Internet Engineering Task Force
INTERNET-DRAFT
draft-ietf-msdp-traceroute-01.ps

W. Fenner
AT&T Labs - Research
D. Meyer
cisco Systems
R. Roberts
Stanford University

March, 2001
Expires September 2001

MSDP Traceroute

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. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a "working draft" or "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`.

Distribution of this document is unlimited.

Abstract

In order to diagnose problems with the Peer-RPF-Flooding mechanisms in the Multicast Source Discovery Protocol [MSDP], we introduce a method of tracing the path that an SA message should have taken, along with collecting statistics along that path. This occurs in a way similar to multicast traceroute [MTRACE], with each router appending information about this hop and forwarding the message towards the desired RP.

This document is a product of the MSDP working group within the Internet Engineering Task Force. Comments are solicited and should be addressed to the working group's mailing list at msdp@network-services.uoregon.edu and/or the authors.

1. Introduction

One of the key problems in the deployment of MSDP [MSDP] infrastructures has been the inability to trace the path taken by MSDP control packets. The protocol specified in the document addresses this problem by providing a traceroute facility for MSDP Source Active (SA) Messages. It is important to note that, with the exception of data encapsulated in SA messages, MSDP traceroute traces a path through the MSDP control plane. It is important to note that the MSDP control path need not be the same as the path followed by data packets.

As is the case for multicast traceroute, the goals of MSDP traceroute include

- o To be able to trace the path that a SA message would take from some source router (RP) to some destination router (RP).
- o To be able to isolate SA Message loss problems
- o To be able to isolate configuration problems (e.g., TTL threshold).
- o To minimize packets sent (e.g. no flooding, no implosion).

1.1. Definitions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Originating an MSDP Traceroute Query

As is the case for other kinds of traceroute, both routers and hosts will want to originate MSDP traceroute queries.

2.1. Routers Originating an MSDP Traceroute Query

When a router originates an MSDP traceroute Query, it formulates the Query by filling its first block and unicasts the Query packet its RPF neighbor toward the specified RP. Note that this is unlike multicast traceroute, in which the Query is sent to the last-hop multicast router for the destination, converted into a Request, and fowarded back towards the source and group.

2.2. Hosts Originating an MSDP Traceroute Query

When a host originates an MSDP traceroute Query, it formulates the Query by filling its first block with the RP (and optionally the source and group), and then unicasts the UDP encapsulated Query packet to the specified RP [means UDP encapsulation has to be supported]. The RP must store state that describes the host that made the query [security stuff here]. If a router receives a Query packet and is not running MSDP, the router discards the packet and returns a response with error code TBD [do we want a notification for this in the MSDP spec?]

3. Packet Formats

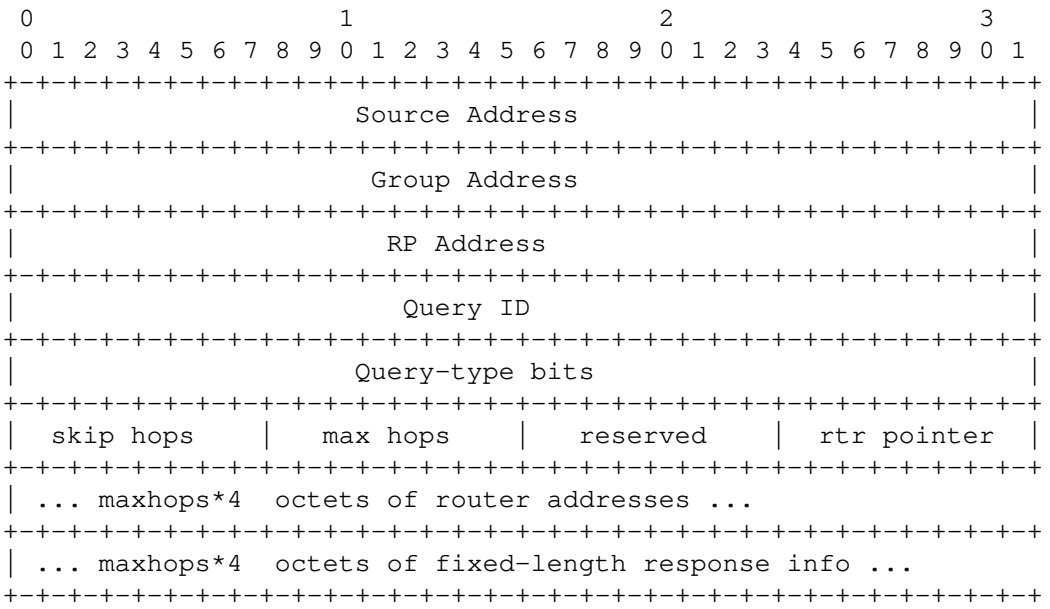
Encapsulated in 2 different MSDP TLVs:

In-progress MSDP traceroute: ...

MSDP traceroute answer: ...

Fixed request header contains S, G, RP, query ID, query-type bits, and the router pointer (for returning packet HBH). Following are two tables, one for router addresses and the other for fixed-length response info filled in at each hop.

The remainder of the packet is available to collect more detailed response info as specified by the query-type bit vector.



3.1. Source Address

Source in (S,G,RP)

3.2. Group Address

Group in (S,G,RP)

3.3. RP Address

RP in (S,G,RP)

3.4. Query ID

This field is used as a unique identifier for this traceroute request so that duplicate or delayed responses may be detected.

3.5. Query-type bits

Bit vector apecifying collection of more detailed response data to be collected for the hops between skip hops and max hops or until the packet size is exceeeded.

3.6. skip hops

The number of hops to be skipped in this request before beginning to collect detailed response data. This allows for the data collection specified by the query-type bit vector to begin at this point in the path. Specified by the requestor.

3.7. max hops

The max. # of hops requested in this traceroute request.

3.8. reserved

This octet is reserved for future use.

3.9. rtr pointer

The rtr pointer is a pointer into both the router addresses table and the fixed-length response info table. When the query is being filled in (outbound) the rtr pointer points to the next available entry. On return to the query originator, it points to the next hop back (toward the originator).

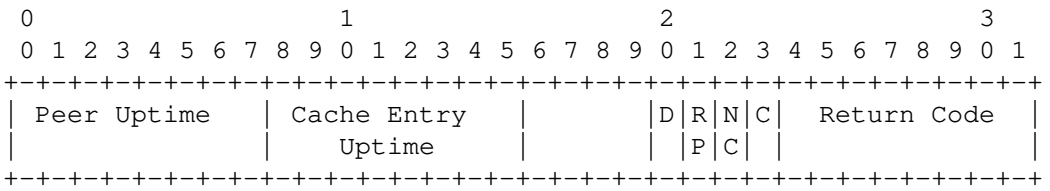
3.10. router addresses

Table of the outgoing interfaces that this request has traversed. Inserted by requestor, max hops * 4 octets.

4. fixed-length response info

Table of fixed response words containing limited status info for each router along the path. Inserted by requestor, max hops * 4 octets.

Fixed response word:



Peer uptime: RPF peer uptime in minutes, capped at 255.

Cache Entry Uptime: cache entry uptime in minutes, capped at 255.

Bits:

D-bit: set if have this (S,G) in cache but with a different RP.

RP-bit: set if this router is an RP

NC-bit: set if this router is not caching SA's

C-bit: set if this (S,G,RP) tuple is in the cache.

Return Code:

0 No-error

1 Hit-src-RP

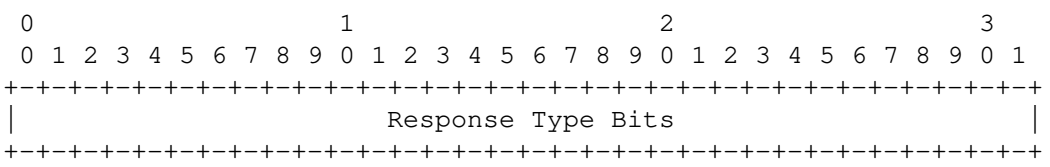
2 Filled-packet

3 No-RPF-neighbor

4 Reached-max-hops

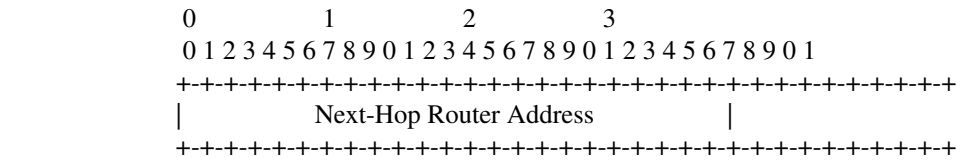
4.1. Detailed response block

Each detailed response block begins with a bit vector describing what types of responses are present (equal to or a sub-set of the query-type bits):

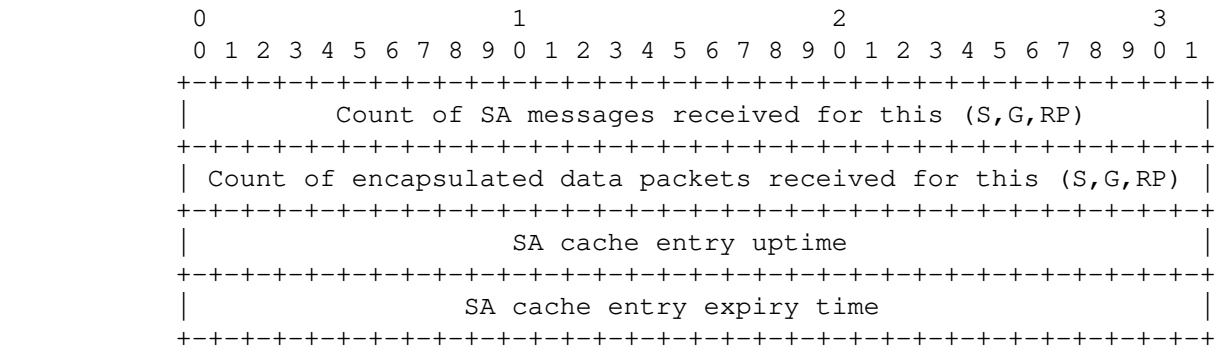


Types of info:

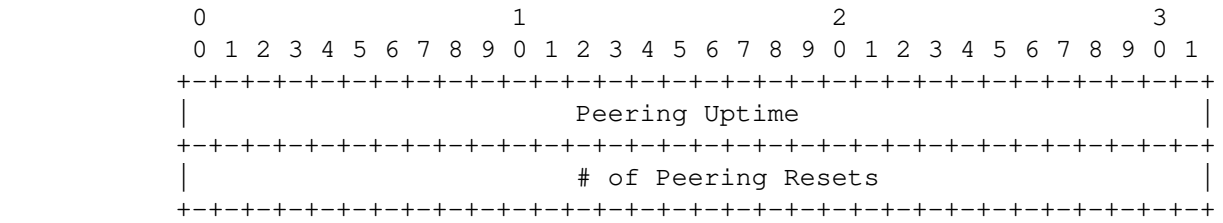
Bit 0: RPF Neighbor info



Bit 1: SA info



Bit 2: Peer info



5. Protocol Processing

5.1. Processing an in-progress traceroute packet

The following steps are taken when receiving an in-progress traceroute:

1. If rtr pointer is greater than max hops, something bad has happened; immediately process as a partial response.
2. If this router is the source RP, set the Hit-src-RP return code and the RP-bit in the fixed-length response info word and store in the fixed-length response info table as offset by rtr pointer. Proceed to Return packet as a response.
3. If this router is an RP, set the RP-bit in the fixed-length response info word. If this router is not doing SA caching, set the NC-bit in the fixed-length response info word and proceed to the RPF peer check. Otherwise look up the SA cache information for the (S,G,RP). If the SA cache entry exists, set the C-bit in the fixed-length response info word and insert the Cache Entry Uptime field in the fixed-length response info word. If no SA cache entry exists for the (S,G,RP), check if there is a cache entry for (S,G) with a different RP. If so, then set the D-bit in the fixed-length response info word.
4. Look up the MSDP RPF neighbor for the RP, using the MSDP peer-RPF-forwarding rules in the MSDP spec [MSDP].
5. If this lookup returns a peering, insert your IP address on this peering into the router address table as offset by the current rtr pointer. Insert the Peer Uptime field in the fixed-length response info word and move it to the fixed-length response info table as offset by the current rtr pointer. If all the query-type bits are zero, proceed to Forward packet to RPF neighbor. Otherwise proceed to Append detailed response data.

6. If the RPF neighbor lookup failed, set the No-RPF-neighbor return code in the fixed-length response info word and move it to the fixed-length response info table as offset by the current rtr pointer and proceed to Return packet as a response.

5.1.1. Append detailed response data

The detailed response data, as specified by the query-type bits, is appended to the received packet.

Until rtr pointer is greater than skip hops, do not append detailed data. If there is a non-zero return code in the fixed-length response info word, proceed to Return packet as a response otherwise proceed to Forward packet to RPF neighbor.

Verify that space remains in the packet to add the detailed response data block(s) as specfied by the query-type bits. If adding these blocks would cause this packet to exceed the MSDP MTU (1400 bytes), set the Filled-packet return code in the fixed-length response info word and move it to the fixed-length response info table as offset by the current rtr pointer. Proceed to Return packet as a respose.

Copy the query-type bits to the response type bits word, first of the detailed response data.

If the Next Hop info query-type bit is set, insert the RPF peer's address into the Next-Hop Router Address word.

If the SA info query-type bit is set, fill in the SA-info query words. If no matching SA entry or the router is not caching SAs, fill words with zeros.,

If the Peer info query-type bit is set, fill in the Peer-info query words.

5.1.2. Forward packet to RPF neighbor

Increment the rtr pointer field.

If the rtr pointer == max hops, decrement the rtr pointer field, set the Reached-max-hops return code in the fixed-length response info word and move it to the fixed-length response info table as offset by the current rtr pointer. Proceed to Return packet as a response.

Otherwise, unicast the current packet over the MSDP peering to the RPF neighbor.

5.2. Return packet as a response

Turn the MSDP type from traceroute request to traceroute response.

Decrement the router pointer.

If the router pointer was already 0, you're the one who requested the trace.

Otherwise, send the message on your MSDP peering with the router at the current router pointer's offset in the router address table.

5.3. Processing a traceroute answer

Decrement the router pointer.

If the router pointer was already 0, you're the one who requested the trace.

Otherwise, send the message on your MSDP peering with the router at the current router pointer's offset in the router address table.

6. Security Considerations

(flesh these out)

7. References

RFC2119 Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", RFC 2119/BCP 14, Harvard University, March 1997.

MSDP Farinacci, D., et. al. "Multicast Source Discovery Protocol (MSDP)", draft-ietf-msdp-spec-05.txt, February, 2000.

MTRACE Fenner, W. and S. Casner, "A "traceroute" facility for IP Multicast", draft-ietf-idmr-traceroute-ipm-06.txt, March, 2000.

8. Author's Addresses

William C. Fenner
AT&T Labs - Research
75 Willow Rd
Menlo Park, CA 94025
Phone: +1 650 330 7893
Email: fenner@research.att.com

David Meyer
Cisco Systems
170 Tasman Drive
San Jose, CA, 95134
Phone: +1 541 915 0094
Email: dmm@cisco.com

Ronald G. Roberts
Stanford University
241 Panama St
Pine Hall Room 175
Stanford, CA 94305-4122
Phone: +1 650 723 3352
Email: rgr@stanford.edu