

Multicast Signaling Conduit Protocol

[draft-keyur-mscp-00.txt](#)

1. Status of this Memo

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/lid-abstracts.txt>

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

2. Abstract

This document describes an approach for solving the "last-mile" problem for multicast, especially SSM-style multicast, called Multicast Signaling Conduit Protocol (MSCP). It is intended to facilitate an endnode whose OS does not support IGMPv3, or who resides in a portion of the Internet infrastructure without multicast, to join multicast groups. It is especially intended for SSM-style groups, although it could be adapted for use for other multicast groups.

[3.](#) Introduction

An impediment to getting SSM Multicast deployed is that, as currently specified, it depends on IGMPv3. Since IGMPv3 is in the IP stack, it requires endnodes to upgrade to an OS that supports IGMPv3 and it requires IGMPv3 support on the first hop multicast capable routers as well. This document proposes a new IP protocol known as Multicast Signaling Conduit Protocol (MSCP) which eliminates the requirement of running IGMPv3 both on endnodes as well as first hop multicast capable routers. This draft is solely for the purpose of joining SSM-style groups, i.e., groups in which the IP address of the root is explicitly known by the joining host. This draft allows hosts to join SSM-style groups even if

- * the host's OS stack does not support IGMPv3
- * the local routers do not support IGMPv3, or even multicast, and/or
- * some routers along the path to the root do not support multicast.

This is achieved by defining a new user level process that runs MSCP. All that is really required to join an SSM-style group with this mechanism is for the joining host and the root (the "Source") to support this. As more routers are upgraded to supporting this, bandwidth use is minimized by utilizing multicast rather than tunnels.

Although in theory this style of host join could be used instead of IGMP to join a non-SSM group, for non-SSM groups this design offers no advantage over IGMP, since only IGMPv2, which is already widely deployed, is required for joining non-SSM groups.

However, if an endnode does reside in a portion of the Internet without any support for multicast, this mechanism could be used for non-SSM groups by using an anycast address in place of an explicit root address.

A router that does not support the messages in this draft will simply forward such messages towards the destination specified in the messages. The first multicast capable router that supports MSCP will notice these messages and will form a tunnel with the end node that generated the message.

We propose a new IP multicast signaling protocol known as Multicast Signaling Conduit protocol (MSCP). This protocol has an ability to form multicast tunnels with specified destination address in the IP header. It runs directly over IP protocol.

We also propose new MSCP messages with destination address in the IP header as either "Source" (root of the tree) address or anycast address to direct all the MSCP packets to a specific place within an AS. A wellknown anycast address defined by IANA will be used to transmit MSCP packets to a specified destination within an AS.

Multicast routers implementing MSCP should treat the receipt of MSCP messages as intended for them, even though the destination address in the IP header is a unicast of the tree root. To allow a router to recognize a MSCP messages addressed to the unicast root address, and place it on the slow path, rather than simply forwarding it towards the specified destination, the packet will have the router alert option, and the IP protocol type="MSCP" with version 1.

Usually a multicast router implementing MSCP that notices an MSCP message would trap the message and create a tunnel to the node that sent the MSCP Host-Join. However, in the case where the IP destination address is "MSCP Anycast", the router should forward the packet towards a router that advertises reachability to that address.

Multicast routers with MSCP support receiving MSCP Join/Prune [5] messages will process them in similar way as PIM Join/Prune messages. However, MSCP routers receiving MSCP Join/Prune will send MSCP (Join/Prune)-Ack message back to the sender of MSCP Join/Prune messages with a 4 byte cookie value. Receivers upon receipt of MSCP (Join/Prune)-Ack message will respond back with MSCP Auth-Ack message in which it will re-send the cookie value sent by the router. MSCP routers receiving Auth-Ack for Join/Prune will then create/delete interface state accordingly.

MSCP routers will create tunnel interfaces (refer [section 6](#)) in their (S, G) oif lists when receiving MSCP Join [5] messages from receivers that are not directly connected. For all the directly connected receivers, MSCP routers will create interfaces in their (S, G) oif lists in the same way as if an IGMP (S,G) join or a PIM Join/Prune message were received. An implementation must have a configured upper bound on number of tunnel interfaces that it can put in oif lists of any multicast route. Multicast routers receiving such MSCP Host-Join/Prune messages will process them and create state in similar way as if a PIM Join/Prune (S, G) message was received.

All the non-multicast capable routers will simply forward such packets towards the unicast source address/anycast address. Current routers which receive an IP unicast packet with IP options process these packets and forward them towards the unicast destination address specified in the

Set to zero on transmission. Ignored upon receipt.

Checksum

The checksum is a standard IP checksum, i.e. the 16-bit one's complement of the one's complement sum of the entire MSCP message. For computing the checksum, the checksum field is zeroed.

For IPv6, the checksum also includes the IPv6 "pseudo-header", as specified in [RFC 2460, section 8.1](#) [5]. This "pseudo-header" is prepended to the MSCP header for the purposes of calculating the checksum. The "Upper-Layer Packet Length" in the pseudo-header is set to the length of the MSCP message. The Next Header value used in the pseudo-header is 103. If the packet's length is not an integral number of 16-bit words, the packet is padded with a byte of zero before performing the checksum.

Encoded format for Unicast Source and Group address is shown below:

```

      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
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Addr Family | Encoding Type |      Unicast Address      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+...

```

Addr Family

The MSCP address family of the 'Unicast Address' field of this address.

Values of 0-127 are as assigned by the IANA for Internet Address Families in [8]. Values 128-250 are reserved to be assigned by the IANA for PIM-specific Address Families. Values 251 though 255 are designated for private use. As there is no assignment authority for this space, collisions should be expected.

Encoding Type

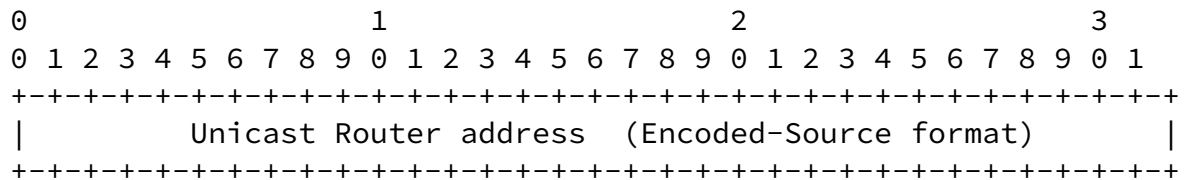
The type of encoding used within a specific Address Family. The value '0' is reserved for this field, and represents the native encoding of the Address Family.

Unicast Address

The unicast address as represented by the given Address Family and Encoding Type.

5.1 MSCP Host Query message format

MSCP Prune-Ack messages are sent in reply to Prune message by routers forming tunnels.

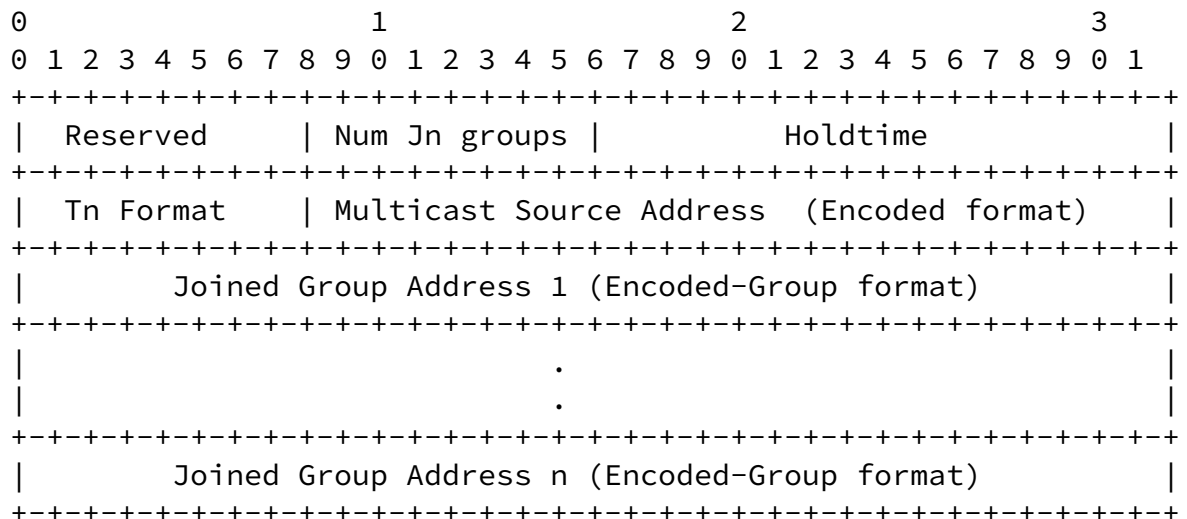


Unicast Router Address

For format description refer to Encoded-Unicast address.

5.2 MSCP Host-Join message format

MSCP Host-Joins are sent to assist building source/core trees (SPT/CBT).



Reserved

Transmitted as zero, ignored on receipt.

Number of Join Groups

The number of multicast group sets contained in the message.

Holdtime

The amount of time a receiver must keep the Join state alive, in seconds. If the Holdtime is set to `'0xffff'`, the receiver of this message should hold the state until canceled by the appropriate canceling Join message, or timed out according to

local policy. This may be used with dial-on-demand links, to avoid keeping the link up with periodic Join/Prune messages.

Note that the Holdtime must be larger than the J/P_Override_Interval(I) (refer to [PIM-SM]).

Tunnel Type

Specific type of Tunnel that receiver wants Router to create
0 (IP-in-IP), 1 (GRE).

Multicast Source address

For format description refer to Encoded-Unicast address.

Join Group Address 1 .. n

This list contains the Groups that the sending router will forward multicast datagrams for if received on the interface this message is sent on.

See Encoded-Source-Address format.

[5.3](#) MSCP Host Join-Ack message format

MSCP Join-Ack messages are sent in reply to Join message by routers forming tunnels.

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			
++	++	++	++
Reserved	Num Pr groups	Holdtime	
++	++	++	++
Tn Format	Multicast Source Address (Encoded format)		
++	++		++
Joined Group Address 1 (Encoded-Group format)			
++			++
.			
.			
++			++
Joined Group Address n (Encoded-Group format)			
++			++
Cookie Value			
++			++

Reserved

Transmitted as zero, ignored on receipt.

Number of Prune Groups

The number of multicast group sets contained in the message.

Holdtime

The amount of time a receiver must keep the Join state alive, in seconds. If the Holdtime is set to '0xffff', the receiver of this message should hold the state until canceled by the appropriate canceling Join message, or timed out according to local policy. This may be used with dial-on-demand links, to avoid keeping the link up with periodic Join/Prune messages.

Note that the Holdtime must be larger than the J/P_Override_Interval(I) (refer to [PIM-SM]).

Tunnel Type

Specific type of Tunnel that receiver wants Router to create
0 (IP-in-IP), 1 (GRE).

Multicast Source address

For format description refer to Encoded-Unicast address.

Join Group Address 1 .. n

This list contains the Groups that the sending router will forward multicast datagrams for if received on the interface this message is sent on.

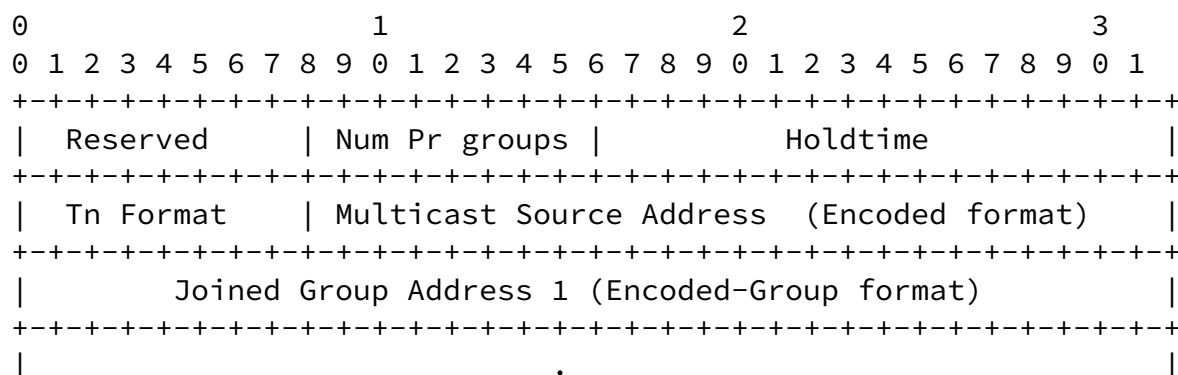
See Encoded-Source-Address format.

Fixed four byte Key

Cookie value that is reflected back in Auth-Ack message.

[5.4](#) MSCP Host-Prune message format

MSCP Host-Prunes are sent to prune source trees when members leave groups.



```

| . |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           Joined Group Address n (Encoded-Group format)           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Reserved

Transmitted as zero, ignored on receipt.

Number of Prune Groups

The number of multicast group sets contained in the message.

Holdtime

The amount of time a receiver must keep the Join state alive, in seconds. If the Holdtime is set to '0xffff', the receiver of this message should hold the state until canceled by the appropriate canceling Join message, or timed out according to local policy. This may be used with dial-on-demand links, to avoid keeping the link up with periodic Join/Prune messages.

Note that the Holdtime must be larger than the J/P_Override_Interval(I) (refer to [[PIM-SM](#)]).

Tunnel Type

Specific type of Tunnel that receiver is using with the Router
0 (IP-in-IP), 1 (GRE).

Multicast Source address

For format description refer to Encoded-Unicast address.

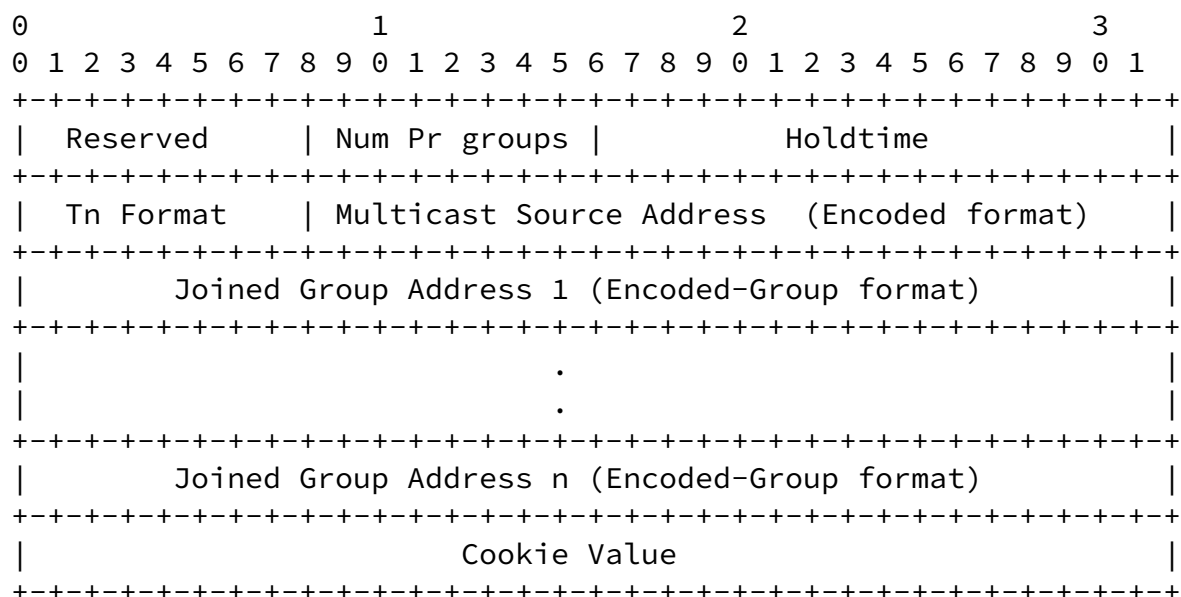
Join Group Address 1 .. n

This list contains the Groups that the sending router will forward multicast datagrams for if received on the interface this message is sent on.

See Encoded-Source-Address format.

[5.5](#) MSCP Host Prune-Ack message format

MSCP Host-Prunes are sent to prune source trees when members leave groups.



Reserved

Transmitted as zero, ignored on receipt.

Number of Prune Groups

The number of multicast group sets contained in the message.

Holdtime

The amount of time a receiver must keep the Join state alive, in seconds. If the Holdtime is set to '0xffff', the receiver of this message should hold the state until canceled by the appropriate canceling Join message, or timed out according to local policy. This may be used with dial-on-demand links, to avoid keeping the link up with periodic Join/Prune messages.

Note that the Holdtime must be larger than the J/P_Override_Interval(I) (refer to [\[PIM-SM\]](#)).

Tunnel Type

Specific type of Tunnel that receiver is using with the Router
0 (IP-in-IP), 1 (GRE).

Multicast Source address

For format description refer to Encoded-Unicast address.

Join Group Address 1 .. n

This list contains the Groups that the sending router will forward multicast datagrams for if received on the interface this message is sent on.

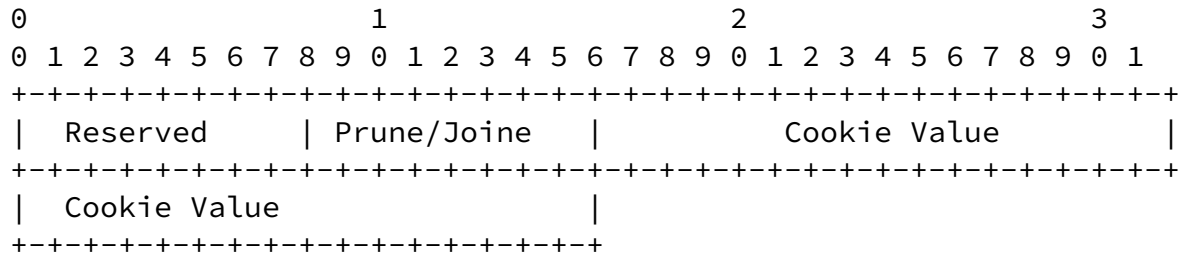
See Encoded-Source-Address format.

Fixed four byte Key

Cookie value that is reflected back in Auth-Ack message.

5.6 MSCP Host Auth-Ack message format

MSCP Prune-Ack messages are sent in reply to Prune message by routers forming tunnels.



Reserved

Transmitted as zero, ignored on receipt.

Prune/Join

Auth message sent in response to 0 (Join-Ack) 1 (Prune-Ack).

Fixed four byte Key

Cookie value that is reflected back in Auth-Ack message.

6. Tunnel Interfaces

Because not all intermediate routers support native ip multicast, MSCP requires all the oifs on which it receives MSCP Host-Join/Prunes messages from receivers which are not directly connected to be created as tunnel interfaces. In practice, tunnels typically use either IP-IP [Perk96] or Generic Routing Encapsulation (GRE) [Han94a,Han94b], although, other encapsulation methods are acceptable.

7. Security Considerations

MSCP does not change any multicast security issues. PIM is open to many types of resource overload. For instance, a node which transmits from many bogus source addresses will cause PIM routers to join to many (S,G) pairs, eventually exhausting state. PIM routers must protect themselves by limiting the amount of state for multicast, so that a denial of service attack on multicast will not destroy unicast.

With LAN-based IGMP joins, only nodes on the LAN can initiate a join, so if the LAN is physically protected, and all routers along the path to the RP (or S in an (S,G) join) are trusted, it might be somewhat harder for a node to create join state untraceably than it would with this tunnel proposal. With this tunnel proposal, the join message is sent over a multi-hop path.

As stated in the document, a router is configured with a maximum number of tunnels it is willing to accept, so an attack to deplete its tunnel resources will only affect attempts to create tunnels, and not cause any other denial of service.

Additionally, the router to whom tunnels will be created is likely to be the entry router at an ISP, and it can consider an entire customer network to be a LAN. If suspiciously many tunnels are being created from that customer network, the router can start discriminating against joins from that customer network, when resources are being depleted.

Eventually, cryptographic authentication can be added between the joiner and the router, by having potential joiners obtain a key (or certificate) from the router before they are allowed to join, and having join messages cryptographically authenticated.

8. Acknowledgements

We express our thanks to Alex Zinin, Dino Farinacci, Lorenzo Vicisano, Liming Wei, Tom Pusateri, and Nidhi Bhaskar for their review and comments.

9. References

- [PIM-SM] Bill Fenner, Mark Handley, Hugh Holbrook, Isidor Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)".
- [Han94a] Hanks, S., Li, T, Farinacci, D., and P. Traina, "Generic Routing Encapsulation", [RFC 1701](#), NetSmiths, Ltd., and cisco Systems, October 1994.
- [Han94b] Hanks, S., Li, T., Farinacci, D., and P. Traina, "Generic Routing Encapsulation over IPv4 networks", [RFC 1702](#), NetSmiths, Ltd., cisco Systems, October 1994.
- [Perk96] Perkins, C., "IP Encapsulation within IP", [RFC 2003](#), October 1996.

10. Author Information

Keyur Patel
Cisco Systems. Inc.
email: keyur@ayrnetworks.com

Radia Perlman
Sun Microsystems. Inc.
email: Radia.Pperlman@sun.com

Dino Farinacci
Procket Networks Inc.
email: dino@procket.com

Greg Shepherd
Procket Networks Inc.
email: shep@procket.com