Multicast Overview

Wyatt Sun

wyatt@kiwi.m8.ntu.edu.tw

NTU EE, Internet Lab.

Multicast Basic Concepts

- One-to-many or many-to-many communication, a packet is sent once by the source, and reaches every destination without unnecessary duplication in the network.
- The only Internet transport protocol that support multicasting is UDP.

Multicast Addressing

- Background: Classful Addressing:
  - Class A: `0.......  ........  ........  ........`
  - Class B: `10......  ........  ........  ........`
  - Class C: `110.....  ........  ........  ........`
  - Class D: `1110....  ........  ........  ........`
  - Class E: `11110...  ........  ........  ........`
- Multicast Addresses: Class D addresses (224.0.0.0 to 239.255.255.255) are called *multicast groups*.

Class D is declared but unused in RFC 820, Jan. 1983.
Class D is allocated for multicast in RFC 966, Dec. 1985.
Class E is declared and reserved in RFC 988, Jul. 1986.

The assignments of Class D addresses are described in [RFC 3171].

Mapping between IP and Ethernet Addr.

Protocol	Address	
Ethernet (48-bit)	00000001 00000000 01011110	0....... ....
IP (32-bit)	1110????	?....... ....

- A block of Etnernet multicast addresses prefixed `01:00:5e` is assigned to the IANA by the IEEE. The IANA allocates half of the block, with the high-ordered bit be set to `0` for IP multicast addresses.
    - 32 IP multicast addresses map to 1 Ethernet multicast address.

IGMP

- *Internet Group Management Protocol*
- IGMP is used to comunicate group membership information on a local network between edge-routers and leaf nodes, multicast routers use it to keep track of group membership on its physically attached networks.
- IGMP is a transport protocol above IP with the protocol number of 2, and its messages are carried in IP datagrams (as with ICMP).
  - IGMP v1: RFC 1112 (Aug. 1989)
  - IGMP v2: RFC 2236 (Aug. 1997)
  - IGMP v3: RFC 3376 (Aug. 2002)

IGMP v1

- Proposed in conjunction with DVMRP in RFC 1112.
  - The router sends an IGMP query periodically.
  - The host replies with an IGMP report when the first process joins a group.
  - The host does nothing when the last process leaves the group.
- If a host is scheduled to send a report, but receives a copy of the same report from another host, the response can be canceled.

Message Format of an IGMPv1 Message:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version| Type  |    Unused     |           Checksum             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Group Address                          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- IGMP Query: Type=1, Group Address=224.0.0.1(all-hosts) or all-zero, TTL=1
- IGMP Report: Type=2, Group Address=interested_group_address, TTL=1

IGMP v2

- Adds fast termination of group subscriptions. When a host leaves a multicast group, if it was the last host to reply to a Query with a Membership Report for the group, it SHOULD send a *Leave Group message* to the all-routers multicast group (224.0.0.2).

Message Format of an IGMPv2 Message:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     | Max Resp Time |           Checksum               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Group Address                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Max Response Time: used only in Membership Query message. It specifies the maximum allowed time before sending a responding report in units of `1/10` second.
- Leave Group: Type=0x17.
- Multicast Router Termination: Type=0x26

IGMP v3

- Adds support for source filtering, that is, the ability for a system to report interest in receiving packets only from specific source address, or from *all but* specific source addresses.
- use Current-State Record, Filter-Mode-Change Record and Source-List-Change Record to include or exclude specified multicast addresses.
- The suppression of membership reports in IGMP v1/v2 is removed, this permits explicit membership tracking.

Message Format of an IGMPv3 Message:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type=0x22   | Max Resp Code |           Checksum              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Group Address                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Resv |S| QRV |      QQIC     |     Number of Sources (N)       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       Source Address [1]                        |
+-                                                               -+
|                       Source Address [2]                        |
+-                              .                                -+
.                               .                                 .
+-                              .                                -+
|                       Source Address [N]                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Multicast Routing Protocols

	Source-Rooted Tree	Shared Tree
intra-domain	DVMRP, MOSPF, PIM-DM	PIM-SM, CBT, OCBT
inter-domain		HIP, KHIP, BGMP

- Intra-domain routing protocols - Source-Rooted Tree
  - DVMRP (RFC 1075, Nov. 1988)
  - MOSPF (RFC 1585, Nov. 1994)
  - PIM-DM (RFC draft-pim-dm-new-v2-04.txt, Nov. 2001 - Sep. 20023)
- Inter-domain routing protocols - Shared Tree
  - HIP (Proc. of PODC, 1998)
  - KHIP (Keyed HIP, SIGCOMM 1999)
  - BGMP (draft-ietf-bgmp-spec-05.txt, Jun. 2003)
- Inter-domain routing protocols - Shared Tree
  - PIM-SM (RFC 1075, Nov. 1988)
  - CBT (RFC 2201, Sep. 1997)
  - OCBT (INFOCOM, 1997)

Source-Rooted Tree

- DVMRP
  - Flood-and-Prune
  - use RPF (Reverse Path Forwarding) for flooding.
  - Not suitable for large scale due to flooding.
- PIM-DM
  - very similar to DVMRP but independent of the underlaying unicast routing protocol.
  - Does not support tunnels.

- MOSPF
  - Base on OSPF, a link-state routing protocol.
  - The per-source tree are computed by each router by routing information from OSPF.
  - forwarding base on (src_addr, dest_addr)

Shared Tree

- CBT
  - Use a bi-directional shared tree for a group.
  - Reduce the number of state information in each router, (S,G) vs. (*,G).
  - Explicit Join and Leave message is needed. (IGMP v3).
- OCBT
  - Makes CBT loop-free and robust.
  - Assign the cores and logical level, indicating the position in the hierarchy.
- BGMP (Border Gateway Multicast Protocol)
  - bi-directional shared trees

- HIP
  - A hierarchical multicast routing scheme that uses OCBT to route between heterogeneous multicast routing domains.
  - Assign the cores and logical level, indicating the position in the hierarchy.
- KHIP (Keyed HIP)
  - Provide efficient mechanism for distributing multicast keys and changing them as necessary.
  - Limit access to the multicast routing tree to authorized members, preventing unauthorized members from sending to or receiving from the multicast session. position in the hierarchy.

- PIM-SM
  - Use a variant of center-base tree algorithm.
  - Use a uni-directional shared tree call RP tree for a group.
  - A set of special routers called Rendezvous Points.
  - RPs play the role like a proxy for both receivers and senders.
  - If the sender’s traffic increase beyond some threshold, the shortest path router is set up between sender and the corresponding RP. In addition, the routes between RP and the receivers switch from the RP tree to a source-based shortest path tree.

MAAA

- Multicast Address Allocation Architecture
- An given multicast address is limited in two dimensions: Lifetime and Scope.
- Considerations: robustness/availablity, timelineness, low probability of clashes, and good address space utilization.
- Good address space packing and constant availability are more important than guaranteeing that address clashes never occur.
- 3 catagories of multicast address allocation:
  1. Static
  2. Scope-relative
  3. Dynamic

- Static addresses:
  1. permanent lifetime.
  2. scope defined by the scope range in which they reside.
  3. for specific protocols that require well-known address to work.

- Scope-relative addresses:
  - permanent lifetime
  - valid in every scope and location
  - RFC 2365 reserves the highest 256 addresses in every administrative scope range for relative assignments.
  - reserved for infrastructure protocols which require an address in every scope.
  - clients can use MZAP or MADCAP to obtain the list of scope in which they reside.
- Dynamic addresses:
  - provided on demand with limited lifetime.
  - for most condition.
  - make address aggregation for routing.

IP Multicast

- Applications:
  - Audio/Video distribution
  - Audio/Video conference
  - File transfer
- Conventional IP Multicast:
  - a set of hosts can be aggregated into a *group* with a single address.
  - lakes basis for *charging*, *access control*, and *scalability*.

Limited Deployment of IP-Multicast

The success of today's Internet results from the unicast-based email and web applications.

- Chick-and-egg problem
- Billing (for ISPs and content providers)
- For ISP, it upsets the router migration model that ISPs follows.
- Group management (for sender and receiver)

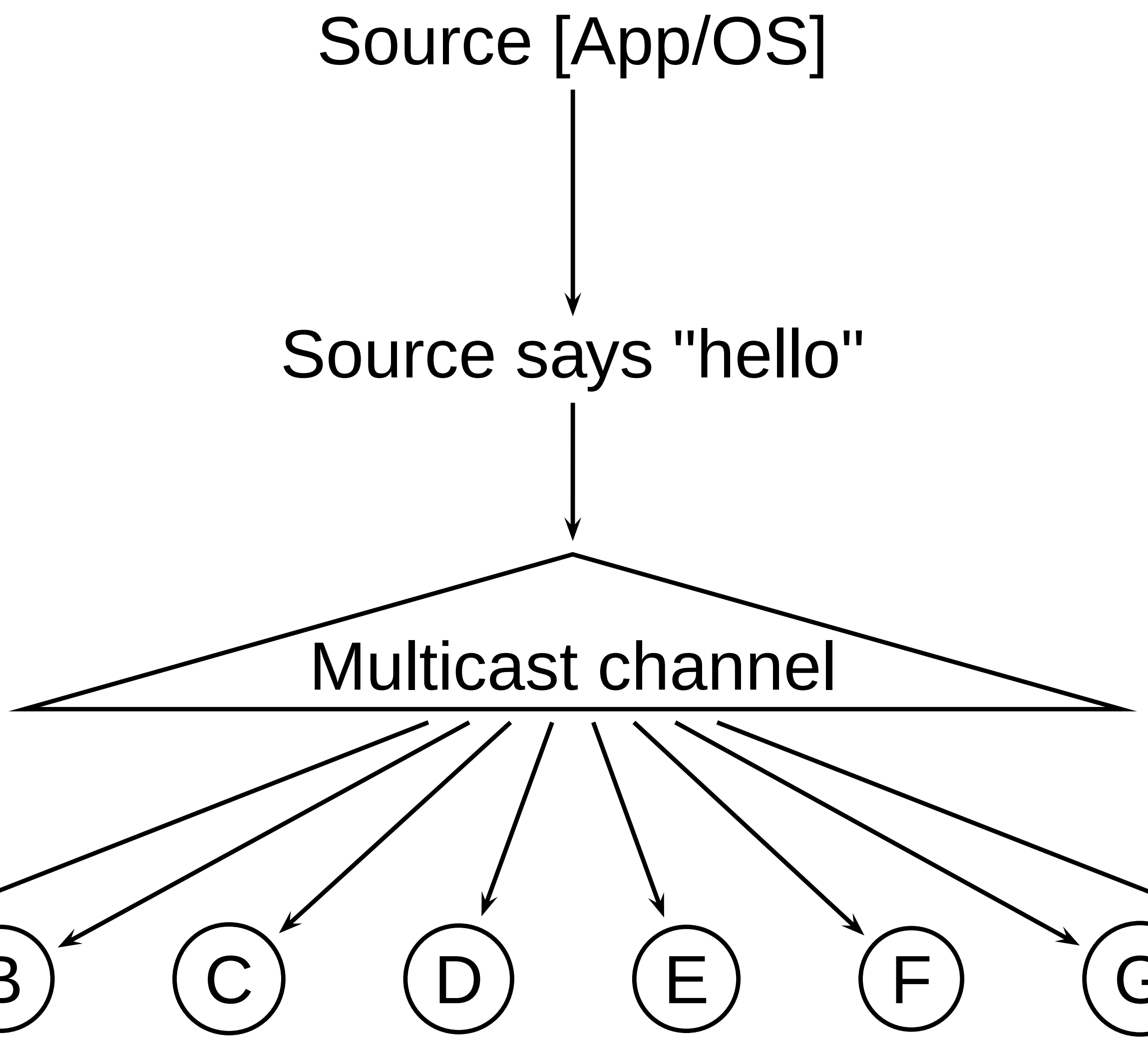
Problems of conventional IP multicast

Violates common ISP billing models, e.g;, 1 million subscribers in Super Bowl.

1. Input data rate is the basis for ISP charging
2. No indication of the group size
3. Can’t restrict sender
4. World-wide unique address is needed for each application, and there are only 256M multicast addresses totally.
5. Routing scalability

- EXPRESS: EXPlicity REquested Single Source
- A multicast *channel* is a datagram service identified by a tuple (S, E).
- In the network: $(S, E) \neq (S', E)$
  On the sender: $(S, E) \neq (S, E')$
- Only the source host S may send to (S,E).
- For each (S,E), it has only one explicitly designated source and 0+ channel *subscribers*.
- Routers identify a datagram by its destination address.

Single Source Multicast



Advantages of EXPRESS

- For source:
  - EXPRESS provides 2^{24} channels per source.
  - Source can control other hosts' ability to subscribe to the channel.
  - The source can use the `CountQuery` to determine the number of subscribers or to take a subscriber vote.
- For subscribers:
  - Only receive traffic from the source, need not explicitly exluce other sources, as with IGMPv3. This reduces the traffic on the receiver's last-hot link.
  - Feedback with the count mechanism may enable new appplication-level features.

- For ISP:
  - Take *channels* as a valuable service.
  - The source is the owner of the channel.
  - The counting machanism can determine the number of subscribers.

All these advantages come from restricting the multicast channel to having a single source.

EXPRESS service interface extensions

- Source host's extended service interface:
`count = CountQuery(channel, countId, timeout)`
- Source informs the network that channel is authenticated by
`channelKey(channel,` $K_{(S,E)}$`)`
The netwrok layer ensures only hosts presenting $K_{(S,E)}$ can subscribe.
- To *subcast* a packet to a subnet, the source can unicast a encapsulated packet to the "on-channel" router.

- Subscriber’s extended service interface:
  `result = newSubscription(channel [,` $K_{(S,E)}$`])`
  `result = deleteSubscription(channel [,` $K_{(S,E)}$`])`
- A subscriber reply to `CountQuery` request with `count(channel, countId, count)`

ECMP (EXPRESS Count Management Protocol)

- Responsibility:
  - Maintains the distribution tree:
    - Subscription
    - Channel maintenance
  - Counting:
    - Supports source-directed counting and voting.
- Function from ECMP:
  - Access control
  - Accounting

- ECMP messages:
  - `CountQuery(channel, countId, timeout)` Originated from source, and forwarded to all the receivers.
  - `Count(channel, countId, count, [`$K_{(S,E)}$`])`
  - `CountResponse(channel, countId, status)`
  - Reserved `countId`:
    - *subscriberId*, the number of subscribers in a subtree.
      1. Originated from source via `countQuery` or
      2. A receiver initiates a `newSubscription` will cause a *subscriberId* `Count` message.
    - *neighbors* and *all channels* for neighbor discovery.

- Distribution Tree Maintenance:
  - A router can select TCP or UDP for ECMP on each interface.

		# of neighbors	# of channels
TCP	core router	less	more
UDP	edge router	more	less

- In TCP, the router maintains:
  1. a TCP connection to each neighbor.
  2. a per-channel subscriber count for each neighbor.
- In UDP, the upstream router periodically multicasts a `CountQuery` request, then all the UDP neighbors respond with `Count` messages for the specified channel.

- Forwarding:
  - A router looks up (S,E) in *Forwarding Information Base (FIB)*, and forwards the packet to the outgoing interfaces.

    The router will drop the packet if:

    1. the incoming interface matches the FIB entry's outgoing interface, or
    2. the packet doesn't match any exact (S,E) entry in the FIB.

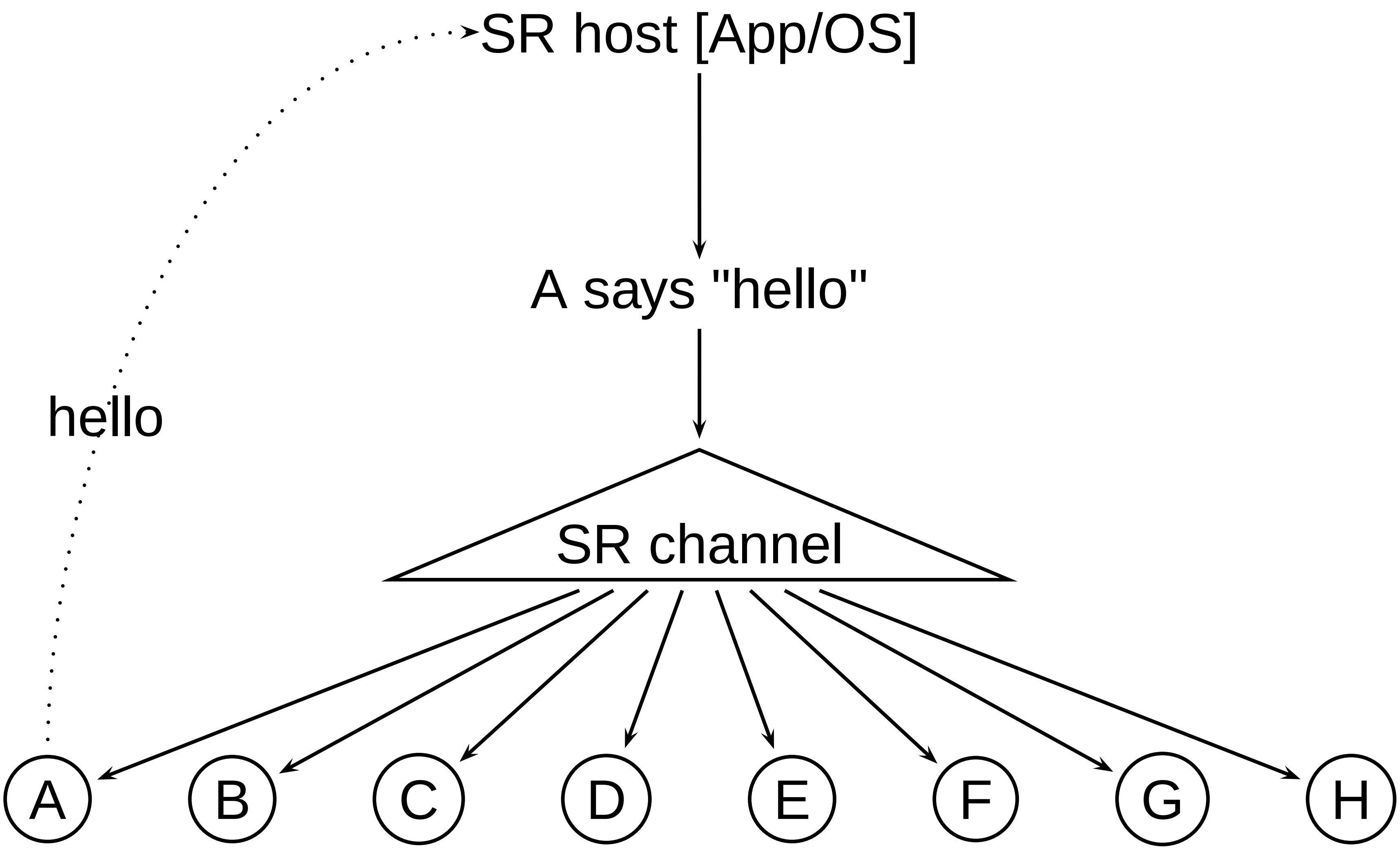
- Multicast can support one-to-many and many-to-many communication.
- Restricting multicast to single source really simplify the model, but how about Multi-Source multicast applications?

Multi-Source Multicast Applications

Large-scale multicast applications are almost single-source, the proposed solution is session relay approach.

- Solutions:
  1. Allowing several sources to share a channel using higher-level relaying through the channel's source host. (Session Relay)
  2. Use multiple channels, one per source.

Session Relay (SR) Approach



- Operation:
    1. Each SR-based application has an associated session relay on an application-selected host SR that acts as the source for the EXPRESS channel (SR,E).
    2. The primary lecturer or speaker either resides on the SR or unicast an enacpsulated packet to the SR.
    3. The SR can use an application-layer relay protocol or an IP-in-IP-like encapsulation.

- Advantages of session Relay:
  1. Reduce traffic by locate the SR near the topology center.
  2. Fault tolerance by additional backup SRs.
  3. provide application-specific function.
- Disadvantage of session Relay:
  1. Extra latency.

Cost Evaluation

- Cost for Router Memory
  - 0.18 cents per subscriber per *year*.
  - A small community cable TV channel can lease for approximately $1.00 per potential viewer per month.
- Cost of Management-Level State
- Cost of State Maintenance

Related Work

- Tree-building: CBT
- Source-specific join: PIM-SM
- Contribution:
  - specifying this model as an extension of IP multicast.
  - demonstrate its advantages for large-scale multicast applications.
  - adding the *accounting* support.
- Simple Multicast Routing Protocol develops an extension to the AppleTalk protocol supports a similar model of multicast.

- IGMPv3: inclusion/exclusion lists allow receiver to enumerate the set of souces that it wishes to hear from or exclude.
  It's more general but more complicated.
- BGMP: in combination with a global address allocation scheme can improve the scalability of IP multicast routing.
- Session Relay approach is similar in some ways to the PIM-SM rendezvous point, but session relays are selected and controlled at the application layer, not the network layer.
- Subcasting in EXPRESS is similar to the `SUBTREE_MCAST` in RMTP.

- Simple Multicast
  - supports multiple sources per multicast tree with bi-cirectional shared trees and appreas to require changes to the multicast forwarding fast path.
  - does Not address accounting or access control.
- RSVP focuses on resource allocation, rather than resource accounting.

Conclusion

- EXPRESS simplifies the implementation by the single-source restriction.
- Contributions:
  1. Addressing
  2. Tree maintenance
  3. Counting
  4. Scalability

References

[1] Christophe Diot, Brian Neil Levine, Bryan Lyles, Hassan Kassem, and Doug Balensiefen, “Deployment Issues for the IP Multicast Service and Architecture”, in *IEEE Network*, 2000.

[2] Kevin Almeroth, “The Evolution of Multicast: From the MBone to Inter-Domain Multicast to Internet2 Deployment Architecture”, in *IEEE Network*, 2000.

[3] Hugh W. Holbrook, David R. Cheriton, “Multicast channels: EXPRESS support for large-scale single-source applications”, in *Proceedings of ACM SIGCOMM*, 1999.