак не связан с префиксами IP-адресов входящих в нее сетей (устарело, теперь AS 32 битные, RFC4893 (May 2007)).
В соответствии с этой концепцией Интернет выглядит как набор взаимосвязанных автономных систем, каждая из которых состоит из взаимосвязанных сетей, соединенными внешними шлюзами. [Олифер, стр. 587]




Таким образом, Интернет состоит не из хостов или сетей; он состоит из AS. Моделью Интернета служит граф, узлами которого являются AS, а ребра соединяют пары AS.

**2.2 Маршрутизация между AS.** Основная цель деления Интернета на автономные системы — обеспечение многоуровневого подхода к маршрутизации. До введения автономных систем предполагался двухуровневый подход, то есть сначала маршрут определялся как последовательность сетей, а затем вел непосредственно к заданному узлу в конечной сети.
Выбор маршрута между AS осуществляют внешние шлюзы, использующие особый тип протокола маршрутизации, так называемый внешний шлюзовой протокол (Exterior Gateway Protocol, EGP). В настоящее время для работы в такой роли сообщество Интернета утвердило стандартный пограничный шлюзовой протокол версии 4 (Border Gateway Protocol, BGPv4), RFC1771 (March 1995) устарел и заменен на RFC4271 (January 2006).
Внутри AS за маршрут отвечают внутренние шлюзовые протоколы (Interior Gateway Protocol, IGP). К числу IGP относятся RIP, OSPF и IS-IS. В случае транзитной автономной системы эти протоколы указывают точную последовательность маршрутизаторов от точки входа в автономную систему до точки выхода из нее. [Олифер, стр. 586 – 588]
Не менее важными являются и «политические моменты». Пускать ли чужой трафик? Стоимость, безопасность.

# 3. Подробно про AS согласно RFC 1930.

[Тут надо понятие префикса, блока IP адресов, CIDR блока; суть технологии CIDR.]

Главным концептуальным документом, который регламентирует вопросы регистрации и использования автономных систем является RFC 1930. Так же RFC 1930 содержит список требований к AS. Конкретные инструкции по регистрации AS содержатся в документах RIR RIPE.
**3.1 Определение.** Автономная система представляет собой элемент политики маршрутизации в современной среде внешней маршрутизации и, в частности, понятие AS применяется к протоколам типа EGP (Exterior Gateway Protocol – устаревший протокол) и BGP.

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).

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.

Короче, AS представляет собой группу из одного или нескольких префиксов IP, работающих у одного или нескольких сетевых операторов, которые имеют единую (SINGLE) и четко определенную (CLEARLY DEFINED) политику маршрутизации.

Политика маршрутизации (routing policy) в данном случае понимается как набор решений о пересылке (routing decisions), принимаемых в современной сети Internet. Эта политика представляет собой обмен маршрутной информацией, которая является субъектом политики маршрутизации, между AS.
**3.2 Классификация**. Трафик, который генерируется или завершается в одной AS (т. е., IP-адрес отправителя или получателя пакетов IP идентифицирует хост, входящий в данную AS), 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. Multihoming implies that you will be talking BGP to at least two other BGP peers and they will be announcing your block of IP nets with your ASN as the source of those nets.(см. также Frequently Asked Questions on Multi-homing and BGP)
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.
Протокол BGP не вносит ограничений в топологию соединений между AS.
Автономная система AS имеет уникальный (в глобальном масштабе) идентификатор, который иногда называют ASN (Autonomous System Number – номер автономной системы); этот идентификатор используется как при обмене маршрутными данными между соседними AS, так и в качестве обозначения самой AS. Присваивается IANA.
**3.3. Когда нужна AS – критерии.** Уникальная политика маршрутизации. AS является необходимой только в тех случаях, когда используется политика маршрутизации, отличающаяся от политики граничного маршрутизатора в соседней системе того же уровня (peer). (В данном случае политика маршрутизации определяет, как остальная часть Internet будет принимать решение о маршрутизации на основе сведений из данной AS.)

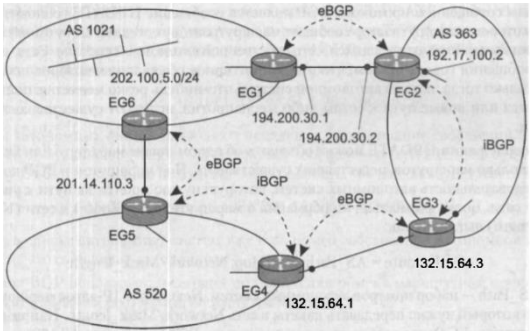
# 4. Как получают AS. RIR, LIR, организации Интернета.

Пример. Организация арендует сети у провайдера и хочет иметь второе подключение. Каковы условия получения номера AS, необходимого для такого подключения?
Во-первых, выделение номера AS требует наличия определённого адресного пространства (не менее класса C), а во-вторых, необходимость регистрации AS должна быть обусловлена наличием у сети собственной политики маршрутизации.
Такие сети называют мультихоумными (multihomed): Multihoming implies that you will be talking BGP to at least two other BGP peers and they will be announcing your block of IP nets with your ASN as the source of those nets.
Формально для регистрации номера AS необходимо знать номера AS своих провайдеров и IP-адреса, которые будут анонсироваться данной AS.
Необходимо помнить, что наиболее распространённой является ситуация, когда организация получает номер AS после получения статуса Локальной Регистратуры и провайдерского блока адресов.
Как правило, сеть, организующая Автономную Систему, располагает адресами, не являющимися подмножеством какого-либо провайдерского блока.

# 5. Подробно про BGP согласно RFC4271.

**5.1. Общее представление о протоколе и пример.**
Пограничный (внешний) шлюзовой протокол (Border Gateway Protocol, BGP) версии 4 является сегодня основным протоколом обмена маршрутной информацией между автономными системами Интернета.
Основной функцией поддерживающей протокол BGP системы является обмен информацией о доступности сетей с другими системами BGP. Информация о доступности сетей включает список автономных систем (AS), через которые проходит эта информация. Этих сведений достаточно для построения графа связности AS, из которого могут исключаться маршрутные петли (routing loop), а также для принятия некоторых решений на уровне политики AS.
BGP-4 обеспечивает новые механизмы поддержки бесклассовой междоменной маршрутизации (CIDR) [RFC1518, RFC1519]. Эти механизмы включают поддержку анонсирования группы адресатов с помощью префикса IP и позволяют обойтись без концепции «класса» сетей в рамках протокола BGP.
BGP-4 также добавляет механизм объединения маршрутов, включающий объединение путей AS.
Маршрутная информация, передаваемая с использованием BGP поддерживает только парадигму пересылки на основе адреса получателя (destination-based forwarding paradigm), которая предполагает, что маршрутизатор пересылает пакеты, опираясь лишь на адрес получателя, содержащийся в заголовке IP-пакета. Это, в свою очередь, отражает набор правил политики, которые могут быть применены (или не применены) с использованием BGP. Протокол BGP может поддерживать только правила, соответствующие парадигме пересылки по адресу получателя.
BGP использует в качестве транспортного протокол TCP [RFC793]. Это избавляет от необходимости реализации явного фрагментирования уведомлений, повторной передачи и порядковых номеров. BGP слушает протокол TCP через порт 179. Механизм уведомлений об ошибках, используемый в BGP, предполагает, что TCP поддерживает аккуратное завершение соединений (т. е., все остающиеся данные будут доставлены прежде, чем соединение будет закрыто). Между парой систем организуется соединение TCP. После этого системы обмениваются между собой стандартными сообщениями для согласования и подтверждения параметров соединения.

Пример. [Олифер, стр. 589]




База маршрутной информации (Routing Information Base, RIB) узла BGP состоит из трех отдельных частей.
a) Adj-RIBs-In – маршрутные данные, полученные из входящих сообщений UPDATE, которые были приняты от других узлов BGP. Эта база представляет маршруты, которые могут использоваться как входные данные для процесса принятия решения (Decision Process).
b) Loc-RIB – локальная маршрутная информация узла BGP, выбранная путем применения локальной политики к маршрутам, содержащимся в Adj-RIBs-In. Эти маршруты будут использоваться локальным узлом BGP. Значения next hop для каждого из этих маршрутов должны быть преобразуемыми с помощью таблицы маршрутизации (Routing Table) локального узла BGP.
c) Adj-RIBs-Out – информация локального узла BGP, выбранная им для анонсирования своим партнерам. Маршрутные данные из Adj-RIBs-Out будут передаваться от локального узла BGP в сообщениях UPDATE для анонсирования партнерам. [RFC 4271, стр. 5]

**5.2. Формализация протокола как конечномерного детерминированного автомата.**
Работа BGP может быть описана в терминах машины конечных состояний (Finite State Mashine FSM).
Idle. FSM отвергает все входящие соединения BGP для данного узла. Никаких ресурсов не выделено.
ManualStart (1) или AutomaticStart (3) -> инициировать соединение TCP с другим узлом BGP; прослушивать соединения, инициированные удаленными узлами BGP. Connect.
ManualStart_with_PassiveTcpEstablishment (4) или AutomaticStart_with_PassiveTcpEstablishment (5) -> прослушивать соединения, инициированные удаленными узлами BGP. Active.
Connect. FSM ожидает завершения процесса организации соединения TCP.
Успешная организация соединения TCP (16 или 17) [if DelayOpen == FALSE] -> передавать партнеру сообщение OPEN. OpenSent.
Получено сообщение OPEN при запущенном таймере DelayOpenTimer (Событие 20) -> завершать инициализацию BGP; передавать сообщение OPEN; передавать сообщение KEEPALIVE. OpenConfirm [Если значение поля AS совпадает с номером локальной автономной системы, для соединения устанавливается статус внутреннего, в противном случае соединение считается внешним].
Active. FSM пытается приобрести партнеров путем прослушивания и восприятия соединений TCP.
OpenSent. FSM ожидает сообщения OPEN от партнера.
Получено сообщение OPEN [if OPEN не содержит ошибок] (Событие 19) -> передавать сообщение KEEPALIVE. OpenConfirm.
OpenConfirm. FSM ожидает приема сообщения KEEPALIVE или NOTIFICATION.
Получено сообщение KEEPALIVE (Событие 26) -> Established.
Получено сообщение NOTIFICATION [например, с кодом ошибки Hold Timer Expired] (Событие 26) -> Idle.
Established. FSM может обмениваться сообщениями UPDATE, NOTIFICATION и KEEPALIVE со своим партнером.
Получено сообщение UPDATE [if нет ошибок] (Событие 26) -> обрабатывать принятое сообщение. Established.
Получено сообщение NOTIFICATION [например, с кодом ошибки NotifMsgVerErr] (Событие 24) -> Idle.

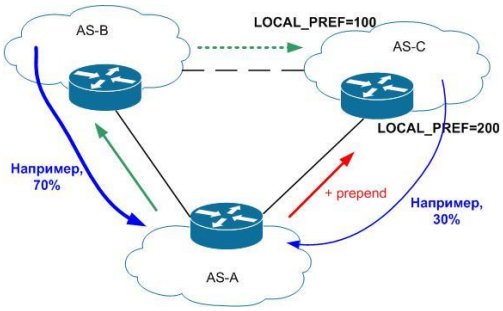
**5.3. Предотвращение конфликтов соединения.**
Реализация BGP должна поддерживать отдельную FSM для каждого включенного в конфигурацию партнера.
Реализация BGP должна подключиться к порту TCP с номером 179 и прослушивать его с целью приема входящих вызовов в дополнение к своим попыткам организовать соединение с партнером. Для каждого входящего соединения должен создаваться экземпляр машины состояний. Существует период, в течение которого соединение с партнером на другой стороне уже организовано, но его идентификатор BGP еще не известен. В течение этого периода могут одновременно существовать входящее и исходящее соединение для одной пары партнеров. Такая ситуация называется конфликтом при соединении.
Если пара узлов BGP пытается одновременно организовать соединение TCP друг с другом, между узлами такой пары могут возникнуть два параллельных соединения. Если IP-адрес отправителя в одном из таких соединений совпадает с IP-адресом получателя в другом соединении и наоборот, возникает конфликт при соединении. При возникновении такого конфликта одно из соединений должно быть закрыто.
Между парой партнеров может существовать несколько соединений, если в них используются различные пары адресов IP. Такая ситуация называется «МНОЖЕСТВЕННЫМ ПАРТНЕРСТВОМ» (multiple "configured peerings").

**5.4. Сообщения BGP.**
OPEN, UPDATE, NOTIFICATION и KEEPALIVE описаны в RFC 4271, пункт 4.

**5.5. Выбор оптимального маршрута.**
Выбор маршрута описан в RFC 4271, пункт 9.

# 6. Пример управления трафиком с использованием протокола BGP.

Каким образом можно влиять на поведение входящего трафика? Классика – искусственно удлинять AS-PATH (prepend), отправлять анонсы провайдеру с некоторыми communities для занижения провайдерского local preference. Однако результат во многом зависит от политик провайдера.




Наша AS-A имеет связи с двумя провайдерами: AS-B (основной) и AS-C (резервный). Свои сети мы анонсируем обоим провайдерам, но в сторону резервного мы специально удлиняем AS_PATH (хотим получить трафик в этот канал только при неисправностях с основным).
Резервный провайдер получает анонсы о наших сетях из двух источников: непосредственно от клиента (от нас) и от своих пиринговых партнеров (пунктирная линия). Во многих случаях приходится сталкиваться с тем, что провайдер считает более приоритетным путем в клиентскую сеть тот путь, который непосредственно соединяет его с клиентом. Для этого он (резервный провайдер) увеличивает значения local preference на анонсы, полученные непосредственно от клиента (в данном случае 200), а не от пира (в данном случае 100). Всем своим соседям он расскажет именно об удлиненном пути (анонсах полученных от клиента), так как BGP маршрутизатор анонсирует дальше только лучший маршрут.
Значит, если трафик будет проходить через автономную сеть провайдера AS-B, то получать мы его будем в основной канал, если через сеть провайдера AS-C – в резервный. В итоге, хотим мы того или нет, но входящий трафик к нам будет «заходить» с обоих каналов. В добавок мы получаем асимметрию: пытаемся отправить трафик в основной канал, а получаем его и с основного, и с резервного.
**Небольшой подитог.** При двух и более провайдерах, трафик будет «литься» со всех сторон.

## Контрольные задания и вопросы.

1. При обнаружении ошибок в процессе обработки сообщений BGP соединение BGP разрывается (the BGP connection is closed). Что означает фраза «соединение BGP разрывается»?
2. Что такое конфликт в соединении BGP? Когда он возникает и как с ним бороться?
3. Среди состояний FSM есть состояние Ative. Приведите примеры событий, переводящих FSM из состояния Active в состояния Idle, Active, OpenSent, OpenConfirm.
4. Может ли маршрутизатор, находящийся на границе AS 1, и получивший информацию о пути к сети, находящейся в AS 2, распределить эту информацию среди маршрутизаторов AS 1, используя для этого протокол внутренней маршрутизации, такой ка OSPF или RIP?
5. Согласно современной трактовке (RFC4271) AS может использовать несколько протоколов внутренней маршрутизации, а в некоторых случаях даже несколько наборов метрик в рамках одной AS. Чем обусловлено использование термина AS в таких случаях?