Internet Engineering Task Force
INTERNET-DRAFT
draft-ietf-magma-msnip-02.ps

MAGMA WG
Bill Fenner/AT&T
Brian Haberman/Caspian Networks
Hugh Holbrook/Cisco
Isidor Kouvelas/Cisco
1 March 2003
Expires: September 2003

Multicast Source Notification of Interest Protocol (MSNIP)

Status of this Document

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This document is a product of the IETF MAGMA WG. Comments should be addressed to the authors, or the WG's mailing list at magma@ietf.org.

Abstract

This document discusses the Multicast Source Interest Notification Protocol (MSNIP). MSNIP is an extension to IGMPv3 and MLDv2 that provides membership notification services for sources of multicast traffic.

Table of Contents

1. Introduction

The Multicast Source Notification of Interest Protocol (MSNIP) is an extension to version 3 of the Internet Group Membership Protocol (IGMPv3 [1]) and version 2 of the Multicast Listener Discovery Protocol (MLDv2 [2]). MSNIP operates between multicast sources and their first-hop routers to provide information on the presence of receivers to the source systems. Using the services offered by MSNIP an application on an IP system wishing to source multicast data can register to be notified when receivers join and leave the session. This enables multicast sources to avoid the work of transmitting packets onto their first-hop link when there are no joined receivers.

A common scenario where MSNIP may be useful is one where there is a multicast server offering a large pool of potential flows that map onto separate multicast destination addresses but receivers exist only for a small subset of the flows. If the source were to continuously transmit data for all the flows that could potentially have receivers, significant resources would be wasted in the system itself as well as the first-hop link. Using a higher level control protocol to determine the existence of receivers for particular flows may not be practical as there may be many potential receivers in each active session.

Information on which multicast destination addresses have receivers for a particular sender is typically available to the multicast routing protocol on the first hop router for a source. MSNIP uses this information to notify the application in the sending system of when it should start or stop transmitting. This is achieved without any destination address specific state on the first-hop router for potential sources without receivers.

2. Routing Protocol Support

For reasons described in this section, MSNIP only supports transmission control for applications that use multicast destination addresses that are routed using Source Specific Multicast (SSM).

Many currently deployed multicast routing protocols, require data from an active source to be propagated past the first-hop router before information on the existence of receivers becomes available on the first-hop. In addition, such protocols require that this activity is repeated periodically to maintain source liveness state on remote routers. All dense-mode protocols fall under this category as well as sparse-mode protocols that use shared trees for source discovery (such as PIM-SM [9]). In order to provide receiver interest notification for such protocols, the default mode of operation would require that the source IP system periodically transmits on all potential destination addresses and the first-hop routers prune the traffic back. Such a flood-and-prune behaviour on the first-hop link significantly diminishes the benefits of managing source transmission.

In contrast, with source-specific sparse-mode protocols such as PIM-SSM [9] availability of receiver membership information on the first-hop routers is independent of data transmission. Such protocols use an external mechanism for source discovery (like source-specific IGMPv3 membership reports) to build source-specific multicast trees.

Clearly these two classes of routing protocols require different handling for the problem MSNIP is trying to solve. In addition the second type covers the majority of applications that MSNIP is targeted at. MSNIP avoids the extra complication in supporting routing protocols that require a flood and prune behaviour.

3. Service Interface for Requesting Membership Notification

Applications within an IP system that wish to source multicast packets and are interested in being notified on the existence of receivers must register with the IP layer of the system. MSNIP requires that within the IP system, there is (at least conceptually) a service interface that can be used to register with the IP layer for such notifications. Dual stack systems supporting both IPv4 and IPv6 need to provide separate service interfaces for each protocol.

A system's IPv4 or IPv6 service interface must support the following operation or any logical equivalent:

IPMulticastsSourceRegister (socket, source-address, multicast-address)
IPMulticastsSourceDeregister (socket, source-address, multicast-address)

In addition the application must provide the following interface for receiving notifications from the IP system:

IPMulticastSourceStart (socket, source-address, multicast-address)
IPMulticastSourceStop (socket, source-address, multicast-address)

where:

socket
is an implementation-specific parameter used to distinguish amongst different requesting entities (e.g., programs or processes) within the system; the socket parameter of BSD Unix system calls is a specific example.

source-address
is the IP unicast source address that the application wishes to use in transmitting data to the specified multicast address. The specified address must be one of the IP addresses associated with the network interfaces of the IP system. Note that an interface in an IP system may be associated with more than one IP addresses. An implementation may allow a special "unspecified" value to be passed as the source-address parameter, in which case the request would apply to the "primary" IP address of the "primary" or "default" interface of the system (perhaps established by system configuration). If transmission to the same multicast address is desired using more than one source IP address, IPMulticastSourceRegister must be invoked separately for each desired source address.

multicast-address
is the IP multicast destination address to which the request pertains. If the application wishes to transmit data to more than one multicast addresses for a given source address, IPMulticastSourceRegister must be invoked separately for each desired multicast address.

3.1. Application Operation

Applications wishing to use MSNIP to control their multicast data transmission to destination G from source address S operate as follows.

Initially the application contacts the IP system to obtain the socket handle for use on all subsequent interactions. The application invokes IPMulticastSourceRegister for the desired S and G and then waits without transmitting any packets for the IP system to notify that receivers for the session exist.

If and when the IP system notifies the application that receivers exist using the IPMulticastSourceStart call, the application may start transmitting data. After the application has been notified to send, if all receivers for the session leave, the IP system will notify the application using the IPMulticastSourceStop call. At this point the application should stop transmitting data until it is notified again that receivers have joined through another

IPMulticastSourceStart call.

When the application no longer wishes to transmit data it should invoke the IPMulticastsSourceDeregister call to let the IP system know that it is no longer interested in notifications for this source and destination. The IPMulticastsSourceDeregister call should be implicit in the teardown of the associated socket state.

4. MSNIP Managed Address Range Negotiation

With current multicast deployment in the Internet, different multicast routing protocols coexist and operate under separate parts of the multicast address space. Multicast routers are consistently configured with information that maps specific multicast address ranges to multicast routing protocols. Part of this configuration describes the subset of the address space that is used by source-specific multicast (SSM) [10]. As described in section 2 MSNIP only tries to control application transmission within the SSM address range.

It is desirable for applications within an IP system that supports MSNIP to have a consistent service interface for multicast transmission that does require the application to be aware of the SSM address range. MSNIP supports this by allowing applications to use the service interface described in section 3 for multicast destination addresses that are outside its operating range. When an application registers for notifications for a destination address that is not managed by MSNIP it is immediately notified to start transmitting. This is equivalent to the default behaviour of IP multicast without MSNIP support which forces multicast applications to assume that there are multicast receivers present in the network.

4.1. Router Coordination

In order for MSNIP to operate on a shared link where more than one multicast routers may be present all multicast routers must be MSNIP capable and have a consistent configuration for the SSM address range. MSNIP enforces these requirements by using two new options for IPv4 in the Multicast Router Discovery protocol [3] and one new option for IPv6 in Neighbor Discovery / ICMPv6 protocol.

4.1.1. MSNIP Operation Option

A multicast router advertises that it is participating in MSNIP using the MSNIP Operation option in either the Multicast Router Discovery protocol for IPv4 or the Neighbor Discovery / ICMPv6 protocol for IPv6. This option MUST be included in all router advertisement messages of a router that is configured for MSNIP. The format of the option is as follows:

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Length=0    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type

The type field is set to WW (TBD by IANA) for IPv4 and ZZ (TBD by IANA) for IPv6.

A multicast router uses received Multicast Router Advertisement and Neighbor Discovery / ICMPv6 messages to determine if all live neighbour multicast routers on each interface are participating in MSNIP. When a router advertisement message not containing an MSNIP option is received by a router participating in MSNIP, a miss-configuration SHOULD be logged to the operator in a rate-limited manner.

If even one multicast router on a link does not have MSNIP capability then ALL routers on that link MUST be configured to not provide MSNIP services and to not advertise the MSNIP Operation option.

4.1.2. SSM Range Option

The SSM Range Multicast Router Discover option advertises the SSM Range with which the router is configured. The option is defined in [4]. This option is only valid in IPv4. SSM support in IPv6 does not allow for alternative SSM address ranges.

4.2. Communicating Range to Source Systems

When an application in an IP system uses the MSNIP service interface to register for notification, the IP system must behave differently depending on whether or not the destination address for which the application registered is operating under SSM (and is being managed by MSNIP). For SSM channels, the IP system should only instruct the application to transmit when there are receivers for the multicast destination. For non-SSM destination addresses the IP system will not be able to determine if there are receivers and should immediately instruct the application to transmit.

The MSNIP managed range discovery mechanism in a source IP system has to deal with three different link configurations:

- A link connected to a multicast routed infrastructure where the first-hop multicast routers are configured for MSNIP operation.
- A link connected to a multicast routed infrastructure where the first-hop multicast routers are not participating in MSNIP.
- A link with no multicast routers.

To be able to differentiate between the three cases and in each case discover the MSNIP managed range an MSNIP capable source IP system must process IGMP Queries as well as Multicast Router Discovery Advertisement messages. The presence of an IGMP querier differentiates between the first two cases, where the host is in a multicast routed network, and the third, where there is no multicast routing.

Multicast Router Discovery Advertisements provide the differentiation between the first two cases. If the MRD Advertisements contain the MSNIP Operation option then the IP system knows that routers on that interface are configured for MSNIP operation.

In each of the three cases the MSNIP managed range is defined as follows:

MSNIP capable multicast routers:
The IP system should use as the MSNIP range the SSM range provided by the last SSM Range option [4] received in a Multicast Router Discovery message.

Multicast routers not participating in MSNIP:
The IP system should use an empty MSNIP managed range. This provides a compatibility mode where all group ranges default to flooding.

Link without multicast routing:

To ensure backwards compatibility with existing receivers, MSNIP does not try to control traffic on a router-less link. It does so by defining the MSNIP managed range to be empty. Although it would be possible to control multicast transmission on a shared link not connected to a multicast routed infrastructure, MSNIP operation would require that receivers were capable of actively participating in the protocol. Source control would work by defining an address range within which sources would not transmit until directly contacted by a receiver (for example the default IPv4 SSM range of 232/8 [6]). The drawback would be that a non MSNIP capable receiver joining a group through IGMP would have no way of getting the source to transmit. Relying on triggered IGMP messages from legacy receivers to control transmission would not be robust either.

5. Requesting and Receiving Notifications

Like IGMP, MSNIP is an asymmetric protocol specifying different behaviour for systems wishing to source traffic and for multicast routers. Host IP systems multicast Host Interest Solicitation messages to register for notification with their first-hop routers. Routers unicast Router Receiver Membership Reports to IP systems to notify them of the arrival of the first or departure of the last receivers for a session. Note that a system may perform at the same time both of the above functions. An example is a router that wishes to source traffic.

5.1. Host Interest Solicitation

Source systems that wish to be managed by MSNIP periodically transmit an Interest Solicitation message. This message is multicast with a multicast destination address of ALL_IGMPv3_ROUTERS (224.0.0.22) or ALL_MLDv2_ROUTERS (TBA) and is transmitted every [Interest Solicitation Interval] seconds. The Interest Solicitation message contains a holdtime which is set to [Interest Solicitation Holdtime] and instructs the multicast first-hop routers to maintain MSNIP state for this IP system for the specified period. A generation ID is also included in the Interest Solicitation message to provide support for routers to detect IP system restarts (see section 8.1). Systems with multiple interfaces or multiple IP addresses per interface must originate separate Host Interest Solicitation messages from each of their IP addresses that they wish to have managed by MSNIP. In practice a system with more than one IP address is treated by MSNIP as multiple IP systems.

When an IP system first comes up it transmits [Robustness Variable] Interest Solicitation messages spaced by [Initial Interest Solicitation Interval] seconds.

All MSNIP capable routers that receive an Interest Solicitation message from an IP system, maintain a system interest record of the form:

(IP system address, holdtime timer)

Each time an Interest Solicitation message is received from the IP system, the holdtime timer is reset to the holdtime in the received message. In addition the router may respond to the solicitation message with a Receiver Membership Report message described in section 5.2. The message contains a TRANSMIT record for each of the multicast destination addresses within the MSNIP managed range for which the routing protocol indicates there are receivers for this source system.

The holdtime timer of a record counts down to zero. When the holdtime timer of a specific system interest record expires, the record is deleted.

5.2. Router Receiver Membership Reports

Receiver Membership Report messages are used by routers to communicate the receiver membership status of particular multicast destination addresses to a specific IP system. Each message contains a list of transmission control records of either TRANSMIT or HOLD type that instruct a system to respectively start or stop sending traffic on this link to the specified multicast destination address. Receiver Membership Report messages are unicast to the target system.

In addition to reports sent in response to Interest Solicitation messages, routers send unsolicited Receiver Membership Reports to IP systems when the receiver membership status reported by the multicast routing protocol changes for a specific source and multicast destination. Such reports are only sent if the destination address is managed by MSNIP and the router has a system interest record created by a previously received Interest Solicitation message with an IP system address equal to the source address. If the source destination pair satisfy these conditions then [Robustness Variable] Receiver Membership Reports are sent out spaced by [Unsolicited Membership Report Interval] seconds. If the membership status changes again for the same destination address and source system while transmission of Receiver Membership Reports is still pending then the pending report messages are canceled and a new set of [Robustness Variable] messages indicating the new state are scheduled.

When an IP system receives a Receiver Membership Report message, for each of the TRANSMIT records listed in the message it creates or updates a transmission record of the form:

(router address, source address, multicast address, holdtime timer)

The router address is obtained from the source address on the IP header of the received message. The source address is obtained from the destination address in the of the IP header. The holdtime timer is set to the value of the holdtime field in the received Receiver Membership Report message.

For each HOLD record present in the message, the system deletes the matching previously created transmission record from its state.

The holdtime timer of a record counts down to zero. When the holdtime timer of a specific transmission record expires, the record is deleted.

Note that creation and deletion of transmission records in an IP systems state may cause local applications to be notified to start and stop transmitting data (see section 6).

6. Application Notification

This section describes the relation between protocol events and notifications to source applications within an IP system. The state machine below is specific to each source address of the IP system and each multicast destination address. The initial state is the No Info state.

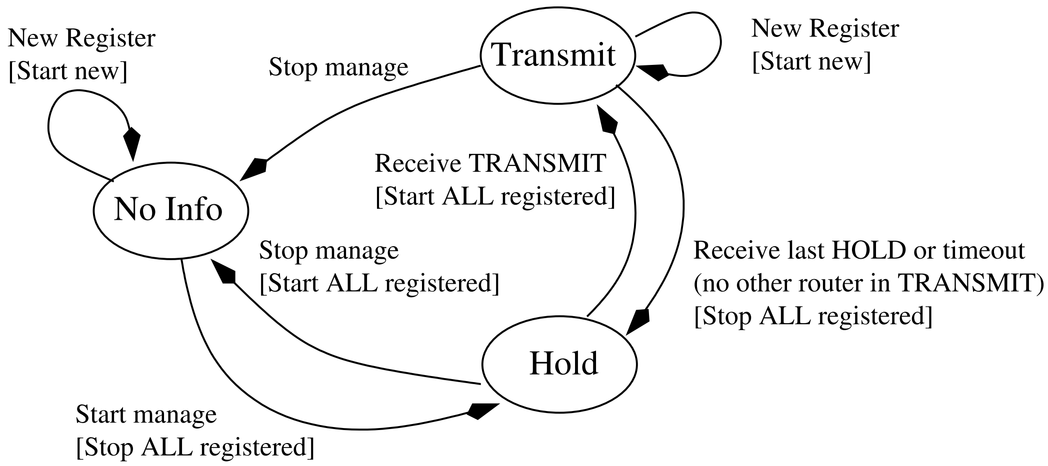

Figure 1: Per source-address (S) and multicast destination address (G) state machine at an IP system

In tabular form, the state-machine is:

Prev State	Event: New Register	Start Manage	Stop Manage	Recv TRANSMIT	Recv last HOLD or timeout
No Info	Start new	-> Hold Stop ALL registered	-	-	-
Hold	-	-	-> No Info Stop ALL registered	-> Transmit Start ALL registered	-
Transmit	Start new	-	-> No Info	-	-> Hold Stop ALL registered

The events in state machine above have the following meaning:

New register
A new application has registered through a call to IPMulticastsSourceRegister for this S and G.

Start manage
We received a SSM Range option in an MRD packet on the interface that S belongs to that changed the status of G from a non-managed to a MSNIP managed destination address. The SSM Range option is only valid in IPv4.

Stop manage
We received a SSM Range option in an MRD packet on the interface that S belongs to that changed the status of G from a MSNIP managed to a non-managed destination address or the mapping state that caused this destination address to be managed expired. The SSM

Range option is only valid in IPv4.

Receive TRANSMIT
We received a Receiver Membership Report with S as the IP destination address that contains a TRANSMIT record for G.

Receive last HOLD or timeout
We either received a Receiver Membership Report with S as the IP destination address that contains a HOLD record for G or the holdtime timer in a transmission record for S and G expired and there are no other transmission records for S and G. This means that the last router that was reporting receivers no longer does so and there are no routers left wishing to receive traffic from this S to destination address G.

The state machine actions have the following meaning:

Start new
Send an IPMulticastSourceStart notification to the application that just registered for this S and G.

Start ALL registered
Send an IPMulticastSourceStart notification to all applications that are registered for this S and G.

Stop ALL registered
Send an IPMulticastSourceStop notification to all applications that are registered for this S and G.

7. Router Processing

This section describes the per-source system tracking state machine operated by each first-hop router. The initial state is No Info.

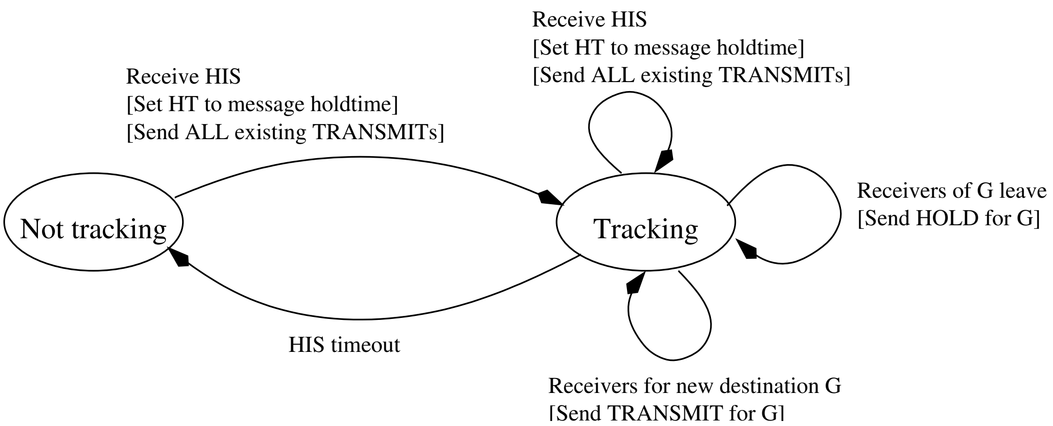

Figure 2: Per IP source system (S) state machine at a router

In tabular form, the state-machine is:

Prev State	Event: Receive HIS	Event: HIS timeout	Event: Receivers for new destination G	Event: Receivers of G leave
Not tracking	-> Tracking Set HT to message holdtime Send ALL existing TRANSMITs	-	-	-
Tracking	Set HT to message holdtime Send ALL existing TRANSMITs	-> Not tracking	Send TRANSMIT for G	Send HOLD for G

The events in state machine above have the following meaning:

Receive HIS
The router has received a Host Interest Solicitation from S.

HIS timeout
The holdtime timer (HT) in the host interest record associated with S has expired.

Receivers for new destination G
The routing protocol has informed MSNIP that it now has receivers for the MSNIP managed destination address G and source IP system S.

Receivers of G leave
The routing protocol has informed MSNIP that all receivers for the MSNIP managed destination address G and source IP system S have left the channel.

The state machine actions have the following meaning:

Set HT to message holdtime
The holdtime timer in the host interest record associated with S is restarted to the value of the holdtime field in the received Host Interest Solicitation message.

Send ALL existing TRANSMITs
The router builds and transmits Receiver Membership Reports to S that contain a TRANSMIT record for each of the MSNIP managed destination addresses that have receivers for S.

Send TRANSMIT for G

The router builds and transmits a Receiver Membership Report to S that contains a TRANSMIT record for the destination address G.

Send HOLD for G

The router builds and transmits a Receiver Membership Report to S that contains a HOLD record for the destination address G.

8. Message Formats

The following packet formats are valid for both IPv4 and IPv6. IP version specific values will be explicitly defined.

There are two message types of concern to the MSNIP protocol described in this document:

Type Number (hex)	Message Name
0xXX	Host Interest Solicitation
0xYY	Receiver Membership Report

8.1. Host Interest Solicitation Packet

A Interest Solicitation packet is periodically multicast by MSNIP capable systems to declare interest in Receiver Membership Reports from multicast routers on the link. The Interest Solicitation message is multicast with a destination address of ALL_IGMPv3_ROUTERS (224.0.0.22) or ALL_MLDv2_ROUTERS (TBA).

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |   Reserved    |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Holdtime            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type

The type field is set to XX (to be assigned by IANA as an IGMP type for IPv4 and an ICMPv6 type for IPv6).

Reserved

Transmitted as zero. Ignored upon receipt.

Checksum

In IPv4, the Checksum is the 16-bit one’s complement of the one’s complement sum of the whole IGMP message (the entire IP payload). In IPv6, the Checksum is the standard ICMPv6 checksum, covering the entire MLDv2 message plus a "pseudo-header" of IPv6 header fields .CITE ICMPv6 . For computing the checksum, the Checksum field is set to zero. When receiving packets, the checksum MUST be verified before processing a packet.

Holdtime
The amount of time a receiving router must keep the system interest state alive, in seconds. The default value for this field is [Interest Solicitation Holdtime].

GenID
Generation ID of the IP system. A number that is selected randomly for each of the [Robustness Variable] initial Interest Solicitation messages when the system comes up and afterwards remains fixed to the value used in the last of the initial messages throughout the system lifetime. The GenID is used by routers to detect system crashes.

8.2. Receiver Membership Report Packet

A Receiver Membership Report packet is unicast by first-hop multicast routers and targeted at potential sources to inform them of the existence or not of receivers for the listed multicast destination addresses.

```
 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
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      Type     |  Dest Count   |           Checksum            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Holdtime            |           Reserved            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Record-Type-1 |            Record-Reserved-1                  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Destination-Address-1                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                               .                               |
|                               .                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Type
The type field is set to YY (to be assigned by IANA as an IGMP type for IPv4 and an ICMPv6 type for IPv6).

Dest Count
The number of multicast destination address records present in this message.

Checksum
In IPv4, the Checksum is the 16-bit one's complement of the one's complement sum of the whole IGMP message (the entire IP payload). In IPv6, the Checksum is the standard ICMPv6 checksum, covering the entire MLDv2 message plus a "pseudo-header" of IPv6 header fields .CITE ICMPv6 . For computing the checksum, the Checksum field is set to zero. When receiving packets, the checksum MUST be verified before processing a packet.

Holdtime
The amount of time in seconds that the target host must keep alive the transmission record state created or updated by the TRANSMIT records in this report. The router originating the

Receiver Membership Report sets this field to the current value of the holdtime timer in the system interest record corresponding to the target host. As a result Receiver Membership Reports sent in response to the reception of a Host Interest Solicitation message have their holdtime set to the value of the holdtime field in the received HIS message.

Reserved
Transmitted as zero. Ignored upon receipt.

Record-Type-1
The type of the first transmission control record in this message. Valid values are:

Record Type	Description	Value
TRANSMIT	Request to start transmitting to destination	1
HOLD	Request to stop transmitting to destination	2

Reserved
Transmitted as zero. Ignored upon receipt.

Destination-Address-1
The multicast destination address of the first record in the message.

8.3. IPv4 Header Fields

Like all other IGMP messages, MSNIP messages are encapsulated in IPv4 datagrams, with an IP protocol number of 2. Every MSNIP message described in this document is sent with an IP Time-to-Live of 1, and carries an IP Router Alert option [RFC-2113] in its IP header.

8.4. IPv6 Header Fields

MLD messages are a sub-protocol of the Internet Control Message Protocol (ICMPv6 [5]). MSNIP messages are identified in IPv6 packets by a preceding Next Header value of 58. All MSNIP messages described in this document are
sent with a link-local IPv6 Source Address (or the unspecified address, if a valid link-local address is not available), an IPv6 Hop Limit of 1, and an IPv6 Router Alert option .CITE RAv6 in a Hop-by-hop Options header.

9. Constants Timers and Default Values

Robustness Variable
The Robustness Variable allows tuning for the expected packet loss on a network. If a network is expected to be lossy, the Robustness Variable may be increased. MSNIP is robust to (Robustness Variable - 1) packet losses. The Robustness Variable MUST NOT be zero, and SHOULD NOT be one. Default: 2

Interest Solicitation Interval
The interval used by MSNIP capable systems between transmissions of Interest Solicitation

messages. Default: 60 secs

Interest Solicitation Holdtime
The interval inserted in Interest Solicitation messages by systems to instruct routers how long they should maintain system interest state for. This MUST be ((the Robustness Variable) times (the Interest Solicitation Interval) plus (one second)).

Initial Interest Solicitation Interval
The interval used by systems to send out the initial Interest Solicitation messages when they first come up. Default: 1 second.

Unsolicited Membership Report Interval
The interval used by routers to send out a set of Membership Report messages when the receiver membership changes for a specific system. Default: 1 second.

10. Possible Optimisations

10.1. Suppressing HIS Messages

A possible optimisation for MSNIP is to suppress the transmission of Host Interest Solicitation messages from the source address of an IP system for which no local application has registered interest. In addition to conserving bandwidth, not transmitting HIS messages prevents remote receivers for groups with no matching source application from creating transmission record state in the host system.

10.2. Host Stack Filtering

Legacy applications that have not been coded with MSNIP support can still be prevented from waisting first-hop link bandwidth by filtering transmitted packets at the operating system level. Even though such applications will not register for MSNIP notifications with the host operating system, if the OS is MSNIP capable and the application is transmitting data to an MSNIP managed group for which there are no transmit records, the OS can safely filter the packets and not transmit them on the wire.

A problem with the filtering approach is that it cannot be combined with the HIS message suppression optimisation (see section 10.1). If there is no registered applications in the system and HIS messages are being suppressed then the first-hop routers will not send any Receiver Membership Reports to the system. As a result knowledge of receiver membership from the presence of transmit records for groups operated by legacy applications will not exist. It therefore becomes unsafe to filter packets from legacy applications.

10.3. Responding to Unexpected IGMP Queries

Under steady state the router side of the IGMP protocol elects a single router on each link that is responsible for issuing IGMP Queries. Routers other than the acting IGMP querier will send an IGMP Query only if they restart and have no IGMP querier election state or if the active Querier crashes and a new election takes place.

MSNIP can take advantage of this mechanism to quickly populate the host interest records of a new router starting up. When the router comes up it will issue an IGMP Query in an attempt to be elected as a Querier. MSNIP capable hosts will notice that the sender of the Query is not the acting Querier. They can use this trigger to respond with Host Interest Solicitation Messages (with transmission randomised over a small interval) to quickly bring the new router up-to-date.

10.4. Host and Router Startup

When a host operating system is restarted there may be applications that are started as part of the initialisation process and want to source IPv4 multicast traffic. It is possible for the applications to register through MSNIP with the IP subsystem and to start transmitting multicast data before the host receives the MSNIP managed range definition through the SSM Range option of the Multicast Router Discovery protocol.

This temporary flooding can be avoided if the host OS holds off notifying MSNIP capable applications that they can transmit until it receives an MRD advertisement and learns the SSM configuration for the network. This behaviour has the drawback that it is not compatible with legacy networks with no MRD deployment. In such a network the host OS has to be able to determine after a configurable period that MRD is not enabled and hence all multicast applications wishing to source traffic should be notified to transmit. A good default value for this period is the MAX_RESPONSE_DELAY of the Multicast Router Discovery protocol [4].

Late router startup is harder to deal with. Hosts that start up before the multicast router may time out waiting for an MRD advertisement and instruct all MSNIP capable multicast source applications to transmit data. One way to work around this problem is to configure the host OS to wait forever for an MRD advertisement before instructing MSNIP applications to transmit.

11. Inter-operation with IGMP / MLD Proxying

MSNIP is intended for use on networks with multicast servers offering a large number of potential sessions. Although unlikely it is possible to deploy such a server behind an IGMP / MLD Proxy [12].

If a device performing IGMP / MLD Proxying wishes to proxy MSNIP, it MUST forward MSNIP Host Interest Solicitation messages that are received on downstream interfaces to its upstream interface. No special treatment is required for MSNIP Receiver Membership Reports as they are unicast to the target host.

In addition to the forwarding of MSNIP messages, an IGMP proxy MUST operate the Multicast Router Discovery protocol [3] on all its downstream interfaces and advertise the MSNIP capability option (section 4.1.1) and SSM address range option (section 4.1.2). The MSNIP capability option should be advertised on downstream interfaces only if it is included in MRD messages received on the upstream interface. The address range to be included in the SSM Range option MUST be determined by MRD and IGMP messages received on the upstream interface of the proxy according to the rules in section 4.2.

In addition to the forwarding of MSNIP messages, an MLD proxy MUST operate the IPv6 Neighbor Discovery protocol. The MSNIP capability option should be advertised on downstream interfaces when it is included in IPv6 Neighbor Discovery messages received on the upstream interface.

12. Security Considerations

We consider the ramifications of a forged message of each type. As described in [1] IPSEC AH can be used to authenticate IGMP messages if desired.

12.1. Receiver Membership Report attacks

A DoS attack on a host could be staged through forged Receiver Membership Report messages. The attacker can send a large number of reports, each with a large number of TRANSMIT records and a holdtime field set to a large value. The host will have to store and maintain the transmission records specified in all of those reports for the duration of the holdtime. This would consume both memory and CPU cycles in the host.

Forged Receiver Membership Report messages from the local network can be easily traced. There are three measures necessary to defend against externally forged reports:

- Routers SHOULD NOT forward Receiver Membership Reports. This is easier for a router to accomplish if the report carries the Router-Alert option.

- Hosts SHOULD ignore Receiver Membership Reports without the Router-Alert option.

Note that a remote attack through the multicast routing protocol is possible. A remote site can originate join state for a large number of groups that will propagate through MSNIP to the target source host. Such attacks are considered a more significant problem for the routers involved and are left up to the routing protocol security.

HOLD records in forged Receiver Membership Report messages are not a significant threat as hosts track the individual interests of each first-hop router separately. Only by forging the source address of the report message so that is appears to have originated from a real first-hop router can the attacker cause the source to stop transmitting to a group that has valid receivers. Such forged messages can be detected by the router itself.

12.2. Host Interest Solicitation attacks

Forged Host Interest Solicitation messages can have two effects:

- When non-existent source addresses are used the solicitation messages can create unwanted host record state on attached routers for the duration of the holdtime specified in the message.

- When a source address corresponding to an existing host is used in the forged HIS message, receipt of the message by attached routers will cause them to transmit Receiver Membership Reports messages for any multicast destination addresses with receivers for the target host. Although no additional state will be created in routers or hosts from this attack, bandwidth and

CPU is wasted in both the first-hop routers and the target host.

Just like for the Receiver Membership Report message, attacks using the Host Interest Solicitation message can be reduced by requiring the use of the Router-Alert option on the message.

12.3. MSNIP Managed Range Discovery

As discussed in [4] it is possible for directly connected systems to send forged Multicast Router Advertisement messages containing the SSM Range Discovery option. As the SSM Range Discovery option determines the MSNIP managed range under IPv4, such forged messages can temporarily replace the managed range map with incorrect information in receiving hosts. An incorrect mapping can have two effects:

- Applications using a multicast destination address within the real SSM range that have no valid receivers can be tricked into thinking that their chosen destination address is no longer an SSM address and will therefore start transmitting data.

- Applications using group addresses outside the valid SSM range can be tricked into thinking that they are using an SSM destination address and therefore prevented from transmitting data.

The Multicast Router Discovery SSM Range Option specification suggests that a router receiving a Multicast Router Advertisement with an inconsistent SSM Range Option log the event to the operator. Such logging will enable tracking of this type of attack.

13. IANA Considerations

This document introduces the following new types and options that require allocation by IANA:

- Two new IGMP messages for Host Interest Solicitation and Receiver Membership Report. Each of these messages requires a new IGMP type value to be assigned by IANA [11].
- The new MSNIP Operation option for the Multicast Router Discovery protocol. This option requires a new MRD type value to be assigned by IANA.

- The new MSNIP Operation option for the Neighbour Discovery / ICMPv6 protocol. This option requires a new NDP / ICMPv6 type value to be assigned by IANA.

14. Acknowledgments

The authors would like to thank Dave Thaler, Jon Crowcroft, Toerless Eckert and Haixiang He for their contribution to this specification.

15. Authors' Addresses

Bill Fenner
AT&T Labs - Research
75 Willow Road
Menlo Park, CA 94025
fenner@research.att.com

Brian Haberman
Caspian Networks
One Park Drive, Suite 400
Research Triangle Park, NC 27709
bkhabs@nc.rr.com

Hugh Holbrook
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
holbrook@cisco.com

Isidor Kouvelas
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95134
kouvelas@cisco.com

16. Normative References

[1] B. Cain, S Deering, W. Fenner, I Kouvelas, A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376.

[2] R. Vida, et al, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", work in progress, <draft-vida-mld-v2-??.txt>, October 2002.

[3] S. Biswas, B. Haberman, "IGMP Multicast Router Discovery", Work In Progress, <draft-ietf-idmr-igmp-mrdisc-08.txt>, 2001.

[4] I. Kouvelas, "Multicast Router Discovery SSM Range Option", work in progress, <draft-ietf-magma-mrdssm-02.txt>, November 2002.

[5] A. Conta, S. Deering, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6)", RFC 2463.

[6] H. Holbrook, B. Cain, "Source-Specific Multicast for IP", work in progress, <draft-ietf-ssm-arch-00.txt>, 21 November 2001.

[7] S. Kent, R. Atkinson, "Security Architecture for the Internet Protocol.", RFC 2401.

[8] C. Partridge, A. Jackson, "IPv6 Router Alert Option", RFC 2711.

17. Informative References

[9] W. Fenner, M. Handley, H. Holbrook, I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", Work In Progress, <draft-ietf-pim-sm-v2-new-??.txt>, 2002.

[10] Z. Albanna, K. Almeroth, D. Meyer, "IANA Guidelines for IPv4 Multicast Address Allocation", Best Current Practices, <draft-ietf-iana-IPv4-mcast-guidelines-00.txt>, 2001.

[11] W. Fenner, "IANA Considerations for IGMP", http://www.iana.org/assignments/igmp-type-numbers, RFC 3228 (BCP 57), February 2002.

[12] W. Fenner, H. He, B. Haberman, H. Sandick, "IGMP / MLD-based Multicast Forwarding (IGMP / MLD Proxying)" draft-ietf-magma-igmp-proxy-01.txt, November, 2002.