Technical Disclosure Commons

October 03, 2018

INTERNET GROUP MANAGEMENT PROTOCOL GENERAL QUERY MECHANISM EXTENSION TO SUPPORT MASS WITHDRAW CASE WITH ETHERNET VIRTUAL PRIVATE NETWORK ALL-ACTIVE MULTI-HOMING

Mankamana Mishra

Ali Sajassi

Al Rocheleau

Samir Thoria

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

INTERNET GROUP MANAGEMENT PROTOCOL GENERAL QUERY MECHANISM EXTENSION TO SUPPORT MASS WITHDRAW CASE WITH ETHERNET VIRTUAL PRIVATE NETWORK ALL-ACTIVE MULTI-HOMING

AUTHORS:
Mankamana Mishra
Ali Sajassi
Al Rocheleau
Samir Thoria

ABSTRACT

Techniques are described herein for providing an Internet Group Management Protocol (IGMP) general query via an Ethernet Virtual Private Network (EVPN) router irrespective of the IGMP querier role in order to prevent multicast service outage. This may provide an all-active multi-home deployment to minimize service impact in case of network failure. Procedures defined herein achieve this for multicast traffic.

DETAILED DESCRIPTION

There exist mechanisms to support multicast receivers behind an all-active multi-homing segment. Because a Customer Edge (CE)'s Link Aggregation (LAG) flow hashing algorithm is unknown, in an all-active redundancy mode it must be assumed that the CE can send a given Internet Group Management Protocol (IGMP) message to any one of the multi-homed Provider Edges (PEs). This may involve a Designated Forwarder (DF) or a non-DF setup (i.e., different IGMP join messages can arrive at different PEs in the redundancy group). Therefore, all PEs attached to a given Ethernet Segment (ES) must coordinate IGMP join state to peers in a redundancy group. This allows the Ethernet Virtual Private Network (EVPN) DF to have the right multicast states and pull the traffic from the core and forward to the access network.

IGMP (V2,V3) is protocol used by Internet Protocol version 4 (IPv4) systems to report their IP multicast group memberships to neighboring multicast routers.

Three variants of the query message have been defined: general query, group-specific query, and group-and-source-specific-query. A general query is sent by a multicast router to learn the complete multicast reception state of the neighboring interfaces. General queries are periodically sent out by multicast routers (querier) on a Local Area Network

5705

(LAN) to refresh membership state. A group-specific query is sent by a multicast router to learn the reception state with respect to a single multicast address of the neighboring interfaces. This query is used while leave processing. A group-and-source-specific-query is sent by a multicast router to learn if any neighboring interface desire reception of packets sent to a specified multicast address from any of a specified list of sources. This is also one of the queries used to process leaves.

All of these queries are generated by the IGMP querier using a pre-defined querier election procedure. Additional background information is available at https://tools.ietf.org/html/rfc2236 and https://tools.ietf.org/html/rfc3376.

Existing IGMP proxy implementations may fail in certain situations, causing traffic loss for a possibly general query interval. Figure 1 below illustrates an example topology. It depicts only the Last Hop Router (LHR) (i.e., towards the receiver) section of the topology. PE1 and PE2 are part of a redundancy group. They are EVPN peers connected to the CE via an all-active LAG. In this example, PE1 is the EVPN DF on the ES.
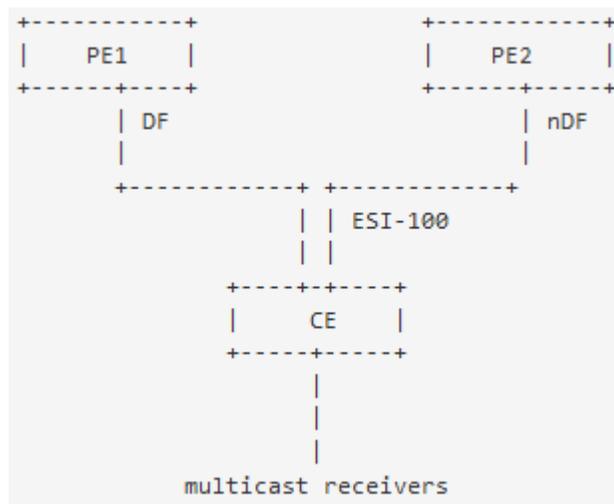
```
+-----------+                  +-----------+
|   PE1     |                  |   PE2     |
+-----+-----+                  +-----+-----+
      | DF                           | nDF
      |                              |
      +-----------+ +-----------+
                  | | ESI-100
                  | |
          +----+-+----+
          |    CE     |
          +-----+-----+
                |
                |
                |
        multicast receivers
```

Figure 1

There may be an independent multicast router (other than PE1 and PE2) on the LAN that is an IGMP querier. Here, the multicast router would be responsible for generating the IGMP query on the LAN.

The multicast receivers may start sending IGMP joins and all of them are hashed to PE2 from CE. Now PE2 synchronizes routes (e.g., EVPN multicast route type 7) to PE1. At this point, PE2 may have a local join state and PE1 may have a remote join state.

A timeline of events in the network is provided as follows. At t0, a general query is sent out with a five-second Maximum Response Time (MRT). At t5, membership is refreshed. At t6, an interface (ES) goes down between CE and PE2. At t125, the next general query is scheduled.

If the interface from CE to PE2 goes down at t6, PE2 is going to remove all local routes. It is also going to withdraw all the routes as well. Since the next general query is scheduled only at t125, PE1 is not going to learn about host before next query. This would cause traffic loss for about 115 seconds.

PE1 may have two possible roles with respect to IGMP: IGMP querier or IGMP non-querier. If PE1 has the querier role, even if PE1 is the querier in network, its next general query is scheduled only at t125. With respect to an IGMP / IGMP snooping interface between PE2 and CE2 going down, this has no impact on sending out the general query. If PE1 has the non-querier role, it is only going to act on EVPN join / withdraw routes. It is not responsible for any query mechanism.

Host are agnostic to ES down, so hosts are also not going to send out membership requests before they receive the query. This would cause multicast service outage. PE1 would have no reason to send out any query irrespective of its role as IGMP querier.

An extension to the current IGMP query mechanism is defined herein to avoid traffic loss in EVPN multi-homing all-active deployments. The extension involves notification via IGMP of a down EVPN ES and an IGMP general query mechanism on an EVPN router irrespective of the IGMP querier role.

General query is used instead of defining a new type of query. Even though an IGMP query mechanism is introduced which would be applicable to certain events and only some of the bridge ports on the Bridge Domain (BD), it would not be good idea to introduce new type of IGMP query message. Since IGMP query type has already been deployed widely, adding a new query type would require upgrading each host in the networks, which would not be acceptable in many deployments.

With respect to the host, it does not matter where the query was originated. As soon as the query is received the host sends out the membership request. So even if EVPN PE is a non-querier and originates the general query it is not going create any impact with

respect to the host. The host may treat it as general query and send out the membership request.

If there are other routers in the network, once they receive the general query they may perform querier election. Thus, it would not have any negative impact in network.

For ES Route (EVPN Route Type 4) withdraw, when a new multi-home Ethernet Segment Identifier (ESI) (non-zero ESI) is configured on network, it is supposed to originate the ES route (EVPN route type 4) to notify other peers about its existence in the network and perform EVPN DF election. When the ESI goes down, route type 4 is withdrawn.

For example, PE1 and PE2 may be an EVPN peer. PE1 and PE2 see each other as multi-home NBRs. Multicast receivers start sending IGMP membership requests which are hashed to PE2. Using the EVPN multicast route, each of the membership requests from the host are synchronized to PE1. Depending on the DF state, the EVPN PE may forward traffic towards the access network.

When the CE to PE2 interface goes down, the ES route (EVPN route type 4) withdraw may be triggered from PE2 to PE1. When the route withdraw is received by PE2, it should send out a general query irrespective of its IGMP querier state. The general query may include an MRT. The general query originating from PE1 may go to each bridge port of each BD associated with the ES for which the route withdraw has been received. The same ES may be part of multiple BDs. So the query has to be sent out for each of the BDs. This query is an extension to the general query, and there is no reason for it to be flooded on all bridge ports of the BD. The BD can have many non-multi-homed ports as well, and there is no need to send out the query. The IGMP procedure on PE1 may not remove the remote routes before the MRT timer expires. Once the MRT timer expires on PE1, it should have already learnt about active multicast host locally. It can then remove remote routes.

In case of an ES flap (e.g., ES route withdraw followed by ES route), first the PE2 - CE interface goes down. PE2 withdraws the ES route, which triggers the IGMP query to go out of PE1. PE2 withdraws all of the IGMP routes. PE1 waits for MRT before removing the routes. When the PE2 - CE links come back up, the host may reply to the previous query hashed back to PE2. Upon reception of IGMP membership request, PE2 sends the join synchronized route to PE1. PE1 receives the routes and stops the timer which was

running to clean up the remote route. There is no need to remove these routes. If routes were already removed, it may add them back.

If there are multiple PEs in a redundancy group, this procedure may still be utilized. For example, PE1, PE2, …, PEn may create a multi-home segment to a CE. If the PE2-CE link goes down, the other PEs get the ES withdraw. Each of them would receive a notification for the multicast mass withdraw. IGMP snooping may send out the query from each of the PEs. Depending on the DF, the state query may be sent out. Membership join may now be hashed to set of available PEs (PE1, PE2, …, PEn).

In this case, (S,G) may have been hashed to PE2 initially and PE1 may have had a remote state for it. After the mass withdraw procedure, this membership request may get hashed to PE4. Now PE4 may originate a new route for the same (S,G). Once PE1 receives the route before the MRT timer expires, it is going to stop the timer and not remove the route. With respect to the IGMP snooping remote route, this is still active and it does not matter if the originator of the route has changed.

The term "query" is widely used in the context of IGMP / Multicast Listener Discovery (MLD) and IGMP / MLD snooping. Queries are generated in the IGMP/MLD context for a startup (generated by each IGMP instance as soon as it comes up), query interval (General Query timer expires), or Group Specific Query (before processing leave). Queries are generated in the IGMP/MLD snooping context based on the Spanning Tree Protocol (STP) state of a port (forwarding/blocking).

In an all-active segment, there is no change of state for the port which is still up. This procedure is not only applicable to snooping, but is also applicable to Layer 3 (L3) ports. Any mechanism that provides support for all-active segments (not only EVPN) may require mechanisms described herein.

The procedure defined herein is not only applicable to situations where the access interface goes down, but also to core isolation. Core isolation occurs when there is no topology change with respect to the access network. When core isolation occurs, each of the PE in the redundancy group cannot see each other as a Border Gateway Protocol (BGP) neighbor any more. State may not be held for long. Each of the PE should update states right away (e.g., if the next query is planned after "x" seconds of core isolation).

If the query is not originated, and entries are not marked and swept within MRT, access traffic may occur. In one example, PE1 and PE2 are part of the same redundancy group. PE1 has a join state, and after core isolation if PE1 does not update its states, and if the host sends a leave message which arrives at PE2, PE1 may not know about it until the next query is sent. Routes may be removed only after two query intervals plus some amount of time. This means that traffic may be sent towards the access network even if the host sent the leave message. If immediate leave was configured, this may be an undesirable state for a user.

During core isolation and before the next query, if the CE hash algorithm sends the same join to a different PE, both PEs may be in a forwarding state and may send duplicate flows for two query intervals, which would not be acceptable to users.

Each of the PEs in the redundancy group are not guaranteed to be queriers in the BD/LAN. The querier may have no information of any of these events. Techniques described herein introduce a new event to trigger the general query which may be sent to all multi-homed segments only. This is optimized to avoid unnecessary flooding of reports in non-multi-homing segments. States may be relearned within the MRT timer. It may be tuned depending on the tolerance of the network. Remote route states that were not refreshed by the MRT timer may be cleaned up. Those routes should be marked as stale and removed.

In summary, techniques are described herein for providing an IGMP general query via an EVPN router irrespective of the IGMP querier role in order to prevent multicast service outage. This may provide an all-active multi-home deployment to minimize service impact in case of network failure. Procedures defined herein achieve this for multicast traffic.

5705

7