Междоменное групповое вещание

По мере роста магистраль MBone сталкивалась со все возрастающим числом проблем. Общей причиной проблем была плоская (не иерархическая) виртуальная топология MBone. Те же проблемы, которые были характерны для основанного на классах механизма маршрутизации уникальных адресов, проявились и в MBone. При плоской топологии сетевые маршруты должны быть известны каждому маршрутизатору, а так как во время своего расцвета MBone включала почти 10 000 маршрутов, то объемы маршрутных данных приблизились к той черте, за которой маршрутизаторы становились практически неуправляемыми.

Большинство из этих маршрутов использовалось крайне неэффективно — записи имели длинные префиксы (между 28 и 32), а это означало, что каждая запись в таблице маршрутизации представляла всего несколько хостов.

Кроме того, при плоской топологии ошибки одного маршрутизатора распространялись по всей сети, вызывая нестабильную работу Mbone в целом.

Опыт маршрутизации уникальных адресов говорил о необходимости применения в групповом вещании механизма агрегирования маршрутов и создания на этой основе иерархической топологии. Действительно, когда-то развитие обычной (использующей уникальные адреса) маршрутизации привело к разделению Интернета на автономные системы (AS), представляющие собой домены маршрутизации. Как отмечено в главе 17, каждой автономной системой управляет одна организация, которая вольна организовывать маршрутизацию внутри этой системы так, как считает нужным, используя для этого протокол RIP, OSPF, IGRP или статические записи в таблицах маршрутизации. По умолчанию считается, что между автономными системами нет отношений доверия, поэтому маршрутная информация через границы AS передается под жестким надзором администраторов, держащих под контролем передачу через свой домен транзитной информации чужих доменов.

Так как структура автономных систем Интернета уже сложилась, при построении иерархической топологии групповой маршрутизации не требовалось создавать новую структуру, достаточно было разработать механизм, способный работать с существующей структурой, основанной на доменах маршрутизации.

Задача создания протоколов междоменной групповой маршрутизации была поставлена сетевым сообществом в 1997 году. На сегодняшний день существуют две группы решений. Первая группа — это так называемые тактические решения, которые могут работать уже сегодня, но не обладают достаточной масштабируемостью, чтобы стать основой развития Интернета на значительную перспективу. Поэтому продолжается работа по поиску долговременных стратегических решений, составляющих вторую группу. Стратегические предложения базируются как на стандартной модели группового вещания протокола IP, так и на более радикальных новых подходах.

Протоколы PIM-SM и BGP в многодоменной сети группового вещания

Примером тактического решения является создание средств маршрутизации группового трафика в многодоменной сети путем расширения функциональных возможностей широко используемого протокола маршрутизации BGP. Действительно, поскольку ставится задача разработки протокола групповой маршрутизации между автономными системами, то естественно в первую очередь обратить внимание на протокол, который уже долгое время успешно служит для маршрутизации трафика с индивидуальными адресами между автономными системами, то есть на протокол BGP (см. главу 17).

Механизм, с помощью которого протокол BGP был доработан для поддержки групповых маршрутов, представляет собой расширение протокола BGPv4. Это расширение, описанное в спецификации RFC 2283, входит в состав протокола BGP4+ (Multiprotocol Extensions for BGPv4 — мультипротокольные расширения для BGPv4). Когда протокол BGP4+ используется для поддержки групповых маршрутов, его часто называют MBGP (Multicast Border Gateway Protocol). Протокол MBGP выполняет достаточно простую функцию для системы группового вещания: он узнает о путях, с помощью которых групповой трафик может достичь других доменов. На рис. 1 показан пример соединения нескольких доменов в сеансах MBGP. В одном случае два домена, связанные вместе, используют разные соединения для индивидуального и группового трафиков, в других случаях — общие соединения для передачи обоих типов трафика.

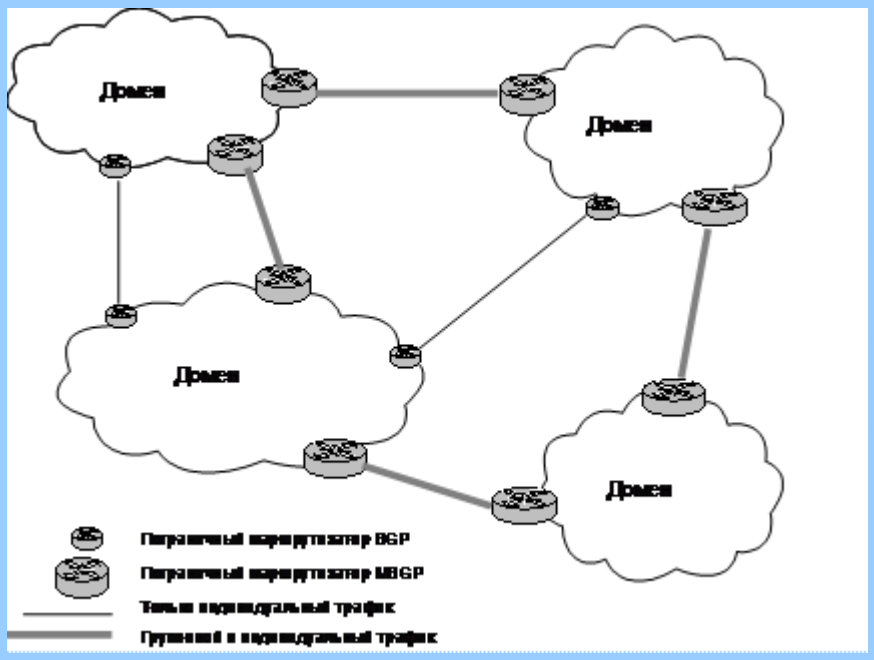

Рис. 1. Пример междоменной топологии группового вещания на базе протоколов BGP и/или MBGP

Разработка протокола MBGP — это только первый шаг на пути реализации группового вещания, и его использование не обеспечивает решение проблемы полностью. Протокол MBGP может выбрать следующий шаг передачи пакетов с индивидуальными и групповыми адресами, но он не может выполнить более сложную функцию — построить дерево группового вещания. Здравый смысл подсказывает, что для этих целей может быть применен только протокол разряженного режима.

Таким тактическим решением стало использование протокола PIM-SM на междоменном уровне, когда домены рассматриваются в качестве узлов сети, а разделяемое дерево строится для тех доменов, в которых содержатся члены определенной группы. Однако полностью повторить на этом уровне протокол PIM-SM не удается из-за междоменных барьеров.

Протокол MSDP

Протокол обнаружения источников группового вещания (Multicast Source Discovery Protocol, MSDP), описанный в спецификации, RFC 3618, позволяет решить еще одну проблему, возникающую при соединении доменов, на которых работают протоколы «разреженного» режима. Проблема в том, как оповестить членов групп в одном домене, что в других доменах есть источники группового вещания для этих групп.

Стандартное решение протокола PIM-SM состоит в отправке источником точке встречи регистрационного сообщения, в ответ на которое RP отсылает источнику сообщение о присоединении, означающее, что источник присоединяется к разделяемому дереву группы. При междоменном взаимодействии возникает вопрос: в каком домене нужно размещать точку встречи? При существовании единственной точки встречи весьма вероятно возникновение так называемой «зависимости от третьей стороны», когда все источники и приемники группы расположены в одном домене, а точка встречи этой группы — в другом. Такая ситуация может привести к организационным и коммерческим конфликтам.

Поставщик услуг автономной системы с источниками и приемниками группы должен полагаться на другого провайдера — своего потенциального конкурента, в домене которого размещена точка встречи.

Поставщик услуг автономной системы, включающей точку встречи, обслуживает трафик группы, для которой у него нет ни источников, ни получателей. В большинстве случаев при отсутствии членов группы нет и финансовой мотивации для обслуживания трафика конкурента.

Более приемлемым решением может показаться дублирование точек встречи в каждом домене. Однако это ведет к другой проблеме — проблеме организации взаимодействия между этими точками, которой не было при работе протокола PIM-SM в пределах одного домена. Именно эту задачу решает протокол MSDP. Протокол MSDP работает в каждом домене на том же маршрутизаторе, что и точка встречи домена. Он выполняет функции представителя домена, объявляющего другим доменам о существовании активных источников групповых данных. Все протокольные модули MSDP, установленные на маршрутизаторах, принадлежащих разным доменам, связаны TCP-соединениями. Алгоритм работы протокола MSDP иллюстрирует рис. 2.

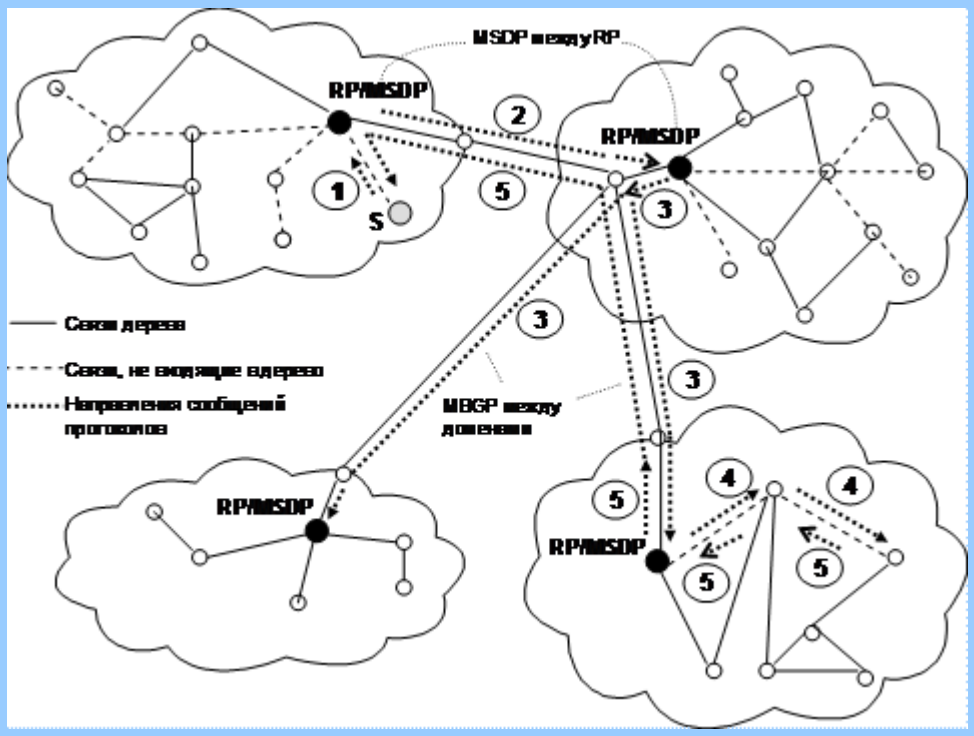

Рис. 2. Схема работы протокола MSDP

Источник генерирует MSDP-сообщение.

MSDP-партнеры, которые получают сообщение, обеспечивают проверку продвижения по реверсивному пути (RPF), то есть проверяют, находится ли пославший это сообщение партнер на «корректном» пути. Когда новый источник для группы становится активным, он регистрируется в соответствующей точке встречи домена.

Протокол MSDP, работающий в данном домене, обнаруживает возникновение нового активного источника и посылает сообщение об активности источника (Source Active, SA) всем узлам MSDP, расположенным в других доменах. После этого данное SA-сообщение распространяется периодически. Если MSDP-партнер получает SA-сообщение через корректный интерфейс, то оно передается всем MSDP-партнерам за исключением того, от которого это сообщение получено.

Внутри домена MSDP-партнер, который является точкой встречи, проверяет, есть ли у него члены группы в домене.

Если члены группы в домене имеются, то RP посылает сообщение протокола PIM о присоединении по адресу источника, объявленному в SA-сообщении (что и было бы сделано точкой встречи в том случае, если бы источник находился в ее домене). Сообщение о присоединении проходит реверсивный путь до источника, используя междоменные пути, определяемые с помощью протокола MBGP. Как только реверсивный путь проложен, точки встречи начинают продвигать групповые пакеты, в том числе между доменами.

Повторение шагов 3 и 4 происходит до тех пор, пока все MSDP-партнеры не получат SA-сообщение и все члены группы во всех доменах не начнут получать данные от источника.

Ограничения и проблемы протоколов PIM-SM/MBGP/MSDP

В ответ на критику, утверждающую, что текущий набор протоколов не отличается простотой, сторонники решения PIM-SM/MBGP/MSDP возражают, что он не более сложен, чем многие решения масштабе всего Интернета. Основное преимущество решения PIM-SM/MBGP/MSDP состоит в том, что оно работает и уже успешно применяется. Основным недостатком является то, что при частых изменениях в составе групп и активности источников масштабируемость этого решения вызывает сомнения.

Когда группы являются динамичными — либо по причине частых изменений в активности источников (пульсирующие источники), либо из-за частых случаев присоединения/отсоединения членов групп — накладные расходы на управление группой могут быть значительными.

У сети возникает трудная задача создания и удаления информации о состояниях тысяч приемников и передатчиков, разбросанных по всему миру.

Кроме того, еще одной специфической проблемой, с которой сталкивается протокол MSDP, является проблема задержки присоединения. Так как SA-сообщение распространяется периодически, то между присоединением новых приемников и получением ими следующего SA-сообщения может возникать значительная задержка. Для решения этой проблемы MSDP-партнеры могут кэшировать SA-сообщение в надежде то, что в присоединения нового приемника источник все еще будет активен. Если MSDP-партнер кэширует SA-сообщение, то другие MSDP-партнеры могут воспользоваться кэшируемой информацией. Некэширующий узел MSDP может послать SA-сообщение запроса узлу MSDP, который выполняет кэширование. К сожалению, минимизация задержки присоединения за счет кэширования приводит к чрезмерному увеличению объема хранимой информации состояния.

Другая проблема вызывается пульсирующими источниками. После проявления активности новым источником проходит значительное время, связанное с установлением между доменами дерева продвижения. Поэтому несколько первых пакетов, передаваемых источником, часто теряются. Решение, предусмотренное в протоколе MSDP для устранения этого недостатка, состоит в том, что перенос первых пакетов данных может осуществляться служебными SA-сообщениями. Это не очень элегантное решение, но оно работает.

Вопрос масштабируемости достаточно важен для MSDP. Принимая во внимание способ работы MSDP, при существовании в сети нескольких тысяч групп накладные расходы на функционирование MSDP становятся слишком большими: в сети постоянно будет циркулировать большое количество SA-сообщений, содержащих групповые данные. Общее мнение состоит в том, что учитывая недостаточную масштабируемость протокола MSDP, его вряд ли можно отнести к стратегическим решениям.

Протоколы BGMP и MASC

Первым реально долговременным предложением по организации междоменного группового вещания в масштабах Интернета стал **протокол BGMP** (Border Gateway Multicast Protocol — пограничный шлюзовой протокол группового вещания). Этот протокол относится к группе, сохраняющей стандартную идеологию группового вещания, предложенную Дирингом. Ключевой идеей BGMP является конструирование двунаправленных разделяемых деревьев между доменами при наличии только одной точки входа. Одной из функций BGMP является принятие решения о том, что какой домен будет служить корнем разделяемого дерева. Предлагаемое решение свободно от проблемы «зависимости от третьей стороны», так как протокол основан на более строгой схеме распределения адресов.

Вообще говоря, распределение групповых адресов стало серьезной проблемой для коммерческих пользователей группового вещания. Протокол BGMP учитывает эту проблему, требуя, чтобы групповые адреса были связаны с определенным доменом. Архитектура BGMP включает собственную схему распределения адресов под названием MASC (Multicast Address-Set Claime), однако может работать и с любой другой, лишь бы она обеспечивала связь групповых адресов с определенной автономной системой.

Помимо описанного требования существует еще одно — необходимо избегать групповых коллизий. Коллизия возникает, когда две группы используют один и тот же групповой адрес, и трафик каждой группы доставляется членам обеих групп. Эффект коллизии групп может проявляться по-разному: от простого неудобства до серьезного нарушения работы сети. Применяемая сегодня схема назначения групповых адресов (имеющих локальное значение) является неформальной: пользователь может просто взять адрес и применить его, то есть вероятность коллизии не нулевая. Именно поэтому процедура распределения групповых адресов приобрела особое значение.

Протокол MASC представляет собой одно из возможных решений и является частью более общей схемы адресации, называемой архитектурой распределения групповых адресов. Имеется три уровня распределения адресов:

- между доменами адреса распределяются по схеме MASC;
- внутри домена распределением адресов занимается протокол распределения адресов (Address Allocation Protocol, AAP);
- протокол MADCAP (Multicast Address Dynamic Client Allocation Protocol) используется хостами для запроса адресов у сервера распределения групповых адресов (Multicast Address Allocation Server, MAAS).

Таким образом, MASC и другие вспомогательные протоколы выполняют функции динамического распределения групповых адресов, необходимые протоколу BGMP. Однако этот подход не является единственно возможным.

Другой подход состоит в статическом распределении групповых адресов. Например, в соответствии с процедурой распределения адресов GLOP каждая автономная система получает фиксированное количество адресов, которые называются GLOP -адресами и в которые как часть входит номер автономной системы. Это предложение становится

довольно популярным, но оно имеет уязвимость. В текущем варианте только 8 бит, или 256 групповых адресов, доступны для автономной системы, что во многих случаях явно недостаточно. Эта проблема может быть решена переходом на систему адресации протокола IPv 6.