Multicast Intra-domain Mini Workshop

@Networkshop 2006

Stig Venaas (UNINETT)
stig.venaas@uninett.no
Tim Chown (Univ. Southampton)
tjc@ecs.soton.ac.uk

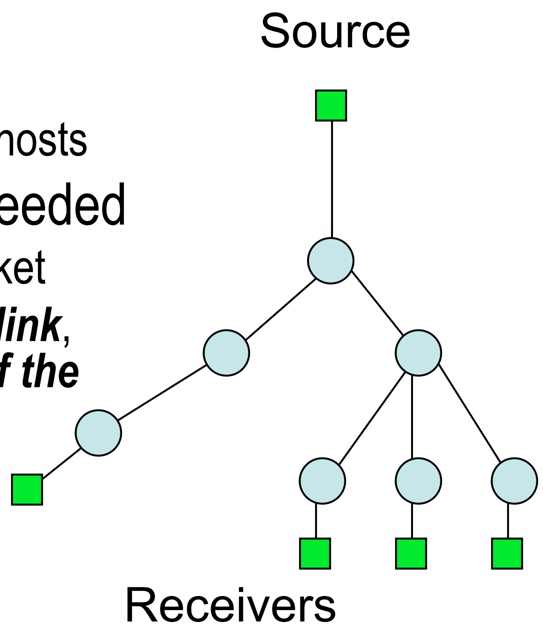
Overview

- Introduction and multicast on the LAN
  - What is multicast and what can it be used for?
  - Addressing and multicast on the LAN, IGMP

- Intradomain multicast
  - Multicast routing – PIM-SM
  - Deploying multicast in a network, e.g. a site

What is IP multicast?

- Usually an IP packet is sent to one specific host
  - The IP destination address specifies which host
- With IP multicast, an IP packet is sent to a group of hosts
  - The IP destination address is a group and not a host address
  - IPv4 multicast addresses, class D. 224.0.0.0 – 239.255.255.255
  - The group can contain any number of hosts (0 to infinity)
  - The group members can be anywhere
  - Like IP subnet broadcast:
    - A single packet is received by all on the subnet.
    - Multicast is not restricted to the subnet, and not sent to all hosts
- Multicast packets will be replicated by routers where needed
  - Routers keep track of which interfaces should forward the packet
  - The same multicast packet is ***never sent twice on the same link***, hence the bandwidth used on a specific link is ***independent of the number of receivers***



Why is it useful?

- Imagine the BBC streaming TV on the Internet to every UK home
  - Multicast only needs a basic machine and typical home Internet connectivity
  - Remember, to send you don’t need more bandwidth than a single receiver
- An ADSL user could send video to thousands of other users
  - The number of receivers is not an issue
- Useful for multi-party applications (conferencing or gaming)
  - Where each participant wants to send the same data to all others
- For financial and gaming applications
  - It may be important to deliver quickly and simultaneously to many recipients
- Multicast also useful for discovery
  - Imagine all printers on your network joining a specific multicast group
  - Can query all printers (and no other hosts) asking them to identify themselves

Service Model and Routing Challenges

- The basic multicast service model is:
  - Anyone can send to the multicast group
    - Senders don't need to know where the receivers are or how many (if any)
  - Hosts interested in the group join it
    - They don't need to know who is sending, where they are, or what other receivers there are
    - They just receive anything sent by anyone to the group while they are members

- Source-Specific Multicast (SSM, RFC 3569)
  - SSM is a new model where receivers specify the sources when joining
  - i.e. receivers need to know who the sources are
- The big challenge is routing
  - If anyone can be anywhere (only telling routers which group they are sending to or joining) how can routers learn from where and to where they should forward the data?

IPv4 Multicast Addressing

- IPv4 multicast addresses: 224.0.0.0 – 239.255.255.255 (224/4, class D)
- These are subdivided in rather complicated ways,
  - see http://www.iana.org/assignments/multicast-addresses/ for details
- Examples:
  - 224.0.0.0 – 224.0.0.255 (224.0.0/24) – Local network control block, never forwarded
    - 224.0.0.1 - All local hosts
    - 224.0.0.2 - All local routers
    - 224.0.0.5 - OSPF
    - 224.0.0.13 - PIM
    - 224.0.0.22 - IGMP
  - 224.0.1.0 – 224.0.1.255 (224.0.1/24) – Internetwork control block, forwarded
  - 224.2/16 – SAP
  - 232/8 – SSM (only to be used for Source Specific Multicast)
  - 233/8 – GLOP
  - 234.0.0.0 – 238.255.255.255 – Reserved
  - 239/8 – Administrative scoping

Address Assignment, SAP and GLOP

- Knowing what addresses to use when creating a session seems rather complicated
- SAP (Session Announcement Protocol, RFC 2974)
  - Announces a session
  - SAP applications also help you pick what addresses to use
  - Uses dynamic groups in range 224.2.128.0 – 224.2.255.255 for global sessions
  - Global announcements sent to 224.2.127.254
  - sdr is the most common SAP application, but not used so much these days
- GLOP (not an acronym)
  - Assignment based on AS numbers, RFC 3180
  - 233.x.x/24 where x.x is an officially assigned AS number
  - For private AS space there is EGLOP (RFC 3138)
    managed by registries, e.g. RIPE (still 233.x.x/24, but with private AS numbers)

Administrative Scoping – 239/8

- Addresses in the range 239/8 are used for administrative scoping
  - Private address space, not to be used globally
  - Different networks can use the same addresses
- 239.255/16 is the smallest administrative scope
  - Sometimes used for site-local
- 239.192/14 is organization-local scope
  - These addresses should work throughout JANET
  - All but 239.194/16 are restricted to JANET
    - 239.192/16, 239.193/16 and 239.195/16 used for sessions visible throughout JANET, but not outside
  - 239.194/16 is used for GÉANT
    - i.e. sessions using these groups are available throughout GÉANT (European academic networks), but not outside
- Multicast distribution can be restricted by specifying a small TTL value for packets
  - Limited use. With routing protocols like PIM-SM and MSDP, packets may travel very far even if TTL is small

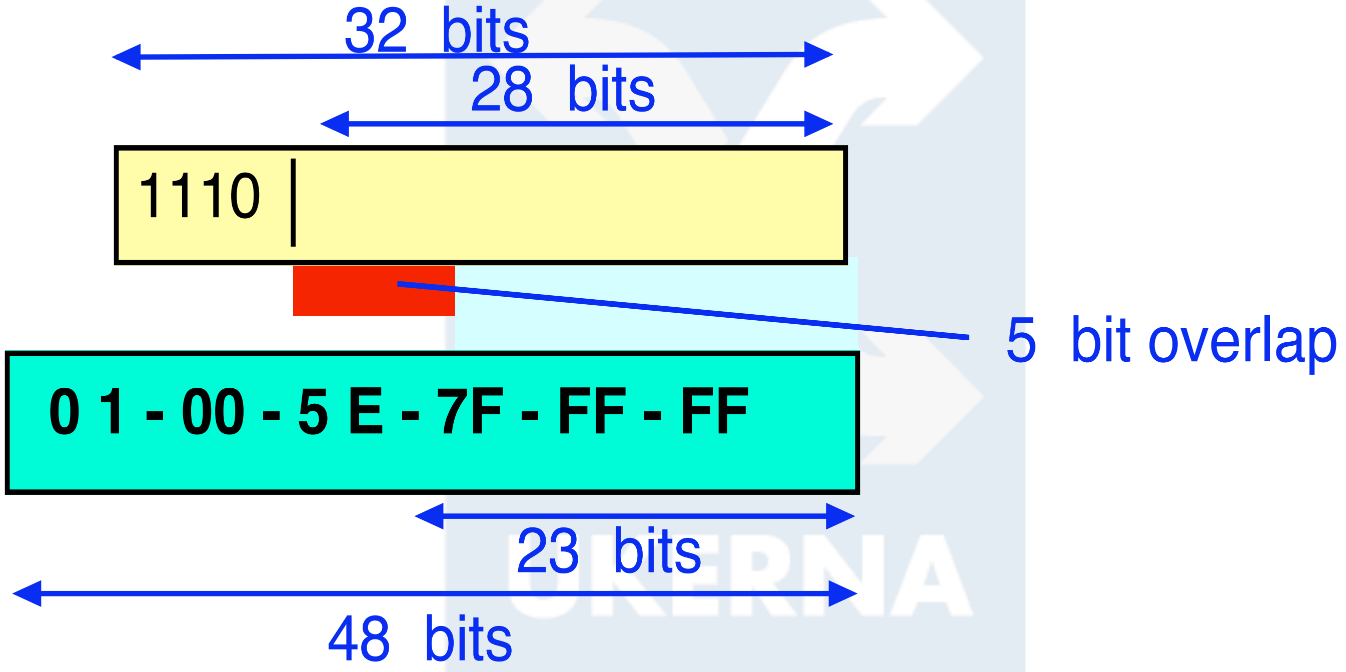
Multicast on the LAN

- Multicast is defined for Ethernet
  - Ethernet multicast is exactly like traditional IP multicast model
  - IP multicast service is based on the Ethernet service model extended from working on a LAN to the Internet
- Originally Ethernet multicast was very simple
  - Any host can send
  - All packets go everywhere (coax cables or hubs)
  - Any host on the LAN can choose to listen, only need to tell NIC what packets to pick up
  - But then came bridges and switches…

Mapping to Ethernet MAC

Class D IPv4 destination address

224.0.0.0-239.255.255.255



MAC hardware destination address

One L2 (MAC) address
may carry multiple L3 (IPv4) addresses

Internet Group Management Protocol – IGMP

- IGMP is a protocol used on the LAN for hosts to tell routers which groups they are interested in
  - May not be necessary to receive from a source on the same LAN
  - However, IGMP is often used by switches to restrict multicast flow on the LAN

- Three versions of IGMP
- IGMPv1 (RFC 1112)
  - Hardly no one uses this anymore, used by e.g. Windows 95
- IGMPv2 (RFC 2236)
  - Maybe the most commonly used version today
- IGMPv3 (RFC 3376)
  - This is the recommended version, backwards compatible with IGMPv2
  - This is needed for SSM (to specify sources to join)
  - Supported by Windows XP, recent Linux and some UNIX systems

IGMP Overview

- Multicast router with the lowest address is elected as *IGMP querier*
- The querier sends periodic queries (default 125s intervals)

- Hosts respond with which groups they want to receive
- Hosts also immediately send a report when host initially joins
  - Do not have to wait for the periodic query
- Hosts immediately send a message when they leave
  - This was not the case for IGMPv1

- State times out if there are no responses to the queries
  - This is important if a host crashes and never says stop

Configuring IGMP on Cisco IOS

For IOS to do IGMP we need to enable multicast routing and enable PIM on the interfaces

```
ip multicast-routing
!
interface …
 ip pim sparse-mode
```

IOS uses IGMPv2 by default, to use IGMPv3:

```
interface …
 ip igmp version 3
```

Checking IGMP state on Cisco IOS 1/2

To see group memberships we can do:

```
cisco> show ip igmp groups
IGMP Connected Group Membership
Group Address    Interface                Uptime    Expires   Last Reporter
239.255.255.255  Vlan26                   7w0d      00:02:32  158.38.63.1
239.255.255.255  Vlan2                    7w0d      00:02:29  128.39.47.90
239.255.255.253  Vlan20                   4d10h     00:02:33  158.38.60.95
239.255.255.253  Vlan10                   1w4d      00:02:33  158.38.60.11
239.255.255.253  Vlan95                   7w0d      00:02:12  158.38.152.196
239.255.255.250  Vlan80                   1d00h     00:02:56  158.38.61.149
239.255.255.250  Vlan10                   5d12h     00:02:26  158.38.60.44
239.255.255.250  Vlan20                   7w0d      00:02:35  158.38.60.95
224.2.127.254    Vlan26                   7w0d      00:02:35  158.38.63.1
224.2.127.254    Vlan2                    7w0d      00:02:32  128.39.47.90
239.255.67.250   Vlan20                   4d10h     00:02:38  158.38.62.97
232.26.17.81     Vlan26                   4d07h     stopped   158.38.63.22
```

Checking IGMP state on Cisco IOS 2/2

To show sources and not just groups for IGMPv3 we can do

```
cisco> show ip igmp groups detail
Flags: L - Local, U - User, SG - Static Group, VG - Virtual Group,
       SS - Static Source, VS - Virtual Source

Interface:      Vlan26
Group:          239.255.255.255
Flags:          L U
Uptime:         7w0d
Group mode:     EXCLUDE (Expires: 00:02:24)
Last reporter:  158.38.63.1
Source list is empty

Interface:      Vlan26
Group:          232.26.17.81
Flags:          SSM
Uptime:         4d07h
Group mode:     INCLUDE
Last reporter:  158.38.63.22
Group source list: (C - Cisco Src Report, U - URD, R - Remote, S - Static,
                    V - Virtual, Ac - Accounted towards access control limit,
                    M - SSM Mapping)
  Source Address   Uptime    v3 Exp   CSR Exp   Fwd  Flags
  129.177.30.248   4d07h     00:02:32  stopped   Yes  R
  129.242.2.140    4d07h     00:02:32  stopped   Yes  R
  152.94.26.6      4d07h     00:02:32  stopped   Yes  R
  158.36.22.14     2d07h     00:02:32  stopped   Yes  R
```

Multicast and Ethernet Switches

- Many switches have features that can restrict multicast flow to only ports where there are members
  - This is usually good, but sometimes switches misbehave

- There are generally three possible methods
  - GARP/GMRP – hosts use l2 protocol to tell switches
    - Supported by very few systems
  - CGMP – routers tell switches what to do
    - A Cisco proprietary protocol and Cisco are dropping support
    - IGMPv3 “leaves” not supported
  - IGMP snooping/proxy
    - The switches snoop the IGMP messages going between hosts and routers

Intradomain Multicast Routing

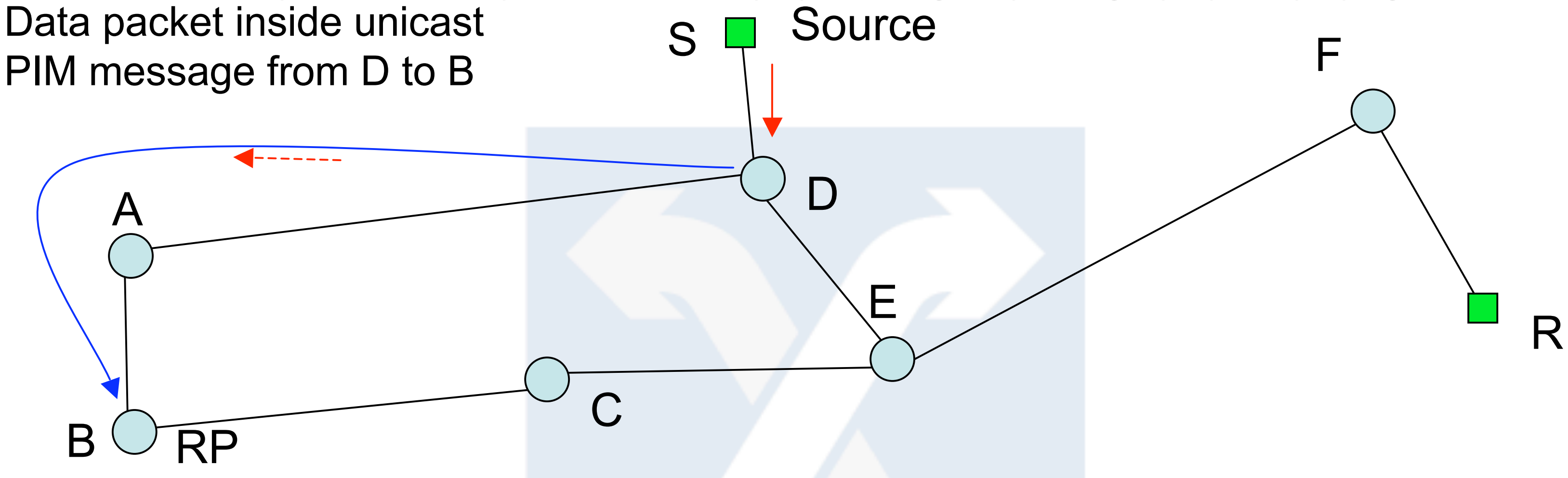
- We will look at how multicast can be deployed in a site network
  - Or any other reasonably small network within one management domain
- We want the traditional service model where anyone in the network can join without knowing where sources are, and sources can send without knowing the receivers
- Multicast routing is all about efficiently creating multicast distribution trees, where each tree is rooted at the source and spreads out to the receivers
- May be difficult since neither sources nor receivers know where everyone else is

- Some multicast routing protocols are based on “flood and prune”
  - Data is initially flooded everywhere
  - Trees are then pruned to be only where needed
  - However does not scale well, unless receivers are almost everywhere
- The most common routing protocols today are PIM-DM and PIM-SM
  - PIM is Protocol Independent Multicast in that it can make use of the unicast routing protocol, and is independent of which Internet Protocol is used

PIM-SM/PIM-DM

- PIM-DM (PIM Dense Mode, RFC 3973)
  - A flood and prune protocol
  - May be okay for a dense population of receivers
- PIM-SM (PIM Sparse Mode, RFC 2362)
  - Does not flood
  - Works with a sparse population of receivers, scales much better
- PIM-SM is by far the most commonly used protocol today
- PIM-SM makes use of a so-called Rendezvous Point where sources and receivers meet
- All routers in the network agree where the RP is for a group
  - Hosts and receivers do not need to know where the others are.
  - Trees, at least initially, pass through the RP

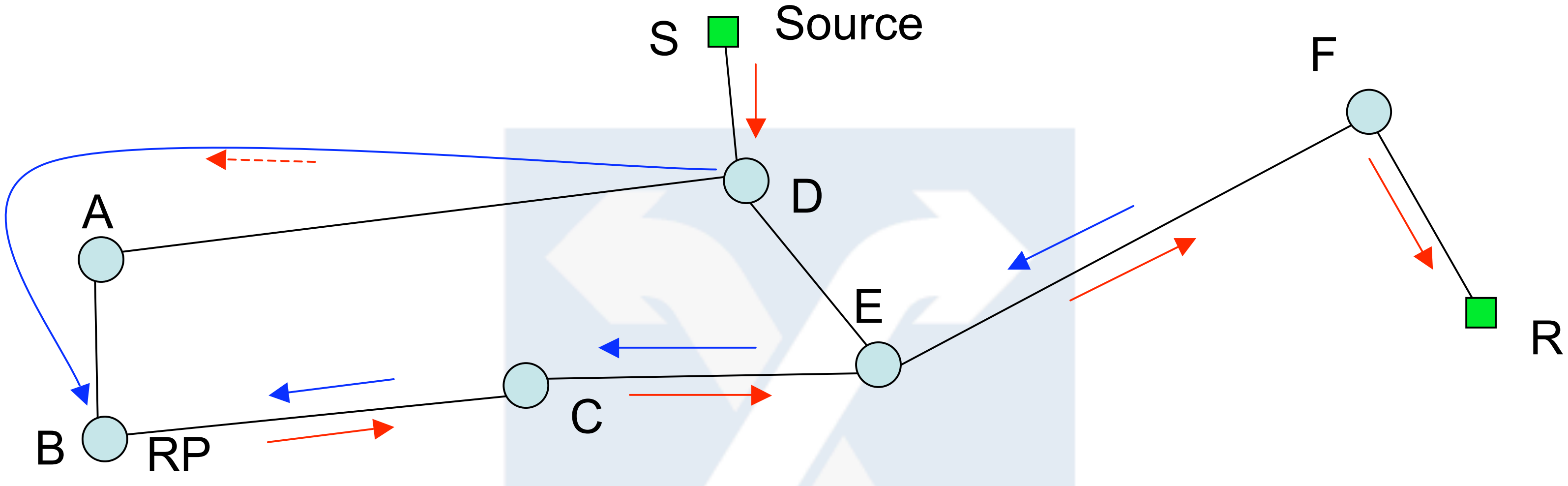
How the RP learns of sources



The blue arrows are PIM messages and the red are data packets

- All the routers are preconfigured with the same RP address B
- We're looking at what happens initially when a source starts sending
- On the link where the source is, one router is elected as Designated Router. It will encapsulate multicast packets into unicast PIM register messages, addressed to the RP. Here router D is the DR
- So in this way, the RP B will learn about the source, as well as receiving the actual multicast data

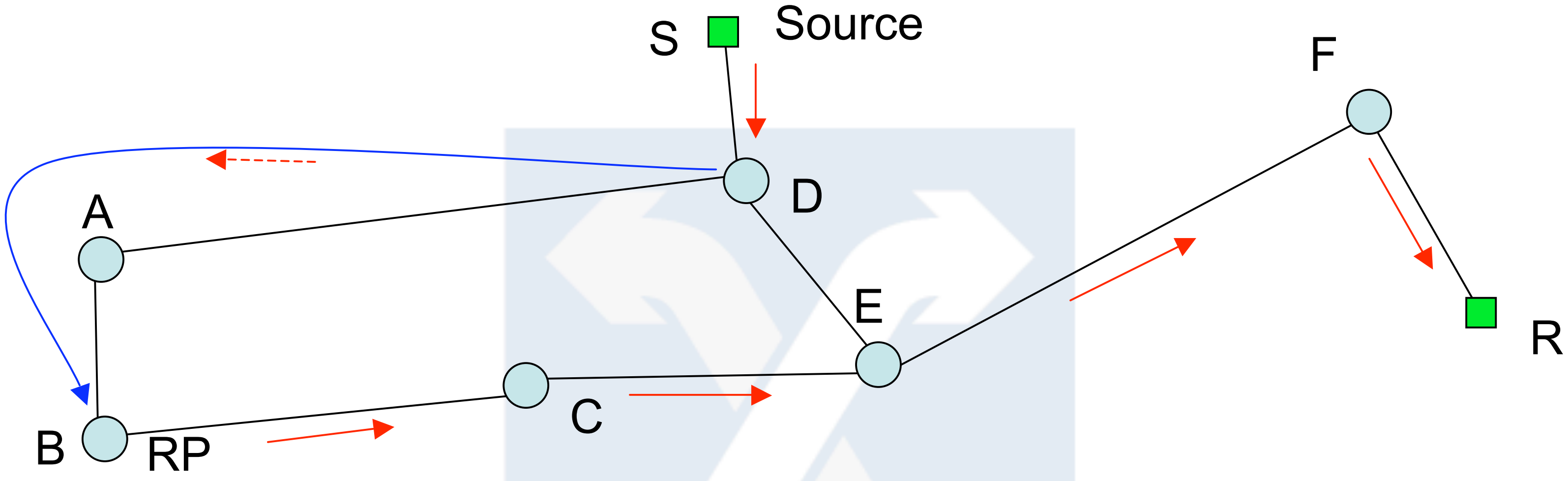
How the RP learns of receivers



The blue arrows are PIM messages and the red are data packets

- All the routers are preconfigured with the same RP address B
- We're looking at what happens when a host R behind router F joins the group
- F should send a PIM join message towards the RP. It checks the routing table, and finds that E is the next hop for reaching B, so it sends the join to E
    - E is said to be F's RPF neighbour for B
- When E receives the join from F, it will find that C is next hop towards the RP, and sends join to C.
- Finally, C will send join to B which is the RP
- If RP is receiving any data (in this case it is), it will as soon as it receives the join (from C here) forward data. Data goes back to receiver following the joins backwards

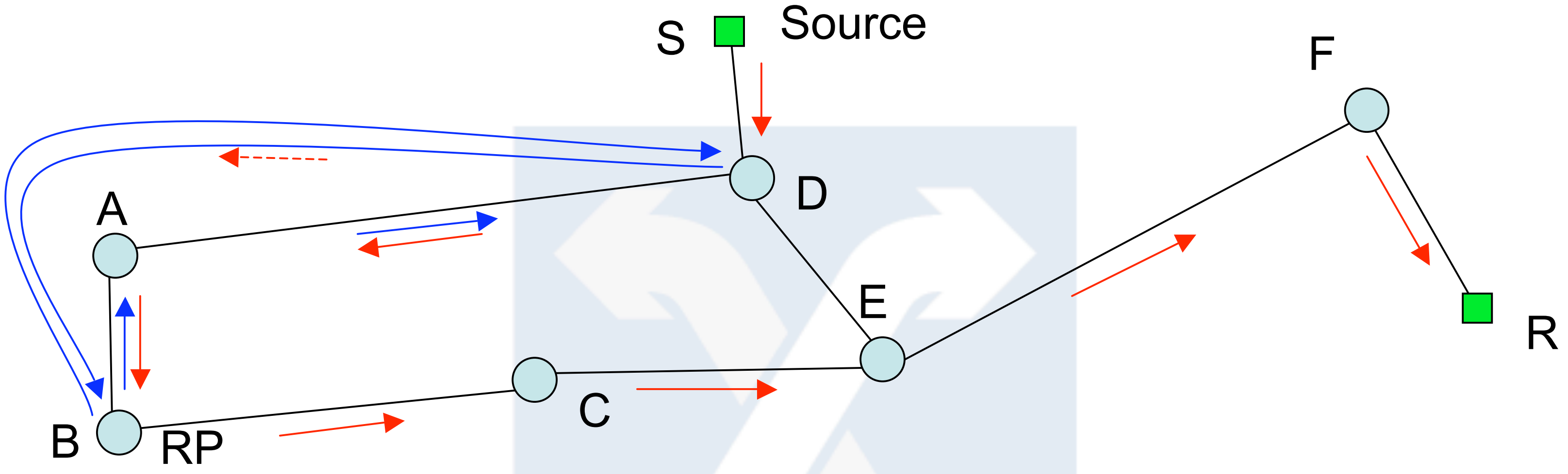
Avoiding unicast encapsulation 1/2



The blue arrows are PIM messages and the red are data packets

- Packets are now flowing from source to interested receivers, but typically wants to avoid encapsulating all multicast into unicast. Requires lots of resources to encapsulate packets at the DR (D) and then decapsulate them at the RP (B)
- PIM-SM allows joining towards a specific source, and this is what the RP will do when it receives encapsulated packets and there is someone wanting to receive them
  - If no one want to receive, there is also a way for the RP to tell the DR to not send anything for a while

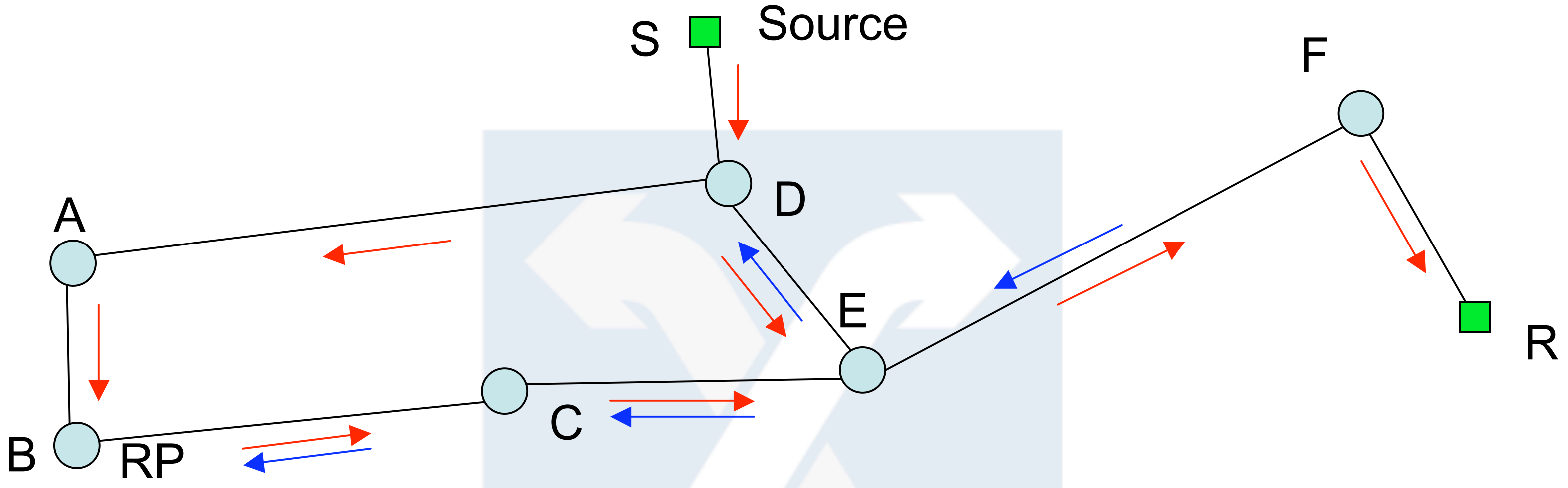
Avoiding unicast encapsulation 2/2



The blue arrows are PIM messages and the red are data packets

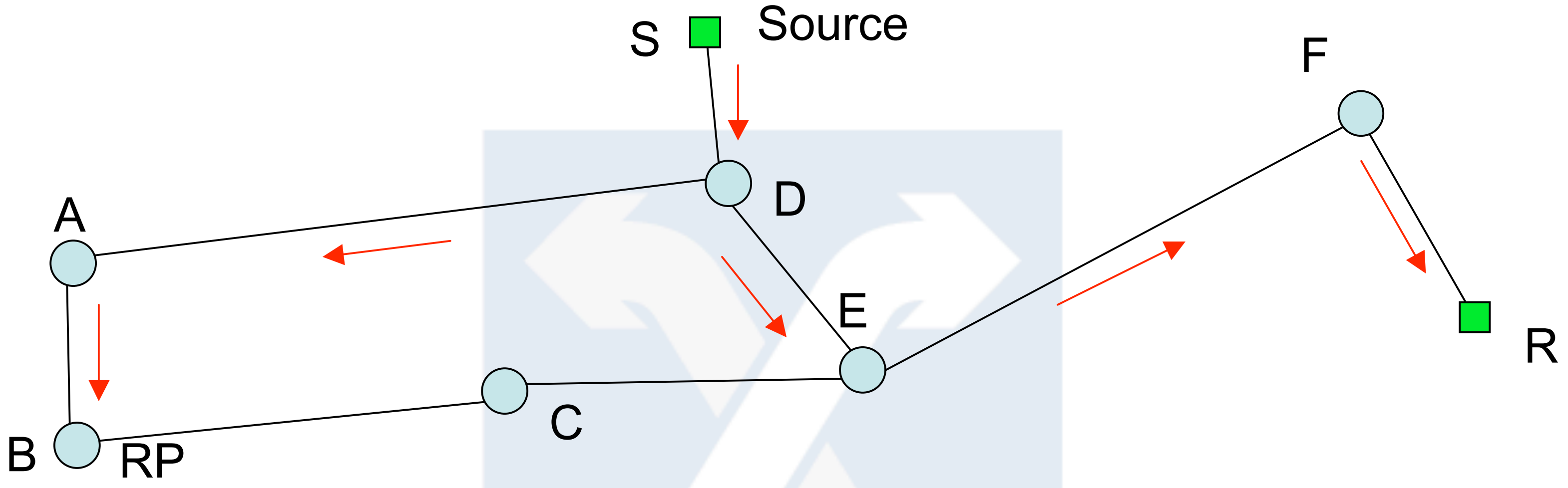
- First B sends source-specific joins towards source S
- B will after joins reach D, start receiving packets natively (not encapsulated)
- B will then send a unicast message (PIM register stop) to D, asking it to stop sending encapsulated packets from S

Optimising forwarding path 1/2



- Now packets are flowing natively, but why send through RP when could go the shorter path from D to E?
- F is often configured to immediately join towards sources
  - It can do this after first packet from S. It sees S's address in the IP header
- So F sends source-specific join towards S
- When these joins reach D, D will also forward multicast packets to E
- When E receives from D, it will send prune messages towards RP saying, don't send me packets from S
  - It might still receive from other sources via the RP

Optimising forwarding path 2/2



- Now almost everything is perfect
- But why should RP receive data if the receivers get it directly?
- So, if no one wants to receive S from RP, it will stop joining S
- Joins are in general sent periodically. If no joins are sent, state will expire and no more multicast will be sent
- Finally, everything is optimal

PIM-SM Summary

- PIM-SM requires an RP for source discovery
- All routers must use the same RP and somehow know the address for it
- Initially packets from a source will be sent to RP
  - Even if no one wants to receive
- Except for this, packets are only sent out on an interface if a join has been received on it
- Initially packets flow from source to receivers via the RP
- Optimal path (not via RP) usually established quickly

Configuring PIM-SM 1/2

- Routers must have:
  - Multicast routing enabled
  - PIM on interfaces where they face one another
  - IGMP on host interfaces
  - Note that on some routers, incl IOS, you need to enable PIM on host interfaces to get IGMP
- Also essential to configure RPs
  - All routers in the domain must agree which RP to use for a group. May have just one RP for all, or different RPs for different ranges
  - RP routers must be configured to be RPs
  - Other routers must know the addresses of the RPs for the different group ranges

Configuring PIM-SM 2/2

- We recommend to statically configure the RP address(es) on each router
  - RP addresses can be configured as additional loopback interfaces on routers and announced as host routes into the routing table
  - You can then move the RP without configuring all the routers
  - Anycast-RP (RFC 3446) allows failover between multiple RPs
    - Today done with MSDP (out of scope of this workshop)
  - Another option is BSR (bootstrap router protocol)
  - Allows routers to dynamically learn which RPs to use
  - A Cisco specific protocol similar to BSR is Auto-RP

Configuring RPs on Cisco IOS

Recommend static RP config, on RPs and other routers you may simply do e.g.

```
ip pim rp-address 192.0.2.1
```

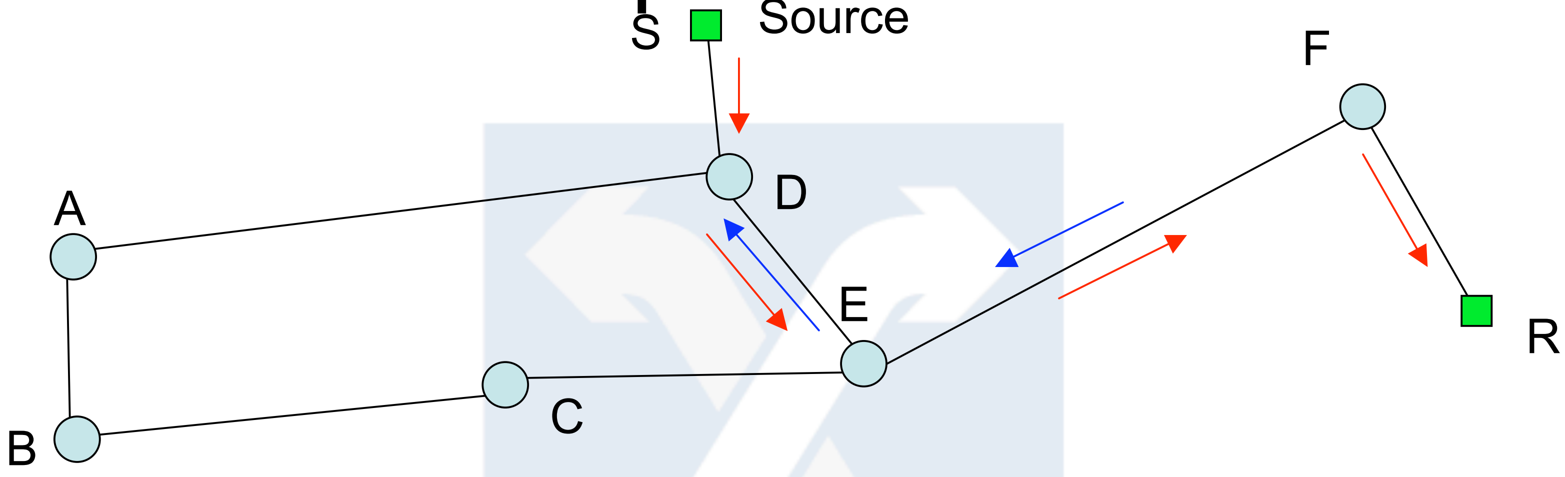
This would specify 192.0.2.1 to be the RP for all multicast groups. In some cases you may prefer to have your own RP for groups used internally, while using your provider’s RP for global groups. Below is an example of that

```
ip pim rp-address 192.0.2.1 21
ip pim rp-address 192.0.2.129 20
!
access-list 20 permit 239.255.0.0 0.0.255.255
access-list 20 permit 229.55.150.208
access-list 20 permit 224.0.1.0 0.0.0.255
access-list 20 deny any
!
access-list 21 deny 239.255.0.0 0.0.255.255
access-list 21 deny 229.55.150.208
access-list 21 deny 224.0.1.0 0.0.0.255
access-list 21 permit any
```

SSM – Source Specific Multicast

- SSM is a new multicast service model
  - Receivers specify the source address(es) in addition to the group
- This avoids rogue sources sending to the group
  - Imagine watching one video stream and someone else also sends their video, or just random data at high rates
- The main benefit is hugely simplified routing
- PIM-SM works very well with SSM
  - Last-hop routers know the sources, can immediately join Shortest-Path Trees.
  - No need for an RP and a shared tree to do source discovery
  - Multicast much easier to deploy and manage
- SSM requires source discovery at the application layer
  - Very easy for streaming with one fixed source (maybe most important use of multicast)
  - May be difficult for some multi-party or discovery applications

Source-Specific Multicast



- With SSM, the receiver somehow knows the source address S
- So host R tells F it wishes to receive a particular group from S
- F then sends source-specific join towards S
- So we immediately get the optimal path and no RP is needed
- There is a problem though. How can the receiver R, or the application at the receiver, learn the source addresses?
  - Source discovery must somehow be done at the application layer

Configuring SSM on Cisco IOS

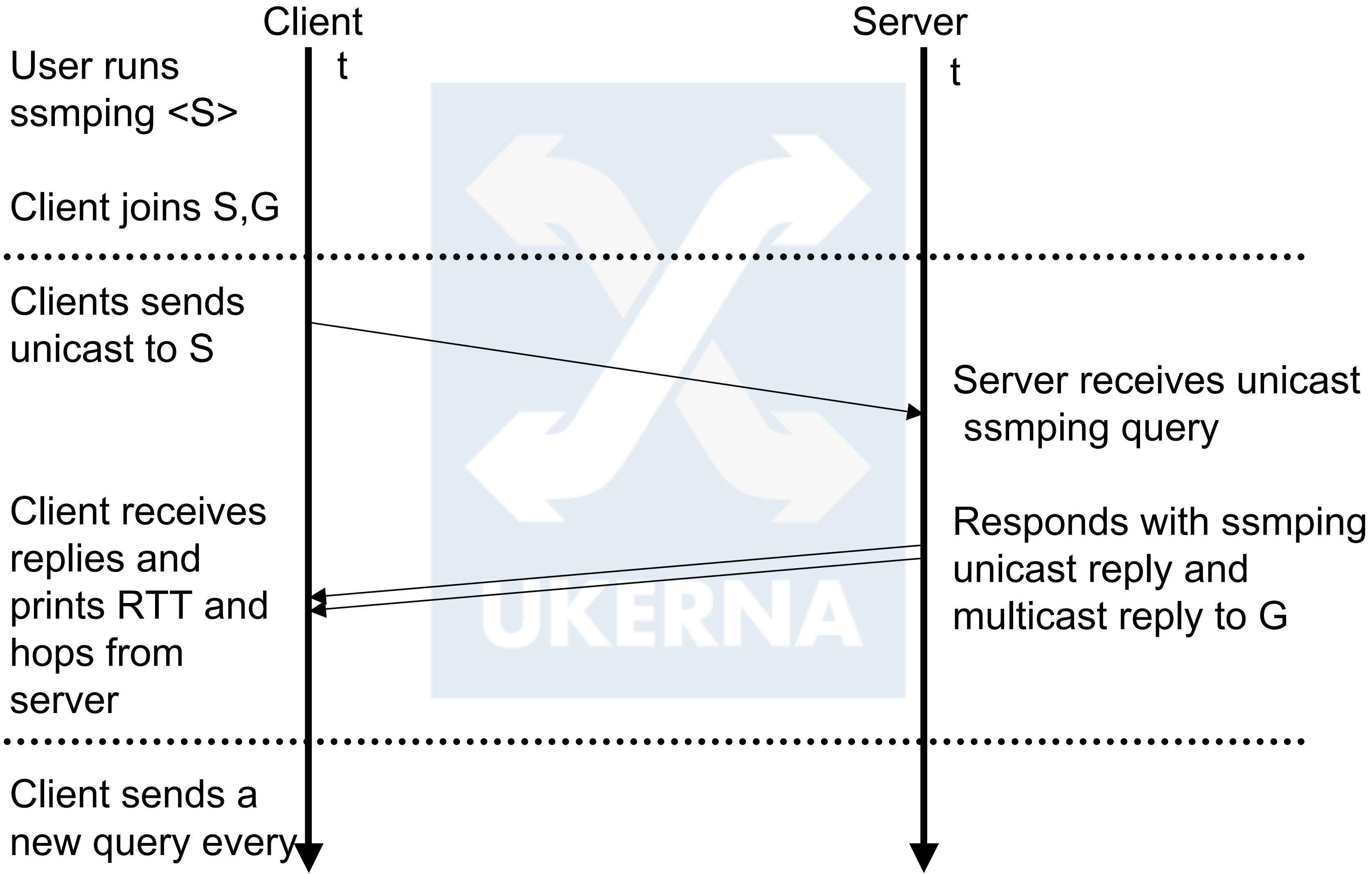
- For SSM to work you need to enable IGMPv3 at the edges (see earlier slide)
- Must say the standard (default) SSM range 232/8 to be treated as SSM groups
  - Do not allow (*,G)-join for SSM groups
  - Never send PIM registers for 232/8
  - RPs ignore (*,G)-joins and registers for such groups

```
ip pim ssm default
```

ssmping

- A tool for testing multicast connectivity
- Behavior is a bit like normal ping
- A server must run ssmpingd
- A client can ping a server by sending unicast ssmping query
- Server replies with both unicast and multicast ssmping replies
- In this way a client can check that it receives SSM from the server
  - also parameters like delay, number of router hops etc.
- JANET is running a server at ssmping.beacon.ja.net
- There is a similar tool called asmping for checking ASM connectivity
- See http://www.venaas.no/multicast/ssmping/

How ssmping works



User runs ssmping <S>

Client joins S,G

Clients sends unicast to S

Server receives unicast ssmping query

Responds with ssmping unicast reply and multicast reply to G

Client receives replies and prints RTT and hops from server

Client sends a new query every second

Example IPv4 ssmping output (v6 supported)

```
$ ssmping -4 -c 5 ssmping.beacon.ja.net
ssmping joined (S,G) = (193.60.199.162,232.43.211.234)
pinging S from 158.38.63.22
  unicast from 193.60.199.162, seq=1 dist=16 time=39.331 ms
  unicast from 193.60.199.162, seq=2 dist=16 time=39.394 ms
multicast from 193.60.199.162, seq=2 dist=16 time=43.905 ms
  unicast from 193.60.199.162, seq=3 dist=16 time=39.542 ms
multicast from 193.60.199.162, seq=3 dist=16 time=39.547 ms
  unicast from 193.60.199.162, seq=4 dist=16 time=39.137 ms
multicast from 193.60.199.162, seq=4 dist=16 time=39.142 ms
  unicast from 193.60.199.162, seq=5 dist=16 time=39.535 ms
multicast from 193.60.199.162, seq=5 dist=16 time=39.539 ms

--- 193.60.199.162 ssmping statistics ---
5 packets transmitted, time 5000 ms
unicast:
   5 packets received, 0% packet loss
   rtt min/avg/max/std-dev = 39.137/39.387/39.542/0.292 ms
multicast:
   4 packets received, 0% packet loss since first mc packet (seq 2) recvd
   rtt min/avg/max/std-dev = 39.142/40.533/43.905/1.958 ms
$
```

What does ssmping output tell us?

- 16 unicast hops from source, also 16 for multicast, might indicate that unicast and multicast follow the same path
- Multicast RTTs are about same for unicast and multicast
  - However, the delay for the first multicast packet is large, would need to send more queries for a proper test
  - Note the difference in unicast and multicast RTT shows one way difference for unicast and multicast replies, since they are replies to the same request packet
- Multicast tree not ready for first multicast reply, ok for 2nd so the tree was in place after about one second when the second packet was sent
- No unicast loss, no multicast loss after tree established

IPv6 multicast

- There is also multicast for IPv6
- Main principles are the same
- Much larger address space
  - Allows many interesting possibilities
  - New ways of allocating addresses, clashes are unlikely
  - e.g. unicast prefix based addresses
  - Also embedded-RP, where group address has the RP address encoded in it so that routers can immediately with no prior configuration know where the RP is
- Well defined scoping
  - Scoping is much easier to understand, 15 scopes
- MLDv1 replaces IGMPv2, MLDv2 replaces IGMPv3
- IPv6 multicast is now supported by a wide range of routers and host operating systems