Лекция 6.

Назначение, разновидности и основные функции маршрутизаторов

Маршрутизатор, это устройство, работающее на 3-м уровне протоколов TCP/IP и обеспечивающие взаимодействие компьютеров и иных устройств, входящих в различные многочисленные сегменты канального уровня, образующие канальную инфрастуктуру компьютерных сетей

6.1. Обзор основных функций маршрутизаторов

К числу основных функций маршрутизаторов относятся:

- маршрутизация обычных (unicast) IP-пакетов
- динамическое изменение таблиц маршрутизации для адаптации к изменению графа связности маршрутизаторов
- маршрутизация групповых (multicast) IP-пакетов
- обеспечение удаленного мониторинга и изменения состояния маршрутизатора и его таблиц.

Маршрутизация обычных (unicast) IP-пакетов выполняется, как мы рассмотрели протокольным модулем IP на основе локальной таблицы маршрутизации.

Начальное состояние этой таблицы задается сетевым администратором и затем может динамически изменяться средствами ***протоколов управления маршрутизацией***, работающих в большинстве своем исключительно на прикладном уровне для внесения в эту таблицу изменений о маршрутах доступа к тем или иным подсетям при изменении состояния графа связности маршрутизаторов. Работа протоколов управления маршрутизации обеспечивается соответствующими демонами (процессами, обычно запускаемыми при загрузке сетевой операционной системы и всегда готовыми к выполнению своих функций). Входящий в эту группу протоколов протокол RIP подробно рассмотрен нами в рамках курса «Компьютерные сети». В рамках отдельных тем настоящей лекции нами рассматриваются также протокол OSPF и протоколы внешней маршрутизации.

Маршрутизация групповых (multicast) IP-пакетов (адрес получателя которых является адресом класса D) выполняется средствами протоколов групповой маршрутизации, работающих на прикладном уровне так же, как и протоколы групповой маршрутизации в виде соответствующего демона. Протоколы групповой маршрутизации будут рассмотрены нами в 9-й лекции настоящего курса.

Удаленный мониторинг и возможность удаленного изменения состояния маршрутизатора (включая состояние всех его таблиц и сетевых интерфейсов) обычно обеспечивается средствами простого протокола управления сетью SNMP, работающего на прикладном уровне в виде демона, называемого агентом SNMP. Указанный протокол вкратце рассматривался нами в курсе «Компьютерные сети». Для оперативного контроля и изменения состояния маршрутизатора могут использоваться также прикладные протоколы удаленного терминала, также рассмотренные нами в курсе «Компьютерные сети».

Кроме перечисленных выше основных функций маршрутизаторы могут выполнять обширный ряд дополнительных функций, которые мы рассмотрим в следующей лекции.

6.2. Разновидности технического воплощения маршрутизаторов

Вкратце рассмотрим основные разновидности технического воплощения маршрутизаторов.

В простейшем случае маршрутизатор, называемый ***ПК-маршрутизатором*** или ***PC-router***'*ом*, может быть построен на базе обычного персонального компьютера или

сервера с несколькими сетевыми интерфейсами путем установки на нем программного обеспечения стека протоколов TCP, требуемого для выполнения основных и, возможно, дополнительных функций маршрутизатора. Отметим, что поскольку реализацию 3-х из упомянутых функций функций маршрутизаторов обеспечивают протоколы, работающие на прикладном уровне стека TCP/IP, выполняется установка не только модулей канального и межсетевого уровней, но и всей реализованной в ядре операционной системы “системной части” стека протоколов TCP/IP, а также демонов необходимых протоколов прикладного уровня. Отметим, что маршрутизаторы, построенные на базе персонального компьютера или сервера не имеют сетевых интерфейсов, обеспечивающих подключение высокоскоростных каналов передачи данных, работающих на основе ряда сетевых технологий. Поэтому такие маршрутизаторы обычно применяются в периферийных подсетях, не использующих такие каналы.

На сколь-либо магистральных участках компьютерных сетей в настоящее время как правило используются так называемые ***аппаратные маршрутизаторы*** различной производительности (для более скоростных магистралей используются более производительные модели таких маршрутизаторов). Фактически аппаратный маршрутизатор представляет из себя специализированный компьютер (с процессором, оперативной памятью, портами сетевых интерфейсов, установленной операционной системой, необходимыми демонами протоколов прикладного уровня и пр.), “заточенный” на решение задач маршрутизации (как основных, рассмотренных выше, так и дополнительных, рассматриваемых в следующей лекции). Как правило, эти маршрутизаторы имеют небольшое число встроенных высокоскоростных портов семейства Ethernet и 1-2 порта WAN (Wide Area Network), обеспечивающих соединение через синхронные каналы, подключаемые через разъемы DCE/DTE. А кроме этих портов на задней панели аппаратного маршрутизатора как правило имеется достаточно большое количество так называемых слотов расширения, в которые могут вставляться интерфейсные модули различных технологий, приобретаемые отдельно для каждой конкретной инсталляции маршрутизатора. Такое техническое решение обеспечивает установку минимально необходимого для конкретной инсталляции комплекта интерфейсных модулей различных типов. В силу достаточно высокой стоимости таких модулей установка лишь действительно необходимых модулей позволяет оптимизировать стоимость всего маршрутизатора. Аппаратные маршрутизаторы конструктивно оформляются в виде более или менее плоских блоков (высотой несколько юнитов), предназначенных для монтажа в коммуникационной стойке.

Третий класс разновидностей маршрутизаторов представлен разнообразными ***специализированными маршрутизаторами***, совмещенными (в одном компактном корпусе, допускающем настольную установку) с одним или более коммуникационных устройств другого типа, связанных с собственно маршрутизатором внутренними каналами передачи данных. Примерами таких маршрутизаторов являются ADSL-router, совмещающий в одном корпусе маршрутизатор и ADSL-модем; WiFi-router, совмещающий маршрутизатор и точку доступа WiFi, а также ADSL WiFi-router, совмещающий все три перечисленных устройства. Указанные типы маршрутизаторов часто используются для создания компьютерных сетей малых офисов и домашних компьютерных сетей.

6.3. Основы протокола управления маршрутизацией OSPF

6.3.1. Недостатки протокола RIP

Протокол RIP был первым протоколом управления маршрутизацией, а, как известно, первое решение далеко не всегда оказывается самым лучшим и/или универсальным. Рассмотрим основные недостатки этого протокола.

1. Относительно большая дополнительная нагрузка на сеть. Довольно часто (раз в 30 секунд) по сети пересылаются практически полные (3 столбца) таблицы маршрутизации для каждого из маршрутизаторов сети.
2. Медленная скорость сходимости. Время сходимости возрастает квадратично с увеличением размеров сети. Поэтому весьма актуальна проблема масштабирования сетей по количеству входящих в них маршрутизаторов.

3. Невозможность распределения трафика между несколькими равноценными каналами. Трафик всегда направляется по первому из равноценных каналов, что может повлечь его перегрузку с одновременным “простоем” остальных равноценных каналов.
4. Невозможность учета уровня текущей загрузки каналов, влияющего на скорость передачи данных. Самый короткий, но перегруженный маршрут, может быть медленнее, чем более длинный, но с более низкой загрузкой из-за увеличения задержек пакетов в очередях к портам перегруженных каналов.

С учетом указанных недостатков протокол RIP применяется в настоящее время только в относительно небольших локальных сетях, основанных на достаточно быстрых и надежных каналах, возможно входящих в состав более крупных сетей, работающих с применением других протоколов управления маршрутизацией, например, OSPF.

Кроме того следует отметить, что возможный потенциал эффективности протоколов класса векторов расстояний протоколом RIP использован далеко не полностью. Более эффективными протоколами этого класса являются разработанные компанией Cisco протокол IGRP (Interior Gateway Routing Protocol - протокол маршрутизации внутренних шлюзов) и его дальнейшее развитие - протокол EIGRP (Extended IGRP - расширенный IGRP). Однако рассмотрение этих протоколов выходит за рамки нашего курса, поскольку рассматриваемый далее протокол OSPF доминирует в корпоративных сетях.

6.3.2. Общие сведения о протоколе OSPF

Протокол OSPF (Open Shortest Path First — открытый протокол выбора кратчайшего пути) практически свободен от всех недостатков протокола RIP. Слово “открытый” в названии протокола означает, что его использование не регламентируется никакими патентными ограничениями. Протокол разработан IETF (Internet Engineering Task Force — инженерным советом интернета) в 1988 году на базе созданного в 1959 году (за 10 лет до постановки задачи на создание ARPAnet) широко известным ученым-программистом Э. Дейкстрой алгоритма SPF поиска кратчайшего пути в графе. В 1991 году протокол OSPF был стандартизован в качестве RFC 1247. Текущая версия протокола OSPFv2 описана в стандарте RFC 2328. Версия протокола, предназначенная для применения в сетях IPv6 называется OSPFv6. Так же, как и RIP, протокол OSPF является протоколом внутренней маршрутизации и может использоваться только внутри сетей, находящихся под единым административным управлением и оформленных как автономные системы, рассматриваемые в параграфе 4.3.

Протокол OSPF в различных источниках относят к следующим классам (иногда упоминая лишь один из них): *протокол состояния связей* или *link-state* (этот класс протокола упоминается практически всегда) и протокол с динамической метрикой (принадлежность к этому классу упоминается реже, причину этого мы рассмотрим).

В отличие от протокола RIP протокол OSPF не пользуется услугами никаких транспортных протоколов (TCP или UDP), а работает на одном уровне с этими протоколами. Пакеты OSPF так же как и для транспортных протоколов TCP и UDP инкапсулируются непосредственно в IP-пакеты, а связь устанавливается через поле кода протокола более высокого уровня из заголовка IP-пакета. Для протокола OSPF это поле имеет значение 89. Модуль OSPF взаимодействует с соответствующим демоном прикладного уровня (gated, ospfd или др.), организующим работу протокола в целом.

Но главным отличием OSPF от протокола RIP, затрачивающего достаточно много времени как на начальное построение таблиц маршрутизаторов сети, так и на повторные перестроения этих таблиц при изменении графа связности маршрутизаторов, является то, что в протоколе OSPF как начальное создание таблиц маршрутизации, так и их последующие модификации выполняются в предельно сжатые сроки за счет использования немедленно выполняемых “наводняющих” (flooding) сеть рассылок относительно небольшого объема (по сравнению с почти полными таблицами маршрутизации в протоколе RIP) информации о состояниях связей каждого маршрутизатора со своими соседями. Эта информация используется для построения в каждом маршрутизаторе графа топологии связей между маршрутизаторами, на базе которого они с использованием алгоритма SPF строят (в начале работы) или модифицируют (при поступлении информации об изменении какой-то связи) свои

таблицы маршрутизации. Это сразу же снимает два первых из упомянутых для протокола RIP недостатков. Способы устранения двух оставшихся недостатков рассматриваются ниже.

6.3.3. Метрики OSPF

Протокол OSPF может использовать несколько различных метрик, основные из которых вкратце рассматриваются ниже. При этом для каждой из метрик в маршрутизаторах сети строятся отдельные таблицы маршрутизации, так что в каждом маршрутизаторе ведется отдельная таблица для каждой из используемых метрик. Способ выбора метрики (и соответствующей ей таблицы маршрутизации) применяемой для маршрутизации пакетов может настраиваться. В частности, для группы метрик, основанных на различных параметрах качества обслуживания (минимизации задержек, минимизации потерь пакетов) выбор может выполняться динамически для каждого IP-пакета на основе значения байта TOS этого пакета.

Для всех типов метрик полагается, что значение метрики, равное 2^{24}-1 = 16777215, не достижимо ни при каких условиях. Это значение используется в качестве признака недоступности канала или сети. Для любого типа метрик метрика маршрута является суммой метрик составляющий этот маршрут связей. И, как мы сейчас увидим при рассмотрении конкретных метрик указанное выше ограничение совершенно не достижимо.

Перейдем к рассмотрению основных типов метрик.

- ***Метрика скорости доставки пакета*** обратно пропорциональна пропускной способности канала. Обычно значение этой скорости для одной связи (иногда называемой стоимостью порта) вычисляется как отношение некоторой эталонной пропускной способности (reference bandwidth - B_R) к пропускной способности канала B: $M = B_R / B$. При этом в коммуникационном оборудовании компании Cisco (коммутаторах и маршрутизаторах) $B_R = 10^8$. Таким образом, метрика канала Fast Ethernet составит 1, а, например, очень медленного канала 64 Кбит/с – 1562,5. Недостижимое значение метрики превышает последнее указанное значение более, чем в 100 раз. Очевидно, что столь длинных маршрутов, состоящих полностью из столь медленных каналов не существует даже в масштабах всего интернета, не говоря уж о внутренней сети какого-либо оператора связи или корпоративной сети крупной организации.
  Отметим, что в оборудовании других производителей значение B_R может отличаться и составлять, например 10^9. Отметим также, что это значение B_R может быть изменено при конфигурировании маршрутизаторов.
  Рассмотренная метрика является *основной* и *практически единственной* используемой метрикой OSPF. Причины редкого использования других метрик указаны при их рассмотрении.
- ***Динамическая метрика свободной пропускной способности*** вычисляется по формуле, похожей на формулу расчета предыдущей метрики $M(t) = B_R / B_F(t)$, где $B_F(t)$ - незанятая (free) в данный момент времени t пропускная способность канала. Очевидно, что при использовании этой метрики наименьшее значение метрики будет у наименее загруженных каналов, что устраняет 4-й недостаток протокола RIP. Поскольку эта метрика явно зависит от времени, она является динамической. Именно благодаря возможности использования этой метрики протокол OSPF относят к классу протоколов с динамическими метриками.
  Очевидно также, что для сколь либо эффективной реализации механизма обновления значений этой метрики изменение значения временного параметра t должно выполняться не непрерывно, а в дискретные моменты времени, разделенные временным интервалом Δt, значительно превосходящим по величине интервал времени, требуемый для лавинообразной рассылки информации об изменении состояний *всех* связей (уровень загрузки *всех* каналов изменяется во времени) и параллельного перевычисления таблиц маршрутизации. В то же время интервал Δt не должен быть и слишком большим, поскольку с его возрастанием уменьшается уровень соответствия текущего значения метрики текущему уровню загрузки канала. Однако достаточно частые

изменения состояния всех связей влекут столь же частые периоды времени, требуемые для работ по приведению таблиц маршрутизации в актуальное состояние. В течение этих периодов времени маршрутизации может выполняться некорректно. Поэтому с учетом того, что протокол OSPF является внутренним протоколом маршрутизации и того, что в настоящее время благодаря стремительному росту пропускных способностей каналов связи каналы передачи данных как сетей операторов связи, так и подключенных к нему сетей как правило имеет существенный резерв свободной пропускной способности, потребность в использовании рассматриваемой динамической метрики фактически отпала.

Метрики параметров QoS (задержек в доставке пакетов, уровня потерь пакетов) имели смысл применения до момента разработки таких служб обеспечения качества сетевого обслуживания (Quality of Service - QoS), как IntServ (1997 год) и DiffServ (1998 год), обеспечивающих более адекватные средства обеспечения требуемого уровня QoS. Поэтому в настоящее время метрики указанных типов практически не применяются.

6.3.4. Логика работы протокола OSPF

Выше отмечалось, что при использовании протокола OSPF выполняется обмен информацией между маршрутизаторами информацией о состоянии связей (link state) каждого маршрутизатора со своими соседями и построение полной базы информации о связях всех маршрутизаторах сети. Как в сообщениях о состоянии связей, так и в базе данных информации о связях маршрутизаторы должны идентифицироваться некоторыми уникальными идентификаторами ID. В OSPF ID маршрутизатора является 4-байтным значением, которое может записываться в десятичном виде, но чаще записывается в виде IP-адреса. При этом в качестве значения ID маршрутизатора обычно выбирается минимальное или максимальное значение из IP-адресов всех интерфейсов этого маршрутизатора.

Перейдем непосредственно к описанию логики работы протокола OSPF.

Вначале (после включения маршрутизатора) этот маршрутизатор при помощи специальных пакетов HELLO определяет соседние маршрутизаторы и фиксирует в своей базе состояния связей LSDB (Link State Data Base) информацию о связях с этими соседями. Информация об одной связи включает, в частности ID двух маршрутизаторов (в данном случае - текущего и его соседа) и метрику расстояния между ними. Затем маршрутизатор при помощи специальных пакетов LSR (Link State Request - запрос о состоянии связи) запрашивает у других маршрутизаторов и получает в пакетах LSA (Link State Advertisement - объявление о состоянии связи) ответ информацию о всех связях из его LSDB и дополняет свою LSDB информацией о связях всех маршрутизаторов сети.

Одновременно с этим все соседние маршрутизаторы, обнаружив появление новой связи со вновь включившимся маршрутизатором, инициируют быструю лавинообразную рассылку информации о новой связи (пакетов LSA) всем маршрутизаторам сети с использованием специального затопляющего (flooding) протокола. Эта рассылка выполняется с максимально возможной для данной сети скоростью. Отметим что обнаружение соседними маршрутизаторами новой связи может происходить в результате получения пакетов HELLO от вновь включенного маршрутизатора, в то время как информация о пропадании связи обнаруживается в результате неполучения ответов на периодически рассылаемые самими этими маршрутизаторами сообщений HELLO.

После этого вновь включенный маршрутизатор строит размеченный граф связей между маршрутизаторами (ребра графа помечены метрикой соответствующей связи), а остальные - модифицируют свой граф связей. Завершив построение или модификацию графа связей каждый маршрутизатор совершенно независимо от других маршрутизаторов применяет алгоритм SPF для поиска кратчайших путей от себя ко всем маршрутизаторам и подключенным к ним сетям. И хотя маршрутизатор выполняет вычисление полного маршрута к каждой сети, но в строку таблицы маршрутизации заносит в качестве адреса шлюза только первый маршрутизатор вычисленного маршрута.

Отметим, что если по алгоритму SPF найдено несколько маршрутов с одинаковым (или очень близким) значением метрики, то в таблицу маршрутизации заносятся столько

же строк о разных равноценных маршрутах к одной и той же сети. А модуль IP, использующий таблицу маршрутизации для маршрутизации, обеспечивает *балансировку загрузки всех этих маршрутов* при пересылке последовательности пакетов, следующих в их общую конечную точку, путем циклического перебора строк таблицы маршрутизации. Это делает протокол OSPF свободным от 3-го недостатка использования протокола RIP, исключая чрезмерную загрузку одного из нескольких равноценных маршрутов при неполной загрузке остальных.

Вернемся к рассмотрению используемого для предельно быстрой рассылки информации об изменении состояния связей протокола затопляющих рассылок. Возможна настройка этого протокола на работу в одном из двух возможных режимов: режима использования одноточечных (unicast) пересылок или групповых (multicast) рассылок .

В режиме одноточечных пересылок, маршрутизатор, на котором произошло изменение состояния какой-либо его связи широковещательно рассылает сообщение LSA об этом изменении всем своим соседям. Каждый из соседей, получив такое сообщение, проверяет не получал ли он его ранее (по другому, более быстрому маршруту) и, если получал, просто игнорирует это сообщение. Если же сообщение получено впервые, то в соответствии с ним модифицируется локальная LSDB и сообщение отправляется далее широковещательно всем соседям, кроме того, от которого получено сообщение.

Такой способ рассылки является предельно быстрым для сетей, не содержащих широковещательных сегментов. Но в широковещательном сегменте, включающем N маршрутизаторов последовательная рассылка сообщения N-1 соседям каждым из N маршрутизаторов (всего N*(N-1) пересылок пакетов) может быть заменена одной групповой (multicast) рассылкой. Для обеспечения возможности групповых рассылок в каждом широковещательном сегменте выделяется специальный “назначенный” (designated) маршрутизатор DR и резервный (backup) назначенный маршрутизатор BDR. Обычно в качестве DR назначается маршрутизатор сегмента с наибольшим значением IP-адреса, а в качестве BDR - следующий за ним. Кроме того создаются групповые адреса всех DR и BDR (124.0.0.6) и всех маршрутизаторов сети OSPF (124.0.0.5). Тогда в каждом широковещательном сегменте множество одноточечных пересылок может быть заменено одной групповой рассылкой пакета с TTL=1 по первому адресу, а рассылка по всей сети - одной групповой рассылкой пакета по 2-му адресу. Первая рассылка практически реализуется одной широковещательной рассылкой по сегменту.

На базе рассмотренной информации можно заключить, что протокол OSPF свободен от всех недостатков протокола RIP. Тем не менее, следует отметить, что поскольку алгоритм SPF имеет высокую вычислительную сложность, квадратично возрастающую при увеличении количества машрутизаторов (вершин коммуникационного графа) OSPF-сети (так для краткости назовем сеть (автономную систему), в которой управление маршрутизацией выполняется средствами протокола OSPF), пределы масштабирования количества маршрутизаторах в OSPF-сетях все-таки являются в некоторой степени ограниченными, так что при выходе за эти пределы время перевычисления таблиц маршрутизации может стать недопустимо большим. При этом квадратичный рост сложности (в зависимости от количества маршрутизаторов в сети) предоставляет довольно ограниченные и не всегда экономически оправданные возможности расширения пределов масштабирования за счет использования более дорогих маршрутизаторов с более быстрыми процессами.

Средством существенного расширения пределов масштабируемости является разбиение всей OSPF-сети на слабо связанные зоны (area), для каждой из которых выполняется построение лишь внутризонных коммуникационных графов и таблиц маршрутизации, размер (количество маршрутизаторов) которых может быть в разы меньше общего количества маршрутизаторов OSPF-сети. При этом вычислительная сложность построения по алгоритму SPF таблиц маршрутизации на маршрутизаторах каждой из таких зон может быть на порядок (и более) меньшей, чем сложность построения таблиц маршрутизации всей OSPF-сети. Более детальное рассмотрение зонной организации OSPF выходит за рамки настоящего курса.

Краткие итоги по OSPF

Благодаря своим достоинствам по сравнению с протоколом RIP и другими протоколами класса векторов расстояний протокол OSPF стал основным и практически единственным протоколом внутренней маршрутизации в абонентских сетях и сетях мелких операторов связи. В сетях крупных операторов связи как правило используется другой протокол внутренней маршрутизации IS-IS, так же как и OSPF являющийся протоколом состояния связей, но более простой, чем OSPF. Поскольку протокол IS-IS практически не применяется в абонентских сетях, включая корпоративные сети крупных организаций, он нами не рассматривается

6.4. Внешние протоколы управления маршрутизацией

6.4.1. Неприемлемость протоколов внутренней маршрутизации для управления маршрутизацией в интерсетях

Рассмотренные в предыдущем курсе и в начальной части текущей лекции протоколы RIP (EIGRP) и OSPF (IS-IS), принадлежащие соответственно к классу векторов расстояний (distance vector) и классу состояния связей (link state) относятся к протоколам *внутренней маршрутизации* (interior gateway protocol - IGP), применимым для управления маршрутизацией лишь в сетях, находящихся под единым административным управлением (управляемых единой командой системных администраторов). К таким сетям относятся сети организаций, не являющихся операторами доступа к интернет (интернет-провайдерами), сети отдельных интернет провайдеров или объединенные сети более или менее крупных провайдеров с сетями их “вассальных” провайдеров.

В любом случае размер таких сетей по сравнению с размером интернета очень и очень невелик, что позволяет использовать для управления маршрутизацией протоколы, не отличающиеся сверхвысокой масштабируемостью (очень низкой для протоколов векторов расстояний и не высокой для протоколов состояния связей). Для протоколов векторов расстояний пределы масштабируемости критически ограничены необходимостью периодического обмена между маршрутизаторами практически полными (за исключением небольшого числа столбцов) маршрутными таблицами и медленной скоростью сходимости, обусловленной относительно большими промежутками времени между рассылкой маршрутных таблиц соседям. Для протоколов состояния связей пределы возможного масштабирования существенно шире, но и они ограничены большим объемом информации о связях всех маршрутизаторов сети, хранящейся в памяти каждого маршрутизатора и используемой для вычисления оптимальных маршрутов по алгоритму, вычислительная сложность которого квадратично зависит от количества компьютеров. Зонная организация OSPF-сетей позволяет “раздвинуть” пределы возможного масштабирования размеров таких сетей не более чем на порядок, чего явно недостаточно для сетей масштаба интернета, включавших уже в начале 1980 годов многие тысячи, а сейчас - многие миллионы маршрутизаторов. Разрыв в уровне масштабируемости, обеспечиваемом рассмотренными протоколами, и требованиями масштабов интернета является первой из главных причин того, что для управления маршрутизацией в сетях масштаба интернета должен применяться качественно иной протокол.

Второй не менее, а более важной причиной применения иного протокола маршрутизации в интернете, состоящем из множества организационно независимых сетей, является то, что в такой совокупности связанных друг с другом сетей не все физически существующие маршруты между отдельными сетями могут быть использованы для передачи данных между конкретными независимыми сетями (называемыми автономными системами) по причинам не технического а, скорее, организационно-политического характера (таких, как отсутствие договоренности между операторами связи сетей A и B о пропуске через сеть B транзитного трафика, адресованного в сеть A). В протоколах же внутренней маршрутизации все физически существующие маршруты могут использоваться для передачи данных без каких бы то ни было организационных ограничений.

Две указанные причины и послужили основанием для создания в 1982-1984 годах первого протокола внешней маршрутизации EGP (Exterior Gateway Protocol - протокол внешнего шлюза), который был впоследствие заменен на более совершенный протокол BGP (Border Gateway Protocol - протокол пограничного шлюза), который и по сей день является единственным протоколом внешней маршрутизации. Оба указанных протокола существенно основаны на использовании официально оформленных сетей, находящейся под единым административным управлением и называемых автономными системами, организация которых заслуживает рассмотрения ее отдельным пунктом.

6.4.2. Автономные системы

Согласно определению, данному в RFC 1930 *автономной системой* (autonomous system, AS) называется система IP-сетей и маршрутизаторов, управляемых одним или несколькими операторами, и имеющими единую *политику маршрутизации* с интернетом. Отметим, что в качестве AS может быть оформлена не только операторская сеть, но и сеть крупной организации (это очень полезно, если эта сеть связана с интернетом каналами нескольких операторов). Сети же нескольких операторов оформляются в качестве единой AS лишь в случае если этими операторами создана совместная единая централизованная команда сетевых администраторов, совместно управляющая объединенной сетью этих операторов. На практике это возможно, например, при объединении управления сетями некоторого ведущего оператора с одной или несколькими сетями его “вассальных” операторов.

Остановимся на понятии политики маршрутизации AS. Такая политика включает единый используемый протокол внутренней маршрутизации (для протокола OSPF он считается таковым вне зависимости от того, какие протоколы маршрутизации используются в тупиковых зонах OSPF-сети) а также политику правил взаимодействия с соседними AS. Определение указанных правил включает перечень соседних IN AS, трафик которых готова принимать определяемая AS, и перечень OUT AS, в которые определяемой AS допускается передавать исходящий трафик. Естественно, эти перечни могут пересекаться, а в случае, когда соседняя AS является единственной, оба указанных перечня по необходимости включают эту соседнюю AS. В указанных перечнях все AS идентифицируются их номерами (см. ниже). Кроме того за каждой AS закрепляется пространство IP-адресов, выделенное для всех подсетей этой AS

Автономные системы являются регистрируемыми объектами и получают свои номера при их регистрации в IANA (Internet Assigned Numbers Authority - агентство по присвоению номеров интернета), которое изначально (до разработки протокола EGP) занималось лишь распределением IP-адресов и, естественно сохранило за собой эту функцию. Отметим, что в настоящее время IANA делегирует предоставление своих функций своим 5-ти региональным представительствам (которые, в свою очередь, могут передавать часть своих полномочий локальным регистраторам IP-адресов). Для региона Европы и России региональным представительством IANA является RIPE NCC (Réseaux IP Européens (фр) + Network Coordination Centre (англ)) - сетевой координационный центр IP-сетей Европы).

При регистрации AS ей выделяется номер. Так, например, корпоративная сеть Южного федерального университета (точнее - Ростовского государственного университета, реорганизованного в Южный федеральный университет в декабре 2006 года), в 1995 году была зарегистрирована в RIPE NCC как автономная система AS 5480. Номера, выделяемые автономным системам вначале были 16-разрядными, но к 2007 году все 65533 возможных 16-разрядных номера (номера 0 и 65535 зарезервированы) были распределены и новым AS стали присваиваться 32-разрядные номера. При регистрации AS ей не просто выделяется номер. Одновременно в базе регистрирующей организации (в нашем случае - в базе RIPE NCC (или другого регионального центра соответствующего региона) сохраняется информация о политике маршрутизации AS, выделенном ей диапазоне IP-адресов, контактная информация о лицах, ответственных с решением технических и административных вопросов, связанных с регистрируемой AS и некоторая дополнительная информация.

6.4.3 Общая логика работы внешних протоколов обмена маршрутной информацией

В отличие от протоколов векторов расстояний, в которых все маршрутизаторы внутренней сети (AS) периодически обмениваются с каждым из своих соседей (и, в результате итеративного повторения этого процесса - со всеми маршрутизаторами сети) практически полными таблицами маршрутизации (векторами расстояний), и от протоколов состояния связей, в которых все маршрутизаторы внутренней сети (AS) немедленно рассылают всем другим маршрутизаторам сети информацию об изменении состояния любой из связей с каким либо из соседей, в протоколах внешней маршрутизации обмен маршрутной информацией выполняется не со всеми маршрутизаторами глобальной сети, а лишь между автономными системами, составляющими эту глобальную сеть. Другим автономным системам передается информация о проходящих через текущую AS путях к IP-сетям. При этом отдельными шагами таких путей являются не каналы между конкретными маршрутизаторами, а автономные системы, через которые проходит путь. Протоколы такого класса называются *протоколами вектора пути* (path vector protocol).

Каждая из автономных систем может быть связана с одной или несколькими другими автономными системами каналами передачи данных, соединяющими 2 пограничных шлюза (маршрутизатора) соседних AS. Количество пограничных шлюзов некоторой AS может совпадать с количеством AS, непосредственно связанной с данной (и тогда через каждый из шлюзов подключается ровно одна соседняя AS соединяется с данной через отдельный шлюз). Но через некоторые шлюзы могут подключаться (отдельными каналами) несколько AS. И тогда количество пограничных шлюзов AS может быть может быть меньшим, чем количество соседних AS.

Очевидно, что взаимодействие соседних AS осуществляется через связанные каналом пограничные шлюзы этих AS. Но для упрощения изложения материала настоящего пункта мы будем везде в этом пункте говорить, что именно AS обмениваются той или иной информацией. А детали взаимодействия пограничных маршрутизаторов соседних AS рассмотрим в следующем пункте.

Отметим, что на ранних стадиях развития интернета (1980 годы) глобальная топология возможных связей между различными AS предполагала существование единой магистральной AS, к которой могли подключаться как одиночные AS, так и древовидные структуры АС. В любом случае, топология связи AS не содержала циклов. Поэтому первый протокол внешней маршрутизации EGP не предусматривал обнаружение циклов в маршрутах (которых в принципе не могло быть). Затем рассмотренное ограничение на топологию связи AS в интернете сняли, допустив произвольные связи между AS, что сделало возможным существования множества циклов в графе топологии связей между AS, а значит - и циклических маршрутов. В связи с этим на базе протокола EGP был создан протокол BGP, способный обнаруживать маршруты, содержащие циклы и выбрасывать циклические участки маршрутов. В остальном указанные протоколы работают идентично.

Рассмотрим общую организацию работы протокола BGP. Когда автономной системе A потребуется сообщить для соседней с ней автономной системы B маршрут к принадлежащей A IP-сети (блока IP-сетей), то A выполняет действие, называемое анонсированием маршрута, а B выполняет прием сделанного анонса. Это действие выполняется не автоматически. Предварительно администраторы автономных систем A и B должны договориться о том, что A анонсирует сеть принадлежащую ей сеть N, а сеть B - принимает анонсы. Факт такого согласия означает, что AS B согласна передавать свой трафик (и, как правило, транзитный трафик других соединенных с B автономных систем) в сеть N автономной системы A; автономная же система A обязуется строго следовать широко известному сетевым администраторам правилу “Мы Вам анонсы - вы нам трафик”, то есть принимать весь трафик из автономной системы B, адресованный в сеть N. Подобная договоренность обычно оформляется соответствующим контрактом, называемым соглашением о присоединении сетей, в котором могут быть оговорены и финансовые условия такого присоединения. Отметим, что по достижении указанной договоренности сетевые администраторы систем A и B должны сконфигурировать на

пограничных шлюзах, связывающих эти сети, возможность установления так называемой BGP-сессии (сеанса связи по протоколу BGP между этими шлюзами, действующего в течение всего промежутка времени действительности анонса). Сама операция анонсирования включает пересылку маршрута (в данном случае этот маршрут имеет вид “A”), в качестве дополнительных атрибутов выступают IP-адрес и маска анонсируемой подсети.

При анонсировании своих сетей администраторы AS обычно анонсируют минимальное число адресных блоков (представляемых в форме IP-адрес-сети/маска-подсети), полностью покрывающих адресное пространство AS. Так, если адресное пространство AS включает, например, 20 подсетей класса C, то обычно это адресное пространство включает 2 смежных блока из 16 подсетей (маска /20) и 4 подсетей (маска /22).

Если администраторы автономных систем A и B договорились о том, что сделанный анонс сетей A в автономную систему B и симметричный ему анонс сетей B в автономную систему A предназначены лишь для для взаимного обмена трафиком, а канал между этими системами не используется ни одной из этих систем для выхода в интернет, то такие отношения между этими автономными системами называются пирингом (piring). Такие отношения зачастую устанавливаются между AS в точках обмена трафиком IX.

Если же администраторы системы A хотят, чтобы их канал к системе B использовался для выхода в интернет, то они должны договориться с администраторами системы B о том, чтобы те в, свою очередь, договорились с одним из своих соседей C, находящемся на пути из B к магистралям интернета об анонсировании сетей системы A в систему C. Соответствующие анонсы получаются добавлением в концу пути, идущему от A (изначально полученному из A) автономной системы B, так что результирующий путь будет иметь вид “A.B”. Далее процесс анонсирования маршрута продолжается уже системой C (по договоренности с B). Если в процессе анонсирования маршрутов текущая AS обнаруживает себя в полученном маршруте, то это означает, что анонсирование прошло по циклическому пути; в этом случае текущая AS удаляет циклический участок пути из маршрута.

Отметим, что в случае, когда сети систем A и B образуют смежный блок большего размера, анонсы сетей A и B могут быть агрегированы в один анонс, исходящий из B.

Каждая из AS (её пограничные шлюзы) запоминает все анонсированные в нее маршруты. На их основе она строит таблицу маршрутизации из кратчайших маршрутов. При этом кратчайшим всегда считается маршрут, содержащий наименьшее количество AS, которые надо пройти для достижения требуемой IP-сети. Такое количество является метрикой протокола BGP. Отметим, что в настоящее время в пограничных маршрутизаторах AS хранится информация об около полумиллионе маршрутов. Для работы с такими объемами информации необходимы мощные и дорогие модели маршрутизаторов. Поэтому неудивительно, например, почему многие мелкие операторы доступа к интернету, подключенные к сети своего “сюзерена” единственных каналом связи предпочитают включать свои сети в состав AS “сюзерена”.

Для повышения уровня надежности (бесперебойности) доступа к интернету во многих автономных системах создаются два канала доступа к магистральному интернету, один из которых как правило является основным, а второй - резервным. Основной маршрут обычно гораздо предпочтительнее резервного по таким причинам, как например пропускная способность каналов этого маршрута (особенно - канала к соседней AS), уровень оплаты трафика, пересылаемого по маршруту (для основного канала такой уровень ниже) и др. Но эти предпочтения могут оказаться в противоречии с используемой в протоколе BGP метрикой - количеством AS, указанных в пути к нашей системе (пусть это будет A). Тогда можно искусственно увеличить слишком маленькую метрику резервного канала, анонсируя соответствующему соседу свои сети с маршрутом, вида “A.A. .A”. Такая добавка в анонсируемый путь нескольких экземпляров собственной AS называется добавкой препендов.

Мы рассмотрели, как формируются таблицы маршрутизации BGP, применяемые при передаче IP-пакетов между автономными системами, не касаясь вопросов, связанных с маршрутизацией пересылаемого через некоторую автономную систему транзитного трафика, который на этом участке должен маршрутизироваться между двумя внешними

шлюзами одной и той же автономной системы. На самом деле протокол внешней маршрутизации BGP включает в себя 2 отдельных протокола: протокол eBGP (exterior - внешний BGP), обеспечивающий управление маршрутизацией между различными автономными системами, и протокол iBGP (interior - внутренний BGP), обеспечивающей управление маршрутизацией транзитного трафика, пересылаемого с одного внешнего шлюза автономной системы на другой ее внешний шлюз. Все, что рассматривалось нами выше, относится к протоколу eBGP.

В протоколе iBGP между парой внешних шлюзов одной и той же AS, используемой для передачи через эту AS транзитного трафика между двумя соседними с ней AS также устанавливается BGP сессия, в рамках которой выполняется анонсирование сетей. Поясним, как выполняется такое анонсирование и в чем состоит его эффект на примере, приведенном на рис. 1. На этом рисунке автономные системы A и B связаны через внешние шлюзы EG_{A1} и EG_{B1}, а системы B и C через внешние шлюзы EG_{B2} и EG_{C1}. Пусть автономная система B анонсирует в систему C сети системы A (маршрут “BA”) по протоколу eBGP. Тогда шлюз должен EG_{B1} анонсировать эти сети шлюзу EG_{B2} по протоколу iBGP, который средствами внутреннего протокола машрутизации, используемого в автономной системе B, должен настроить таблицы маршрутизации внутренних маршрутизаторов системы B так, чтобы пакеты с адресами получателей из подсетей AS A, поступающие извне (из AS C) на шлюз EG_{B2} маршрутизировались в шлюз EG_{B1} для дальнейшей пересылки их в AS A. А поступающие в шлюз EG_{B2} из AS C пакеты, адресованные в подсети AS A, будут маршрутизироваться этим шлюзом в шлюз EG_{B1} для их пересылки в AS A.

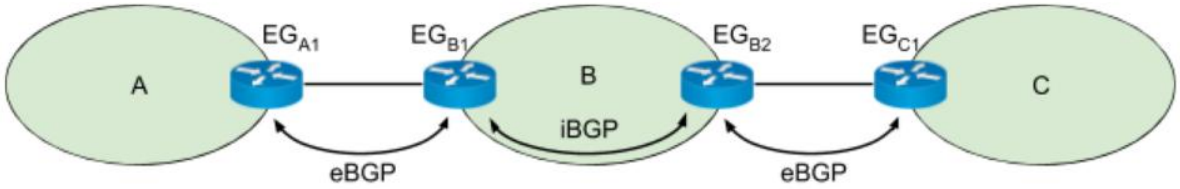

Рис.1. Схема взаимодействия AS по протоколу BGP