Internet Engineering Task Force Inter-Domain Multicast Routing WG
INTERNET-DRAFT W. Fenner
draft-ietf-idmr-membership-reports-00.txt November 12, 1997
Expires May 1998

Domain Wide Multicast Group Membership Reports

Status of this Memo

Abstract

When running a multi-level multicast routing protocol, upper layers need to know about group memberships in lower layers in a protocol-independent fashion. Domain Wide Multicast Group Membership Reports allow this information to be learned in a fashion similar to IGMP[Fenn97] at the domain level.

This document is a product of the IDMR working group within the Internet Engineering Task Force. Comments are solicited and should be addressed to the working group's mailing list at idmr@cs.ucl.ac.uk and/or the author.

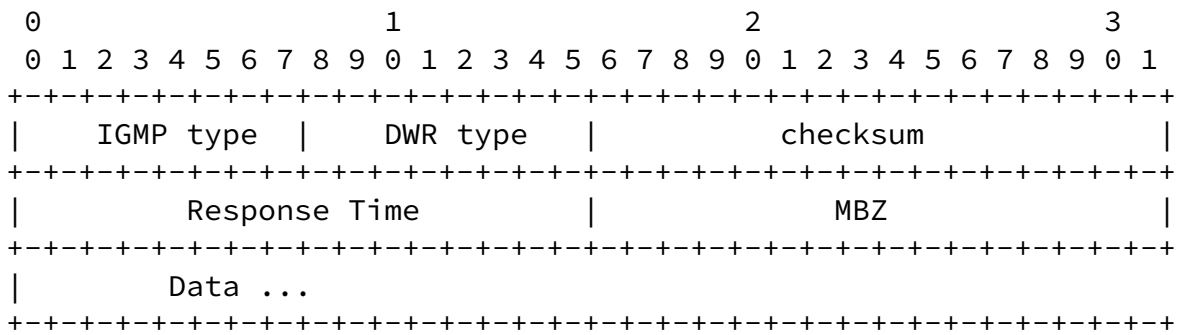
1. Introduction

Domain-Wide Multicast Group Membership Reports (DWRs) are a group membership protocol at the domain level. When using a heirarchical multicast routing protcol [Thya94],[Thal97], the inter-domain protocol needs to learn of group memberships inside domains. Although some intra-domain routing protocols can provide this information easily to the domain border routers, some cannot. DWRs specify a protocol-independent manner to learn group membership inside a domain.

This document specifies the DWR protocol, as used by border routers and interior routers. It specifies a behavior that can be used with any intra-domain protocol, along with optimizations for certain intra-domain protocols, and a transition scheme so that all interior routers need not be updated.

2. Packet Format

DWR packets are sent as IGMP packets (IP protocol #2) when using IPv4. It's not yet clear what IPv6 type will be used, or if they should have their own IP protocol number which can be the same for v4 and v6.



2.1. IGMP Type: 8 bits

The IGMP type field is defined to be 0xXX.

2.2. DWR Type: 8 bits

The following DWR types are defined:

Type	Name
0x00	Domain-Wide Query
0x01	Domain-Wide Membership Report
0x02	Domain-Wide Leave
0x03	Non-authoritative Domain-Wide Leave*

---
  * Note that the requirement for a Non-authoritative

2.3. checksum

IP-style checksum over the whole packet, zeroed for computing checksum. This checksum MUST be computed when transmitting packets, and MUST be verified when receiving.

2.4. Response Time

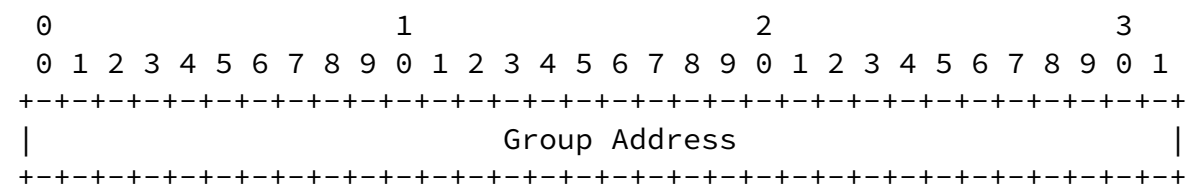
The time, in units of 100ms, allowed for responses. Varying this field allows tuning the burstiness of DWR traffic at the cost of higher latencies.

2.5. MBZ

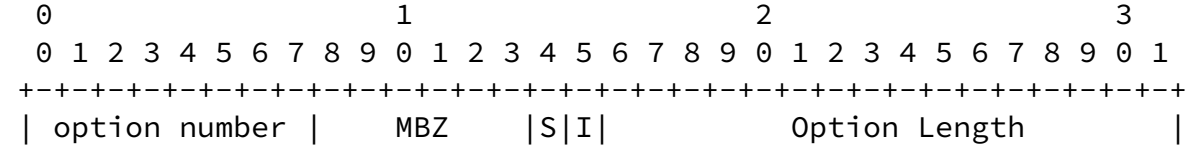
This field must be zeroed on transmission and ignored on reception.

2.6. Data

This field consists of the rest of the packet. It may consist of one of two forms; a Group Address:



>>>Include mask with group??? TBD<<< or an option header:



```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Option Data...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The two forms may be told apart because option numbers are assigned in the range [0,223], the first byte of an IPv4 group address is in the range [224,239], and the first byte of an IPv6 group address is 255.

DWL has been questioned.

2.6.1. Data Description

2.6.1.1. Group Address

The group address being reported or queried.

2.6.1.2. Option Number

The number of the option.

2.6.1.3. MBZ

Must be zeroed on transmission and ignored on reception.

2.6.1.4. I

Ignore this entry for group membership purposes if the option is not recognized.

2.6.1.5. S

Ignore this entry for report suppression purposes (on Replies or Leaves) or for reply purposes (on Queries) if the option is not recognized.

2.6.1.6. Option Length

The number of words, excluding the initial word, of option data following the option header.

2.6.1.7. Option Data

Option Length words of data.

2.6.2. Option Processing

Options which precede any group addresses are called Global Options. Options which follow a group address are associated with that group address and are called Group Options. Groups with options attached should be ignored as if they were not present if there are unrecognized Group Options with the S bit set. Packets with Global Options should be ignored as if they were not received if there are unrecognized Global Options with the S bit set.

2.6.3. Defined Options

One envisioned use for options is to expand DWR's to include

IGMPv3 source-specific information, using the S and I bits to provide full backwards compatibility.

2.6.3.1. Unicast-reply

This option has the S bit turned off (e.g. may be ignored if not understood) [but it's in the original spec so should be implemented]. If a query is received with this option, the reply should be unicasted to the source of the query. If the option carries a unicast address, it is the address to be unicasted to.

If a unicast address is specified in the option, the option MUST be ignored if the receiver does not have an IPSEC[Atki95] relationship with the source of the packet.

3. Message Descriptions

3.1. Domain-Wide Query

A Domain-Wide Query is sent periodically by one of the domain's border routers. >>>Querier Election TBD<<<. The default period is >>>TBD<<<, and MUST be configurable. Domain-Wide Query messages are sent to the well-known Domain-Wide Query multicast group. This group is in the range of addresses that are scoped to a single domain, which has not yet been allocated by the IANA.

A Domain-Wide Query with no additional data is a request for knowledge of all multicast group memberships in the domain. The Domain-Wide Query may be restricted by including groups or options in the data portion of the packet. If group addresses or DWR options are specified in the packet, the Query is restricted to those groups or other options as specified in the packet.

3.2. Domain-Wide Report

A Domain-Wide Report is sent by a router in response to a Domain-Wide Query message, or in response to the appearance of a new group member in the domain. The latter messages are called "Unsolicited" Domain-Wide Reports.

A Domain-Wide Report message includes a list of group memberships and other options in the additional data portion of the packet.

3.3. Domain-Wide Leave

A Domain-Wide Leave is sent by a router when it knows that there are no more members in the domain of a group or groups. The group

or groups listed in the additional data portion of the packet are considered by the border routers to have no more members.

3.4. Non-Authoritative Domain-Wide Leave

There are tradeoffs to providing a Non-Authoritative Domain-Wide Leave message. A Non-Authoritative Domain-Wide Leave is sent by a router when it has no more members of the group but cannot be sure there are no more members in the domain, and causes the elected border router to send a Domain-Wide Query message for the group(s) mentioned in the Non-Authoritative Domain-Wide Leave message. This adds to the time before the border routers can signal that there

are no group members inside the domain. With DVMRP and PIM-DM (the two current multicast routing protocols that would require a Non-Authoritative Domain-Wide Leave), it is possible for the border routers to notice that all sources for the group have been pruned to all border routers, and determine that there are no internal members in that way. However, if there are no active senders, this method doesn't work. In certain Inter-Domain multicast routing protocols, group members in domains cause state to exist whether there are active senders or not, so it is advantageous to know as soon as possible (e.g. this is an argument for N-A D-W L). The disadvantage is yet more protocol mechanism (mostly involving more timers on the border routers) and more protocol activity inside domains requiring this message.

4. Interior Router Behavior

4.1. General Behavior

This section describes the general behavior of interior routers, or of proxies running inside domains. Some of these behaviors may be optimized when running multicast routing protocols with more centralized knowledge of group memberships inside the domain; these optimizations will be described later.

If a router has not yet been upgraded to perform domain-wide reports, a proxy may be placed on each of its connected networks. This proxy must partipicate in the networks's group membership protocol (e.g. IGMPv2[Fenn97]). For example, it may perform only the duties of a Non-Querier in IGMPv2, which allow it to passively learn all of the group members on a network. The proxy can then respond to Domain-Wide Query messages just as the interior router would.

4.1.1. Reception of Queries

All routers in the domain join the Domain-Wide Query well-known

multicast group. Upon reception of a Domain-Wide Query message, a router sets a timer to a value randomly chosen from the range (0, Response time] as specified in the packet. The Data section of the Query should be saved to be used when the timer expires.

4.1.2. Transmission of Reports

Upon the expiry of a Domain-Wide Query timer, a router assembles a packet containing the list of group memberships on subnets directly connected to this router, excluding those that were suppressed by previous reports, and send this message to the Domain-Wide Report well-known multicast group. If the Domain-Wide Query contained a list of groups or options, the Report should be restricted to those groups in the list in the Query message.

4.1.3. Reception of Reports

All routers in the domain join the Domain-Wide Report well-known multicast group in order to perform suppression, as follows. Upon reception of a Domain-Wide Report message, a router processes the list of groups in the message. For each group, if the group has unrecognized options, it should be skipped if any of the unrecognized options have their S bit set. Otherwise, if the router has attached members of the group, it should mark that record as being suppressed by another report, and record the existence of this group membership in the domain. This record MUST time out after >>>XXX<<<, and MUST be cancelled by reception of a Domain-Wide Leave or Non-Authoritative Domain-Wide Leave message mentioning this group.

4.1.4. Reception of Leaves

Upon the reception of a Domain-Wide Leave, a router should process the list of groups in the message. For each group, if the group has unrecognized options, it should be skipped if any of the unrecognized options have their S bit set. Otherwise, the router should remove its record of the existence of another group membership in the domain.

4.1.5. Transmission of Leaves

A router sends a Non-Authoritative Leave when all directly-attached group members leave the group. This message MAY be suppressed if some other router was the last to report group membership with a DWR.

4.2. Optimizations

In explicit group membership protocols like PIM, CBT and MOSPF, there is a set of routers smaller than "all routers in the domain" which knows of group memberships in the domain. This section describes the optimizations possible when running a protocol like this.

In PIM and CBT, only RP's and Cores must participate. MOSPF is a special case, in that all routers in the MOSPF domain know of all group memberships in the domain. In this case, the DWR protocol may degenerate to a virtual protocol run entirely inside the elected border router.

4.2.1. Reception of a Query Message

Only participating interior routers must join the Domain-Wide Query well-known multicast group.

4.2.2. Transmission of a Report Message

Report messages contain all memberships that this router knows about (e.g. in MOSPF, it's all memberships in the domain; in PIM, it's all groups that I'm the RP for).

4.2.3. Reception of a Report Message

If there is no overlap of the groups being reported by each participant, the interior routers need not join the Domain-Wide Report well-known multicast group so will not receive Report messages. E.g. if R1 and R2 each handle one half of the multicast group address space, there is no need for them to join the Domain-Wide Report group.

4.2.4. Reception of a Leave Message

As with Reports, if there is no overlap, the interior routers need not join the DWR group so will not receive these messages.

4.2.5. Transmission of a Leave Message

If it is possible to know when the last member in the domain goes away, send authoritative Domain-Wide Leave messages, instead of Non-Authoritative Domain-Wide Leave messages.

5. Unsolicited messages

When a new group member appears in the domain, a Report message

SHOULD be transmitted (called an Unsolicited Report). Interior routers MAY track the presence of group members inside the domain; a router which is doing this SHOULD suppress its unsolicited Report if it knows of another group member inside the domain.

6. Border Router Behavior

6.1. Send Periodic Queries

The elected border router should send periodic Queries.

6.2. Querier Election

All border routers should join the Domain-Wide Queries well-known multicast group, in order to perform Querier Election. All routers start up thinking they are the elected Querier. If a router hears a DWQ from a router with a lower IP address, it elects that router as Querier. It is recommended to have an IPSEC[Atki95] relationship among the domain border routers so that Querier Election can not be fouled by a forged packet.

6.3. Send Group-Specific Queries in response to Non-Authoritative Leaves

6.4. Request Unicast Replies

If a Border Router wishes to reduce the amount of DWR multicast traffic in a domain, it may add the "Reply via Unicast" option to its DWQ's. This has the advantage of reducing the amount of state kept inside the domain for forwarding packets destined to the DWR-reply multicast group, at the cost of eliminating suppression. The border router must multicast DWR's sumarrizing the replies it got via multicast to the DWR-reply multicast group when the response interval is over.

7. Miscellaneous issues

7.1. Unsolicited Reports

What about unsolicited reports when there's no domain-wide querier? Does a router wait to hear a query before it sends its unsolicited DWR's? If not, how do we protect the current MBone which has no

domain boundaries against unsolicited reports? If so, what about a router that has just rebooted?

7.2. Non-authoritative leave messages

In DVMRP and PIM-DM domains, where a router knows that its

directly-attached group member has left a group but does not know whether or not there is another member in the domain. Therefore, it could send a non-authoritative leave message which solicits a group-specific query from the querying border router, similar to Leave messages in IGMPv2[Fenn97].

Some argue that it's possible to implement this functionality with the natural Prune messages inside the domain -- if a group gets pruned for all sources all the way to the border routers, then the border routers could send an authoritative leave message, and there would be no need for non-authoritative messages (which simplifies the protocol). However, a) the border routers have to agree that they have all received Prune messages, which is potentially complicated, and b) it's possible for all DVMRP or PIM-DM state to time out for a quiescent sender, so there are no prunes sent when the last group member in the domain leaves the group. Although it's true that no traffic is flowing, and when traffic does flow prunes will occur and the border routers will notice and send an authoritative domain-wide leave as above, the fact that this domain is a member of the group will take state in the next-level multicast routing protocol. Non-authoritiative leaves allow the removal of said state when the last member leaves, which is potentially much sooner than the next traffic to the group.

7.3. Elect Reporters?

In order to help with suppression in a domain, a Querier might choose to elect a "Designated Reporter" for a group for a certain duration. The Designated Reporter is the only router which should send Reports for the designated group(s) for the designated time. All others should act as though they have been suppressed for the designated time. The Querier should cancel a router's Designated Reporter status when that router sends a Leave message, or when it hasn't heard a reply from that router for <N> Query intervals. When cancelling a router's Designated Reporter status, a Querier

should send a Group-Specific Query for the group(s) in question and can optionally elect one of the responders as the new Designated Reporter.

7.4. Both I and S?

Both I and S bits are required. Consider a Border Router B, and interior routers X and Y. X understands and generates the source-specific membership option, and Y does not. If there were only one bit, X should set it so that Y's whole-group membership report was not suppressed by X's source-specific membership report. However, if B doesn't understand a source-specific membership report and X is the only member in the domain, B will ignore X's source-specific

membership report. This motivates the seperate S bit.

The seperate I bit was originally motivated by the idea of routers announcing group/RP mappings using DWR's and an option (while routers may not necessarily have members, they might still want to announce the mapping). With the evolution of PIM and CBT this particular use is no longer required but is perhaps general enough to motivate inclusion of the bit for future use.

8. Acknowledgements

The ideas of unicasting DWR replies and of electing a "designated reporter" came from a discussion on the IDMR mailing list with Jeffrey Zhang and Yunzhou Li of Bay Networks. This discusison is still ongoing =)

9. Security Considerations

Many same as IGMPv2[Fenn97]

9.1. Unicast Responses

Sending a DWQ requesting a unicast response can cause many DWR's to be unicasted to the sender. Requiring IPSEC authentication on the DWQ only if it requests unicast to a different address may not be strong enough - for example, someone at the other end of a slow link may swamp the link by sending a DWQ.

10. Author's Address

William C. Fenner
Xerox PARC
3333 Coyote Hill Road
Palo Alto, CA 94304
Phone: +1 650 812 4816
Email: fenner@parc.xerox.com