

# Лекция 9

## Протоколы групповой маршрутизации

В курсе «Компьютерные сети» при рассмотрении классов IP-адресов упоминался класс D с признаком начала класса “1110” и диапазоном возможных значений 224.0.0.0 - 239.255.255.255. Это класс групповых адресов, содержащих единственное (кроме признака класса) 28-разрядное поле, значением которого является уникальный в интернете номер группы компьютеров. Рассмотрение того, что представляют из себя такие группы, и как обеспечивается доставка IP-пакетов групповому получателю, является предметом настоящей лекции.

### 9.1. Модель и общая организация группового вещания

Групповое или multicast вещание, это вещание “один ко многим” или “многие ко многим” в отличие от одноадресного (unicast) вещания “один к одному” или широковещания (broadcast) “один ко всем”. Потребность в групповом вещании возникает, например, при проведении на базе соответствующих служб компьютерных сетей селекторных совещаний, видеоконференций и других подобных мероприятий, требующих поочередной (дуплексной) или одновременной (симплексной) рассылки информации соответственно от одного или нескольких участников мероприятия всем остальным его участникам (примером таких мероприятий являются и онлайн-лекции) .

Если такую рассылку, предназначенную всем участникам, выполнять по их unicast адресам, то на начальных, возможно общих для всех участников мероприятия шагах маршрута доставки, вместо каждого пакета с общей для всех получателей информацией потребуется передавать N копий такого пакета, где N - количество участников - получателей информации. Эта ситуация иллюстрируется на рис.1.

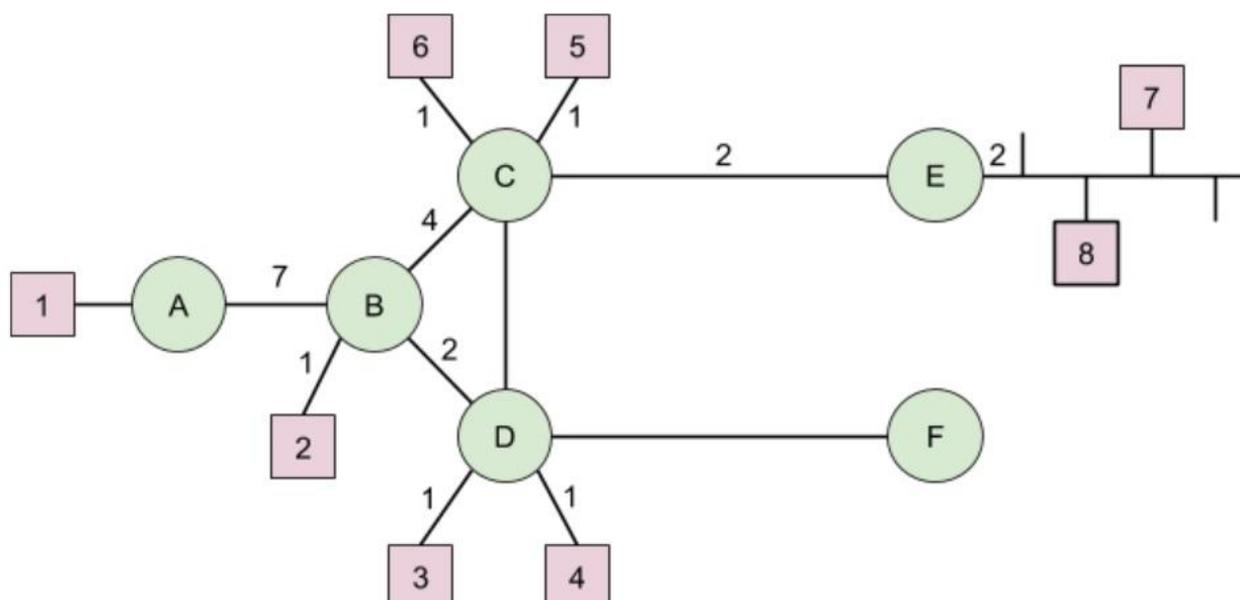


Рис. 1. Иллюстрация рассылки многочисленных экземпляров идентичных пакетов при организации групповой рассылки на базе unicast рассылок

На этом рисунке к 6-ти маршрутизаторам сети (A - F) подключены компьютеры 8-ми участников мероприятия. Тогда при рассылке каждого IP-пакета 1-м участником (подключенным к маршрутизатору A) остальным семи участникам между маршрутизаторами A и B будут передаваться 7 пакетов (это отмечено соответствующей пометкой над соединяющим эти маршрутизаторы каналом); между маршрутизаторами B и C - 4 пакета; между парами маршрутизаторов B и D, а также C и E - по 2 пакета.

Очевидно, что такая организация рассылки крайне не экономна. Оптимальным решением является следующее. Маршрутизатор A пересылает маршрутизатору B единственный пакет, который “разветвляется” маршрутизатором B в трёх направлениях: к компьютеру участника 2 и к маршрутизаторам C и D. Маршрутизатор C также “разветвляет” пакет, направляя его маршрутизатору E и участникам 5 и 6. Маршрутизатор D направляет 2 копии каждого пакета участникам 3 и 4. А маршрутизатор E, к которому участники 7 и 8 подключены через один и тот же сегмент канального уровня, направляет каждый полученный групповой пакет в указанный сегмент в широковещательном режиме.

Обеспечение рассылки информации группе получателей по подобной оптимальной схеме выполняется протоколами групповой маршрутизации. Для поддержки этих протоколов на маршрутизаторах должен устанавливаться демон `mroute`. Подсети, на всех на маршрутизаторах которых установлен этот демон,

называются **областями группового вещания**. Совокупность областей группового вещания, связанных между собой через IP-туннели, образуют **магистраль группового вещания Mbone**.

Как говорилось выше мультикастовая группа адресуется адресом класса D. **Каждый из участников группы** в принципе **может выступать как в роли источника данных, рассылаемых всем остальным участникам группы, так и в роли получателя данных**, рассылаемых остальными участниками этой же группы. Однако в состав группы могут входить и **пассивные участники (наблюдатели)**, никогда не выступающие в роли источника рассылки данных.

При этом состав группы удовлетворяет двум следующим свойствам.

Во-первых, **состав группы является открытым**: каждый из включенных в группу компьютеров ничего не знает о других участниках группы, кроме их группового адреса.

И, во-вторых, **состав группы является динамическим**: в любой момент времени к группе могут подключаться новые участники, а кто-то из "старых" участников выходит из ее состава.

**Технология группового вещания была предложена Стивом Диригом в 1988 году. К 1992 году на базе этой технологии была создана экспериментальная магистраль группового вещания Mbone, объединившая участников из 20 сетей интернета. С использованием этой магистрали была впервые проведена аудиоконференция, позволившая транслировать ход проводимого в Сан-Диего собрания инженерного совета сети Интернет IETF для участников этого совета, расположенных в разных странах мира.** На этой стадии развития Mbone все конечные точки туннелей располагались непосредственно на компьютерах участников конференций. Это позволяло легко связывать в единую магистраль любые пары компьютеров вне зависимости от того, в каких автономных системах они находились. И **лишь к 1997 году был разработан протокол междоменного группового вещания, позволивший создавать Mbone как совокупность связанных областей группового вещания** (а не отдельных компьютеров), расположенных в нескольких автономных системах.

В состав протоколов групповой маршрутизации входят

**протокол IGMP определения наличия членов мультикастовых групп** в подсетях, непосредственно подключенных к маршрутизатору и **ряд протоколов собственно групповой маршрутизации**. К числу последних относятся

- протокол DVMRP класса векторов расстояний использующий в своей работе таблицы маршрутизации RIP;

- протокол MOSPF состояния связей, использующий таблицы OSPF,

- независимый от протокола управления обычной маршрутизацией протокол PIM

и междоменный протокол групповой маршрутизации BGMP.

Алгоритмы работы протоколов групповой маршрутизации существенно разнятся от "режима заполнения" магистрали MBone членами группы. В число этих режимов входят режим плотного заполнения Dense Mode (DM), при котором члены группы подключены к очень многим или даже к большинству маршрутизаторов MBone, и режим разреженного заполнения Sparse Mode (SM), при котором члены группы подключены к редким из маршрутизаторов MBone.

Отметим, что при проведения аудио и видеоконференций количество членов группы обычно не очень велико, а масштаб сети, соединяющей всех членов группы могут быть весьма большим. Так в крупнейших конференциях, в которых довелось участвовать автору, и проводимых Минобрнауки РФ было свыше 90 участников (по одному участнику из региональных министерств образования и еще несколько участников). При этом сеть, к которой были подключены эти участники, имела масштаб страны. Таким образом, режим заполнения сети участниками был более чем разреженным. Ввиду типичности этого примера можно прийти к выводу, что при проведения широкомасштабных групповых мероприятий режим SM является гораздо более востребованным, чем режим DM.

Алгоритмы группового вещания для режима плотного заполнения мы рассмотрим на примере протокола DVMRP, не предусматривающего специальных эффективных алгоритмов работы для режима разреженного заполнения. На примере протоколов MOSPF и PIM-SM рассмотрим различные методы организации групповых рассылок для разреженных групп. Рассмотрение протокола междоменной групповой маршрутизации BGMP выходит за рамки настоящего курса.

## 9.2. Протокол IGMP

Протокол IGMP (Internet Group Management Protocol) межсетевой протокол управления группами) предназначен для определения наличия членов различных групп в подсетях, непосредственно подключенных к каждому маршрутизатору

областей группового вещания. Для такого определения маршрутизатор выполняет следующие действия:

1. Маршрутизатор периодически рассылает **запросы HMQ (Host Membership Query)** во все подключенные к нему сегменты **в режиме широковещательной рассылки**. Это выполняется путем широковещательной рассылки группового пакета **по групповому адресу 224.0.0.1 с TTL=1**. Эти пакеты доставляются только конечным компьютерам. Любой промежуточный маршрутизатор их отбрасывает (TTL=1). Периодическая рассылка запросов (и обработка ответов на них) позволяет учитывать динамическое изменение состава групп.
2. Компьютер по получении HMQ отправляет маршрутизатору **(но широковещательно)** сообщение **HMR (Host Membership Report)** для каждой группы, членом которой он является. При этом различные сообщения HMR посылаются **с некоторой задержкой** относительно предыдущих, *чтобы не создавать пиковый трафик*. **Если компьютер, обнаруживает, что HMR посылаются и другими компьютерами того же сегмента** (а обнаружить это он в состоянии, поскольку HMR посылаются широковещательно) – **задержки между пакетами дополнительно увеличиваются**.
3. Маршрутизатор по получении HMR **не запоминает никакой информации о приславшем его компьютере**. Вместо этого он **запоминает ТОЛЬКО информацию о наличии членов различных групп в КАЖДОМ из подключенных к нему сегментов**. Тогда при получении IP-пакета с групповым адресом получателя этот пакет широковещательно направляется в каждый из сегментов, в которых есть члены группы.

### 9.3. Протокол DVMRP

При использовании протокола **DVMRP (Distance Vector Multicast Routing Protocol)** - протокол групповой маршрутизации класса векторов расстояний), как и для других (не всех!) протоколов групповой маршрутизации **для каждого члена группы, который может выступать как источник групповой рассылки, создается остовое дерево (spanning tree) сети маршрутизаторов**, через которые выполняется такая рассылка **ко всем получателям**. Существует несколько методов построения остового дерева, один из которых был рассмотрен нами при обсуждении протокола STP (Spanning Tree Protocol) в лекции 2.

В DVMRP применяется другой метод, называемый *широковещанием с отсечением* (broadcast and prune), основанный на изначально широковещательной рассылке данных по всем маршрутам от источника с последующим отсечением избыточных длинных маршрутов, для определения которых используется метод RPF (Reverse Path Forwarding - “продвижение” по обратному пути). Рассмотрим укрупненный алгоритм этого метода.

1. **Источник групповых данных (IP-пакетов) широковещательно рассылает их по сегменту** (*широковещательный адрес используется на канальном уровне*, в то время как в качестве адреса получателя в IP-пакетах указывается групповой адрес, группу, определяемую этим адресом ниже, будем называть *текущей*). При этом групповые пакеты доставляются как к другим членам группы, расположенных в том же сегменте (а длина таких сегментов при использовании современных технологий их построения может быть сколь угодно большой), так и маршрутизаторам, связывающим эту сеть с остальными сетями.
2. **Каждый маршрутизатор**, получивший пакет не от источника, находящемся в собственном сегменте, проверяет, **проходит ли через интерфейс, с которого был получен пакет через кратчайший путь от источника к данному маршрутизатору**. Для этого он **на основе таблицы маршрутизации протокола RIP находит строку, соответствующую кратчайшему пути от себя к источнику** (в этом и состоит “продвижение” по обратному пути RPF). **Если значение поля идентификатора интерфейса в этой строке отлично от идентификатора интерфейса, с которого был принят пакет** (так будет, например на рис. 4.4 для пакета, принятого маршрутизатором С от маршрутизатора В, или для пакета, принятого В от С), **то путь, по которому следовал этот пакет не является кратчайшим**, данные, принятые по этому пути надо игнорировать и, кроме того, **надо “отсечь” этот путь от остового дерева путем отсылки отсекающего сообщения *prune* в интерфейс, через который был принят пакет**.

*Маршрутизатор, принявший такое сообщение через некоторый интерфейс должен пометить соответствующий интерфейс, как исключенный из широковещательной рассылки для групповых сообщений текущей группы.*

**Если же путь получения пакета является кратчайшим** путем от источника, **то пакет широковещательно направляется во все интерфейсы**

маршрутизатора, кроме следующих: того интерфейса, с которого поступил пакет; интерфейсов к тупиковым сегментам (не содержащих других маршрутизаторов), которые по данным протокола IGMP не содержат членов текущей группы, и интерфейсов, отсеченных от передачи сообщений текущей группы ранее выполненными сообщениями prune.

3. В результате выполнения предыдущего шага алгоритма для всех маршрутизаторов, в которые поступает групповой пакет, этот пакет достигнет тупикового маршрутизатора, ко всем интерфейсам которого, кроме того, с которого был принят групповой пакет, подключены тупиковые сегменты. Если хотя бы в одном из этих сегментов зарегистрированы члены текущей группы, то пакет направляется во все такие интерфейсы широковещательно.

Если таких сегментов нет (как для маршрутизатора F) - то в интерфейс, из которого был получен пакет, направляется отсекающее сообщение prune.

Отметим, что информация об отсечении каналов связи между маршрутизаторами периодически сбрасывается и остовое дерево строится вновь. Это позволяет поддерживать соответствие структуры остового дерева текущей конфигурации динамически изменяемой конфигурации мультикастовой группы.

Рассмотренный алгоритм построения остового дерева эффективно работает в сетях, плотно насыщенных членами группы. В случае же разреженного размещения членов сколь либо большого числа групп волны распространения множества широковещательных фронтов построения остовых деревьев способны чрезмерно загрузить сетевые каналы и маршрутизаторы "паразитным" трафиком широковещательной рассылки групповых сообщений в подсетях, не содержащих членов тех или иных групп.

## 9.4. Протокол MOSPF

Протокол MOSPF (Multicast Extensions to OSPF - мультикастовые расширения к OSPF, стандартизованы в RFC 1584) для поддержки группового вещания использует несколько расширенные обычные механизмы протокола OSPF, такие как распространение информации о связях между соседними маршрутизаторами во все маршрутизаторы OSPF сети (с возможным разбиением этой сети на зоны с ограниченным межзональным обменом маршрутной информации и самостоятельное построение в каждом из этих маршрутизаторов графа связей и создаваемой на его основе таблицы маршрутизации).

Мультикастовые расширения протокола OSPF состоят в том, что при рассылке информации о связях каждого маршрутизатора к этой информации добавляются данные о состоянии членства в мультикастовых группах подсетей, непосредственно подключенных к этому маршрутизатору. Эта информация рассылается в специальном сообщении **Group Membership**. Таким образом, каждому маршрутизатору **MOSPF** дополнительно к топологии связей маршрутизаторов OSPF сети (или зоны такой сети) становится известным состояние получателей данных по каждой группе во всей этой сети (или ее зоне). На базе этой информации маршрутизатор строит остовые деревья кратчайших путей каждой группы. В результате необходимость в широковещательных рассылках полностью отпадает: групповые сообщения любого члена группы сразу распространяются по дереву кратчайших путей от него до всех других членов группы.

В силу рассмотренного метода его работы **протокол MOSPF хорошо** (без накладных расходов на широковещательные рассылки) **работает как для плотного, так и для разреженного режимов заполнения сети членами группы.**

## 9.4. Протокол PIM

Протокол **PIM (Protocol Independent Multicast** - независимая от обычного протокола управления маршрутизации мультикастовая рассылка) фактически представляет из себя **два совершенно независимых протокола PIM-DM и PIM-SM**, предназначенных для работы с группами соответственно плотно (DM) или разреженно (SM) заполняющих сеть.

Алгоритм работы протокола **PIM-DM** совершенно аналогичен алгоритму протокола DVMRP с той лишь разницей, что **для выполнения процедуры определения кратчайшего маршрута к источнику пакета по методу RPF могут использоваться не обязательно таблицы маршрутизации RIP - вместо них могут использоваться маршрутные таблицы любых других протоколов** обычной маршрутизации.

**Протокол PIM-SM** также может работать с маршрутными таблицами любых обычных (юникастовых) протоколов управления маршрутизации. Но, **в отличие от протоколов DVMRP и PIM-DM, он строит для каждой группы не N остовых деревьев** (где N - количество источников данных в группе) **а одно единственное дерево, корень которого обычно расположен где-то "вблизи центра" сети,**

соединяющих участников группы. Такое дерево называется разделяемым (shared) остовым деревом.

В качестве корня разделяемого дерева выбирается специально сконфигурированный маршрутизатор, называемый точкой встречи RP (Rendezvous Point - точка рандеву). От этой точки ведут пути к каждому члену группы, так что пути от конкретного источника ко всем получателям являются сочленениями пути от этого источника к RP и путей, ведущих от RP к получателям. Точек RP в сети может быть несколько, так что для разных групп могут использоваться различные RP. В этом случае каждая из RP при помощи специального протокола конфигурируется для работы с определенными мультикастовыми группами.

Рассмотрим алгоритм построения разделяемого остового дерева (для одной мультикастовой группы), применяемый в протоколе PIM-SM. Это алгоритм включает два крупных шага.

1. Каждый получатель групповой информации направляет в RP по ее уникальному адресу специальный пакет *присоединения* к группе, который, естественно, маршрутизируется с использованием обычных таблиц маршрутизации (созданных любым из протоколов управления маршрутизации). На каждом из промежуточных маршрутизаторов, которые пакет *присоединения* проходит на своем пути к RP, для указанной в этом пакете группы фиксируется направление продвижения ее мультикастовых пакетов, совпадающее с тем направлением, откуда прибыл пакет. При прохождении через такой маршрутизатор пакетов *присоединения* от других членов группы, пришедших с других направлений, будут зафиксированы соответствующие дополнительные направления продвижения мультикастовых пакетов группы. Таким образом, к моменту прихода в RP пакетов *присоединения* от всех получателей группы будет сформировано остовое дерево доставки из RP мультикастовых сообщений всем получателям группы.

2. Каждый источник (точнее - ближайший к нему маршрутизатор) посылает в RP по ее уникальному адресу пакет, называемый пакетом *регистрации* и содержащий в качестве данных инкапсулированный мультикастовый пакет.

Маршрутизатор RP, получив такой пакет, извлекает из него мультикастовый пакет и проверяет, зарегистрированы ли у него какие-либо получатели соответствующей группы, отличные от источника пакета

регистрации. *Если таких получателей нет, пакет регистрации просто игнорируется. В противном случае инкапсулированный групповой пакет рассылается по сформированному для соответствующей группы остовому дереву, а источнику инкапсулированного пакета направляется пакет сообщения о его присоединении к группе.*

*Когда ближайший к источнику пакета регистрации маршрутизатор получает сообщение о присоединении, он может направлять групповые данные непосредственно по их групповому адресу, ибо ветвь остового дерева, ведущая от источника мультикастового трафика к RP, уже была сформирована первым шагом алгоритма, когда источник группового трафика выступал в роли его получателя.*

В завершение отметим, что протокол PIM-SM в настоящее время является наиболее распространённым протоколом групповой маршрутизации.