Network Working Group S. Deering
Request for Comments: 1112 Stanford University
Obsoletes: RFCs 988, 1054 August 1989

Host Extensions for IP Multicasting

1. STATUS OF THIS MEMO

This memo specifies the extensions required of a host implementation of the Internet Protocol (IP) to support multicasting. It is the recommended standard for IP multicasting in the Internet. Distribution of this memo is unlimited.

2. INTRODUCTION

IP multicasting is the transmission of an IP datagram to a "host group", a set of zero or more hosts identified by a single IP destination address. A multicast datagram is delivered to all members of its destination host group with the same "best-efforts" reliability as regular unicast IP datagrams, i.e., the datagram is not guaranteed to arrive intact at all members of the destination group or in the same order relative to other datagrams.

The membership of a host group is dynamic; that is, hosts may join and leave groups at any time. There is no restriction on the location or number of members in a host group. A host may be a member of more than one group at a time. A host need not be a member of a group to send datagrams to it.

A host group may be permanent or transient. A permanent group has a well-known, administratively assigned IP address. It is the address, not the membership of the group, that is permanent; at any time a permanent group may have any number of members, even zero. Those IP multicast addresses that are not reserved for permanent groups are available for dynamic assignment to transient groups which exist only as long as they have members.

Internetwork forwarding of IP multicast datagrams is handled by "multicast routers" which may be co-resident with, or separate from, internet gateways. A host transmits an IP multicast datagram as a local network multicast which reaches all immediately-neighboring members of the destination host group. If the datagram has an IP time-to-live greater than 1, the multicast router(s) attached to the local network take responsibility for forwarding it towards all other networks that have members of the destination group. On those other member networks that are reachable within the IP time-to-live, an attached multicast router completes delivery by transmitting the datagram as a local multicast.

This memo specifies the extensions required of a host IP implementation to support IP multicasting, where a "host" is any internet host or gateway other than those acting as multicast routers. The algorithms and protocols used within and between multicast routers are transparent to hosts and will be specified in separate documents. This memo also does not specify how local network multicasting is accomplished for all types of network, although it does specify the required service interface to an arbitrary local network and gives an Ethernet specification as an example. Specifications for other types of network will be the subject of future memos.

3. LEVELS OF CONFORMANCE

There are three levels of conformance to this specification:

Level 0: no support for IP multicasting.

There is, at this time, no requirement that all IP implementations support IP multicasting. Level 0 hosts will, in general, be unaffected by multicast activity. The only exception arises on some types of local network, where the presence of level 1 or 2 hosts may cause misdelivery of multicast IP datagrams to level 0 hosts. Such datagrams can easily be identified by the presence of a class D IP address in their destination address field; they should be quietly discarded by hosts that do not support IP multicasting. Class D addresses are described in section 4 of this memo.

Level 1: support for sending but not receiving multicast IP datagrams.

Level 1 allows a host to partake of some multicast-based services, such as resource location or status reporting, but it does not allow a host to join any host groups. An IP implementation may be upgraded from level 0 to level 1 very easily and with little new code. Only sections 4, 5, and 6 of this memo are applicable to level 1 implementations.

Level 2: full support for IP multicasting.

Level 2 allows a host to join and leave host groups, as well as send IP datagrams to host groups. It requires implementation of the Internet Group Management Protocol (IGMP) and extension of the IP and local network service interfaces within the host. All of the following sections of this memo are applicable to level 2 implementations.

4. HOST GROUP ADDRESSES

Host groups are identified by class D IP addresses, i.e., those with "1110" as their high-order four bits. Class E IP addresses, i.e., those with "1111" as their high-order four bits, are reserved for future addressing modes.

In Internet standard "dotted decimal" notation, host group addresses range from 224.0.0.0 to 239.255.255.255. The address 224.0.0.0 is guaranteed not to be assigned to any group, and 224.0.0.1 is assigned to the permanent group of all IP hosts (including gateways). This is used to address all multicast hosts on the directly connected network. There is no multicast address (or any other IP address) for all hosts on the total Internet. The addresses of other well-known, permanent groups are to be published in "Assigned Numbers".

Appendix II contains some background discussion of several issues related to host group addresses.

5. MODEL OF A HOST IP IMPLEMENTATION

The multicast extensions to a host IP implementation are specified in terms of the layered model illustrated below. In this model, ICMP and (for level 2 hosts) IGMP are considered to be implemented within the IP module, and the mapping of IP addresses to local network addresses is considered to be the responsibility of local network modules. This model is for expository purposes only, and should not be construed as constraining an actual implementation.

```
         |                                                           |
         |                Upper-Layer Protocol Modules               |
         |___________________________________________________________|

     --------------------- IP Service Interface -----------------------
          ___________________________________________________________
         |                            |              |               |
         |                            |     ICMP     |     IGMP      |
         |             IP             |______________|_______________|
         |           Module                                          |
         |                                                           |
         |___________________________________________________________|

     ---------------- Local Network Service Interface -----------------
          ___________________________________________________________
         |                            |                              |
         |           Local            | IP-to-local address mapping  |
         |          Network           |           (e.g., ARP)        |
         |          Modules           |______________________________|
         |      (e.g., Ethernet)                                     |
         |                                                           |
```

To provide level 1 multicasting, a host IP implementation must support the transmission of multicast IP datagrams. To provide level 2 multicasting, a host must also support the reception of multicast IP datagrams. Each of these two new services is described in a separate section, below. For each service, extensions are specified for the IP service interface, the IP module, the local network service interface, and an Ethernet local network module. Extensions to local network modules other than Ethernet are mentioned briefly, but are not specified in detail.

6. SENDING MULTICAST IP DATAGRAMS

6.1. Extensions to the IP Service Interface

Multicast IP datagrams are sent using the same "Send IP" operation used to send unicast IP datagrams; an upper-layer protocol module merely specifies an IP host group address, rather than an individual IP address, as the destination. However, a number of extensions may be necessary or desirable.

First, the service interface should provide a way for the upper-layer protocol to specify the IP time-to-live of an outgoing multicast datagram, if such a capability does not already exist. If the upper-layer protocol chooses not to specify a time-to-live, it should default to 1 for all multicast IP datagrams, so that an explicit choice is required to multicast beyond a single network.

Second, for hosts that may be attached to more than one network, the service interface should provide a way for the upper-layer protocol to identify which network interface is be used for the multicast transmission. Only one interface is used for the initial transmission; multicast routers are responsible for forwarding to any other networks, if necessary. If the upper-layer protocol chooses not to identify an outgoing interface, a default interface should be used, preferably under the control of system management.

Third (level 2 implementations only), for the case in which the host is itself a member of a group to which a datagram is being sent, the service interface should provide a way for the upper-layer protocol to inhibit local delivery of the datagram; by default, a copy of the datagram is looped back. This is a performance optimization for upper-layer protocols that restrict the membership of a group to one process per host (such as a routing protocol), or that handle loopback of group communication at a higher layer (such as a multicast transport protocol).

6.2. Extensions to the IP Module

To support the sending of multicast IP datagrams, the IP module must be extended to recognize IP host group addresses when routing outgoing datagrams. Most IP implementations include the following logic:

```
     if IP-destination is on the same local network,
        send datagram locally to IP-destination
     else
        send datagram locally to GatewayTo( IP-destination )
```

To allow multicast transmissions, the routing logic must be changed to:

```
if IP-destination is on the same local network
or IP-destination is a host group,
   send datagram locally to IP-destination
else
   send datagram locally to GatewayTo( IP-destination )
```

If the sending host is itself a member of the destination group on the outgoing interface, a copy of the outgoing datagram must be looped-back for local delivery, unless inhibited by the sender. (Level 2 implementations only.)

The IP source address of the outgoing datagram must be one of the individual addresses corresponding to the outgoing interface.

A host group address must never be placed in the source address field or anywhere in a source route or record route option of an outgoing IP datagram.

6.3. Extensions to the Local Network Service Interface

No change to the local network service interface is required to support the sending of multicast IP datagrams. The IP module merely specifies an IP host group destination, rather than an individual IP destination, when it invokes the existing "Send Local" operation.

6.4. Extensions to an Ethernet Local Network Module

The Ethernet directly supports the sending of local multicast packets by allowing multicast addresses in the destination field of Ethernet packets. All that is needed to support the sending of multicast IP datagrams is a procedure for mapping IP host group addresses to Ethernet multicast addresses.

An IP host group address is mapped to an Ethernet multicast address by placing the low-order 23-bits of the IP address into the low-order 23 bits of the Ethernet multicast address 01-00-5E-00-00-00 (hex). Because there are 28 significant bits in an IP host group address, more than one host group address may map to the same Ethernet multicast address.

6.5. Extensions to Local Network Modules other than Ethernet

Other networks that directly support multicasting, such as rings or buses conforming to the IEEE 802.2 standard, may be handled the same way as Ethernet for the purpose of sending multicast IP datagrams. For a network that supports broadcast but not multicast, such as the Experimental Ethernet, all IP host group addresses may be mapped to a single local broadcast address (at the cost of increased overhead on all local hosts). For a point-to-point link joining two hosts (or a host and a multicast router), multicasts should be transmitted exactly like unicasts. For a store-and-forward network like the ARPANET or a public X.25 network, all IP host group addresses might be mapped to the well-known local address of an IP multicast router; a router on such a network would take responsibility for completing multicast delivery within the network as well as among networks.

7. RECEIVING MULTICAST IP DATAGRAMS

7.1. Extensions to the IP Service Interface

Incoming multicast IP datagrams are received by upper-layer protocol modules using the same "Receive IP" operation as normal, unicast datagrams. Selection of a destination upper-layer protocol is based on the protocol field in the IP header, regardless of the destination IP address. However, before any datagrams destined to a particular group can be received, an upper-layer protocol must ask the IP module to join that group. Thus, the IP service interface must be extended to provide two new operations:

```
JoinHostGroup  ( group-address, interface )

LeaveHostGroup ( group-address, interface )
```

The JoinHostGroup operation requests that this host become a member of the host group identified by "group-address" on the given network interface. The LeaveGroup operation requests that this host give up its membership in the host group identified by "group-address" on the given network interface. The interface argument may be omitted on hosts that support only one interface. For hosts that may be attached to more than one network, the upper-layer protocol may choose to leave the interface unspecified, in which case the request will apply to the default interface for sending multicast datagrams (see section 6.1).

It is permissible to join the same group on more than one interface, in which case duplicate multicast datagrams may be received. It is also permissible for more than one upper-layer protocol to request membership in the same group.

Both operations should return immediately (i.e., they are non-blocking operations), indicating success or failure. Either operation may fail due to an invalid group address or interface identifier. JoinHostGroup may fail due to lack of local resources. LeaveHostGroup may fail because the host does not belong to the given group on the given interface. LeaveHostGroup may succeed, but the membership persist, if more than one upper-layer protocol has requested membership in the same group.

7.2. Extensions to the IP Module

To support the reception of multicast IP datagrams, the IP module must be extended to maintain a list of host group memberships associated with each network interface. An incoming datagram destined to one of those groups is processed exactly the same way as datagrams destined to one of the host's individual addresses.

Incoming datagrams destined to groups to which the host does not belong are discarded without generating any error report or log entry. On hosts with more than one network interface, if a datagram arrives via one interface, destined for a group to which the host belongs only on a different interface, the datagram is quietly discarded. (These cases should occur only as a result of inadequate multicast address filtering in a local network module.)

An incoming datagram is not rejected for having an IP time-to-live of 1 (i.e., the time-to-live should not automatically be decremented on arriving datagrams that are not being forwarded). An incoming datagram with an IP host group address in its source address field is quietly discarded. An ICMP error message (Destination Unreachable, Time Exceeded, Parameter Problem, Source Quench, or Redirect) is never generated in response to a datagram destined to an IP host group.

The list of host group memberships is updated in response to JoinHostGroup and LeaveHostGroup requests from upper-layer protocols. Each membership should have an associated reference count or similar mechanism to handle multiple requests to join and leave the same group. On the first request to join and the last request to leave a group on a given interface, the local network module for that interface is notified, so that it may update its multicast reception filter (see section 7.3).

The IP module must also be extended to implement the IGMP protocol, specified in Appendix I. IGMP is used to keep neighboring multicast routers informed of the host group memberships present on a particular local network. To support IGMP, every level 2 host must join the "all-hosts" group (address 224.0.0.1) on each network interface at initialization time and must remain a member for as long as the host is active.

(Datagrams addressed to the all-hosts group are recognized as a special case by the multicast routers and are never forwarded beyond a single network, regardless of their time-to-live. Thus, the all-hosts address may not be used as an internet-wide broadcast address. For the purpose of IGMP, membership in the all-hosts group is really necessary only while the host belongs to at least one other group. However, it is specified that the host shall remain a member of the all-hosts group at all times because (1) it is simpler, (2) the frequency of reception of unnecessary IGMP queries should be low enough that overhead is negligible, and (3) the all-hosts address may serve other routing-oriented purposes, such as advertising the presence of gateways or resolving local addresses.)

7.3. Extensions to the Local Network Service Interface

Incoming local network multicast packets are delivered to the IP module using the same "Receive Local" operation as local network unicast packets. To allow the IP module to tell the local network module which multicast packets to accept, the local network service interface is extended to provide two new operations:

```
JoinLocalGroup  ( group-address )

LeaveLocalGroup ( group-address )
```

where "group-address" is an IP host group address. The JoinLocalGroup operation requests the local network module to accept and deliver up subsequently arriving packets destined to the given IP host group address. The LeaveLocalGroup operation requests the local network module to stop delivering up packets destined to the given IP host group address. The local network module is expected to map the IP host group addresses to local network addresses as required to update its multicast reception filter. Any local network module is free to ignore LeaveLocalGroup requests, and may deliver up packets destined to more addresses than just those specified in JoinLocalGroup requests, if it is unable to filter incoming packets adequately.

The local network module must not deliver up any multicast packets that were transmitted from that module; loopback of multicasts is handled at the IP layer or higher.

7.4. Extensions to an Ethernet Local Network Module

To support the reception of multicast IP datagrams, an Ethernet module must be able to receive packets addressed to the Ethernet multicast addresses that correspond to the host's IP host group addresses. It is highly desirable to take advantage of any address filtering capabilities that the Ethernet hardware interface may have, so that the host receives only those packets that are destined to it.

Unfortunately, many current Ethernet interfaces have a small limit on the number of addresses that the hardware can be configured to recognize. Nevertheless, an implementation must be capable of listening on an arbitrary number of Ethernet multicast addresses, which may mean "opening up" the address filter to accept all multicast packets during those periods when the number of addresses exceeds the limit of the filter.

For interfaces with inadequate hardware address filtering, it may be desirable (for performance reasons) to perform Ethernet address filtering within the software of the Ethernet module. This is not mandatory, however, because the IP module performs its own filtering based on IP destination addresses.

7.5. Extensions to Local Network Modules other than Ethernet

Other multicast networks, such as IEEE 802.2 networks, can be handled the same way as Ethernet for the purpose of receiving multicast IP datagrams. For pure broadcast networks, such as the Experimental Ethernet, all incoming broadcast packets can be accepted and passed to the IP module for IP-level filtering. On point-to-point or store-and-forward networks, multicast IP datagrams will arrive as local network unicasts, so no change to the local network module should be necessary.

APPENDIX I. INTERNET GROUP MANAGEMENT PROTOCOL (IGMP)

The Internet Group Management Protocol (IGMP) is used by IP hosts to report their host group memberships to any immediately-neighboring multicast routers. IGMP is an asymmetric protocol and is specified here from the point of view of a host, rather than a multicast router. (IGMP may also be used, symmetrically or asymmetrically, between multicast routers. Such use is not specified here.)

Like ICMP, IGMP is a integral part of IP. It is required to be implemented by all hosts conforming to level 2 of the IP multicasting specification. IGMP messages are encapsulated in IP datagrams, with an IP protocol number of 2. All IGMP messages of concern to hosts have the following format:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Version| Type  |    Unused     |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Group Address                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Version

This memo specifies version 1 of IGMP. Version 0 is specified in RFC-988 and is now obsolete.

Type

There are two types of IGMP message of concern to hosts:

1 = Host Membership Query
2 = Host Membership Report

Unused

Unused field, zeroed when sent, ignored when received.

Checksum

The checksum is the 16-bit one's complement of the one's complement sum of the 8-octet IGMP message. For computing the checksum, the checksum field is zeroed.

Group Address

In a Host Membership Query message, the group address field is zeroed when sent, ignored when received.

In a Host Membership Report message, the group address field holds the IP host group address of the group being reported.

Informal Protocol Description

Multicast routers send Host Membership Query messages (hereinafter called Queries) to discover which host groups have members on their attached local networks. Queries are addressed to the all-hosts group (address 224.0.0.1), and carry an IP time-to-live of 1.

Hosts respond to a Query by generating Host Membership Reports (hereinafter called Reports), reporting each host group to which they belong on the network interface from which the Query was received. In order to avoid an "implosion" of concurrent Reports and to reduce the total number of Reports transmitted, two techniques are used:

1. When a host receives a Query, rather than sending Reports immediately, it starts a report delay timer for each of its group memberships on the network interface of the incoming Query. Each timer is set to a different, randomly-chosen value between zero and D seconds. When a timer expires, a Report is generated for the corresponding host group. Thus, Reports are spread out over a D second interval instead of all occurring at once.

2. A Report is sent with an IP destination address equal to the host group address being reported, and with an IP time-to-live of 1, so that other members of the same group on the same network can overhear the Report. If a host hears a Report for a group to which it belongs on that network, the host stops its own timer for that group and does not generate a Report for that group. Thus, in the normal case, only one Report will be generated for each group present on the network, by the member host whose delay timer expires first. Note that the multicast routers receive all IP multicast datagrams, and therefore need not be addressed explicitly. Further note that the routers need not know which hosts belong to a group, only that at least one host belongs to a group on a particular network.

There are two exceptions to the behavior described above. First, if a report delay timer is already running for a group membership when a Query is received, that timer is not reset to a new random value, but rather allowed to continue running with its current value. Second, a report delay timer is never set for a host's membership in the all-hosts group (224.0.0.1), and that membership is never reported.

If a host uses a pseudo-random number generator to compute the reporting delays, one of the host's own individual IP address should be used as part of the seed for the generator, to reduce the chance of multiple hosts generating the same sequence of delays.

A host should confirm that a received Report has the same IP host group address in its IP destination field and its IGMP group address field, to ensure that the host's own Report is not cancelled by an erroneous received Report. A host should quietly discard any IGMP message of type other than Host Membership Query or Host Membership Report.

Multicast routers send Queries periodically to refresh their knowledge of memberships present on a particular network. If no Reports are received for a particular group after some number of Queries, the routers assume that that group has no local members and that they need not forward remotely-originated multicasts for that group onto the local network. Queries are normally sent infrequently (no more than once a minute) so as to keep the IGMP overhead on hosts and networks very low. However, when a multicast router starts up, it may issue several closely-spaced Queries in order to build up its knowledge of local memberships quickly.

When a host joins a new group, it should immediately transmit a Report for that group, rather than waiting for a Query, in case it is the first member of that group on the network. To cover the possibility of the initial Report being lost or damaged, it is recommended that it be repeated once or twice after short delays. (A simple way to accomplish this is to act as if a Query had been received for that group only, setting the group's random report delay timer. The state transition diagram below illustrates this approach.)

Note that, on a network with no multicast routers present, the only IGMP traffic is the one or more Reports sent whenever a host joins a new group.

State Transition Diagram

IGMP behavior is more formally specified by the state transition diagram below. A host may be in one of three possible states, with respect to any single IP host group on any single network interface:

- Non-Member state, when the host does not belong to the group on the interface. This is the initial state for all memberships on all network interfaces; it requires no storage in the host.

- Delaying Member state, when the host belongs to the group on the interface and has a report delay timer running for that membership.

- Idle Member state, when the host belongs to the group on the interface and does not have a report delay timer running for that membership.

There are five significant events that can cause IGMP state transitions:

- "join group" occurs when the host decides to join the group on the interface. It may occur only in the Non-Member state.

- "leave group" occurs when the host decides to leave the group on the interface. It may occur only in the Delaying Member and Idle Member states.

- "query received" occurs when the host receives a valid IGMP Host Membership Query message. To be valid, the Query message must be at least 8 octets long, have a correct IGMP checksum and have an IP destination address of 224.0.0.1. A single Query applies to all memberships on the interface from which the Query is received. It is ignored for memberships in the Non-Member or Delaying Member state.

- "report received" occurs when the host receives a valid IGMP Host Membership Report message. To be valid, the Report message must be at least 8 octets long, have a correct IGMP checksum, and contain the same IP host group address in its IP destination field and its IGMP group address field. A Report applies only to the membership in the group identified by the Report, on the interface from which the Report is received. It is ignored for memberships in the Non-Member or Idle Member state.

- "timer expired" occurs when the report delay timer for the group on the interface expires. It may occur only in the Delaying Member state.

All other events, such as receiving invalid IGMP messages, or IGMP messages other than Query or Report, are ignored in all states.

There are three possible actions that may be taken in response to the above events:

- "send report" for the group on the interface.

- "start timer" for the group on the interface, using a random delay value between 0 and D seconds.

- "stop timer" for the group on the interface.

In the following diagram, each state transition arc is labelled with the event that causes the transition, and, in parentheses, any actions taken during the transition.

```
                         ________________
                        |                |
                        |                |
                        |                |
                        |                |
              --------->|   Non-Member   |<---------
             |          |                |          |
             |          |                |          |
             |          |                |          |
             |          |________________|          |
             |                   |                  |
             | leave group       | join group       | leave group
             | (stop timer)      |(send report,     |
             |                   | start timer)     |
     ________|________           |          ________|________
    |                 |<---------          |                 |
    |                 |                    |                 |
    |                 |<-------------------|                 |
    |                 |   query received   |                 |
    | Delaying Member |    (start timer)   |   Idle Member   |
    |                 |------------------->|                 |
    |                 |   report received  |                 |
    |                 |    (stop timer)    |                 |
    |_________________|------------------->|_________________|
                           timer expired
                           (send report)
```

The all-hosts group (address 224.0.0.1) is handled as a special case. The host starts in Idle Member state for that group on every interface, never transitions to another state, and never sends a report for that group.

Protocol Parameters

The maximum report delay, D, is 10 seconds.

APPENDIX II. HOST GROUP ADDRESS ISSUES

This appendix is not part of the IP multicasting specification, but provides background discussion of several issues related to IP host group addresses.

Group Address Binding

The binding of IP host group addresses to physical hosts may be considered a generalization of the binding of IP unicast addresses. An IP unicast address is statically bound to a single local network interface on a single IP network. An IP host group address is dynamically bound to a set of local network interfaces on a set of IP networks.

It is important to understand that an IP host group address is NOT bound to a set of IP unicast addresses. The multicast routers do not need to maintain a list of individual members of each host group. For example, a multicast router attached to an Ethernet need associate only a single Ethernet multicast address with each host group having local members, rather than a list of the members' individual IP or Ethernet addresses.

Allocation of Transient Host Group Addresses

This memo does not specify how transient group address are allocated. It is anticipated that different portions of the IP transient host group address space will be allocated using different techniques. For example, there may be a number of servers that can be contacted to acquire a new transient group address. Some higher-level protocols (such as VMTP, specified in RFC-1045) may generate higher-level transient "process group" or "entity group" addresses which are then algorithmically mapped to a subset of the IP transient host group addresses, similarly to the way that IP host group addresses are mapped to Ethernet multicast addresses. A portion of the IP group address space may be set aside for random allocation by applications that can tolerate occasional collisions with other multicast users, perhaps generating new addresses until a suitably "quiet" one is found.

In general, a host cannot assume that datagrams sent to any host group address will reach only the intended hosts, or that datagrams received as a member of a transient host group are intended for the recipient. Misdelivery must be detected at a level above IP, using higher-level identifiers or authentication tokens. Information transmitted to a host group address should be encrypted or governed by administrative routing controls if the sender is concerned about unwanted listeners.

Author's Address

Steve Deering
Stanford University
Computer Science Department
Stanford, CA 94305-2140

Phone: (415) 723-9427

EMail: deering@PESCADERO.STANFORD.EDU