IDMR Working Group
Internet Draft
February 2001
Expiration Date: August 2001

M. Christensen
Exbit Technology
F. Solensky
Gotham Networks

IGMPv3 and IGMP Snooping switches
<draft-ietf-idmr-snoop-00.txt>

Status of this Memo

Abstract

This memo describes the interoperability problems and issues that can arise when a mixed deployment of IGMPv3 and IGMPv2 capable hosts and routers are interconnected by a switch with IGMP snooping capabilities. It is intended as a accompanying document to the IGMPv3 specification.

The memo contains a brief IGMP walk through followed by a description of the IGMP snooping functionality. Specific examples are given which are all based on Ethernet as the link layer protocol. Finally recommendations are given for the behavior of IGMP snooping switches.

The purpose of this document is twofold:

- We want to summarize IGMP snooping induced problems so that IETF can take appropriate actions when deciding on new protocols and behaviors.

- We also hope to bring the attention to switch vendors so that we can minimize the interoperability problems in the future.

1. Introduction

In recent years, a number of commercial vendors have introduced products described as "IGMP snooping switches" to the market. These devices do not adhere to the conceptual model that provides the strict separation of functionality between different communications layers in the ISO model and instead utilizes information in the upper level protocol headers as factors to be considered in the processing at the lower levels. This is analogous to the manner in which a router can act as a firewall by looking into the transport protocol's header before allowing a packet to be forwarded to its destination address.

In the case of multicast traffic, an IGMP snooping switch provides the benefit of conserving bandwidth on those segments of the network where no node has expressed interest in receiving packets addressed to the group address.

2. IGMP snooping overview

For a full description of IGMP we refer to [IGMPv3], however, IGMP operation can be summarized in the following:

* Hosts send IGMP Membership Report messages to inform neighboring routers that they wish to join a specific IP multicast group.

* IGMPv3 Membership Reports may be qualified with a list of allowed or forbidden source addresses.

* Routers periodically send IGMP Group Query messages to Hosts in order to maintain group membership state information. These queries can be either general or group specific queries.

* Hosts respond to queries with membership Reports.

* Hosts running either IGMPv2 or IGMPv3 may also send a Leave Group

message to routers to withdraw from the group.

A traditional Ethernet network may be separated into different network segments to prevent placing too many devices onto the same shared media. These segments are connected by bridges and switches.

When a packet with a broadcast or multicast destination address is received, the switch will forward a copy into each of the remaining network segments in accordance with [BRIDGE]. Eventually, the packet is made accessible to all nodes connected to the network.

This approach works well for broadcast packets that are intended to be seen or processed by all connected nodes. In the case of multicast packets, however, this approach could lead to less efficient use of network bandwidth, particularly when the packet is intended for only a small number of nodes. Packets will be flooded into network segments where no node has any interest in receiving the packet. While nodes will rarely incur any processing overhead to filter packets addressed to unrequested group addresses, they are unable to transmit new packets onto the shared media for the period of time that the multicast packet is flooded.

The problem of wasting bandwidth is even worse when the LAN segment is not shared, for example in Full Duplex links. Full Duplex is standard today for most switches operating at 1Gbps or above. In this case the bandwidth that is wasted is proportional to the number of attached nodes.

Allowing switches to snoop IGMP packets is a creative effort to solve this problem. The switch uses the information in the IGMP packets as they are being flooded throughout the network to determine which segments should receive packets directed to the group address.

IGMP snooping is being implemented slightly different by different switch vendors. We will not address specific implementations here as documentation is not widely available. For details of one implementation we refer to [CISCO].

In the following we will describe problems in relation to IGMP snooping with the following constraints, which we believe are the most common cases.

1. Group membership is based on multicast MAC addresses only.

2. Forwarding is based on port masks for each supported multicast group.

3. The switch is equipped with a CPU for maintaining group membership information.

Constraint 3 above is not a strict requirement as IGMP snooping could be accomplished entirely in hardware. However it becomes more difficult to support future modifications to the protocol.

IGMP snooping switches build forwarding lists by listening for (and in some cases intercepting) IGMP messages. Although the software processing the IGMP messages may maintain state information based on the full IP group addresses, the forwarding tables are typically mapped to link layer addresses. An example of such a forwarding table is shown in Figure 1.

Multicast MAC address	Member ports
01-00-5e-00-00-01	2, 7
01-00-5e-01-02-03	1, 2, 3, 7
01-00-5e-23-e2-05	1, 4

Figure 1.

Because only the least significant 23 bits of the IP address are mapped to Ethernet addresses [RFC1112], there is a loss of information when forwarding solely on the destination MAC address. This means that for example 224.0.0.123 and 239.128.0.123 and similar IP multicast addresses all map to MAC address 01-00-5e-00-00-7b (for Ethernet). As a consequence, IGMP snooping switches may collapse IP multicast group memberships into a single Ethernet multicast membership group.

Finally, it should be mentioned that in addition to building and maintaining lists of multicast group memberships the snooping switch should also maintain a list of multicast routers. When forwarding multicast packets they should be forwarded on ports which have expressed joined using IGMP but also on ports on which multicast

routers are attached.

2.1. Problems in older networks

The drawback of using IGMP snooping switches to make the flooding of multicast traffic more efficient is that the underlying link layer topology is required to remain very stable. This is especially true in IGMP versions 1 and 2 where group members do not transmit membership report messages after having overheard a report from another group member.

This problem can be demonstrated with an example. In the topology illustrated in figure 2, a topology loop exists between four IGMP snooping switches labeled A, B, C and D.

- The spanning tree algorithm would detect this loop and disable one of the links; for example, the link connecting ports B3 and

C1.

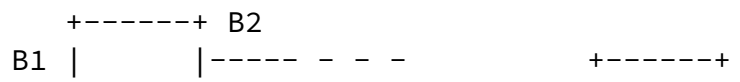
- Node N1 transmits a group membership report which will be flooded throughout the network.

- When switch A hears the report, it determines that packets addressed to the group should be forwarded to port A3.

- Router R hears the Join message and starts forwarding packets with the multicast destination address into the network. Node N1 is now part of the group.

- The link between D2 and C2 is broken. The spanning tree algorithm reactivates the blocked link B3-C1.

- If switch A relies solely on the exchange of IGMP messages to alter its forwarding behavior, node N1 will be unable to receive packets forwarded to the group address until router R sends out another membership query request.



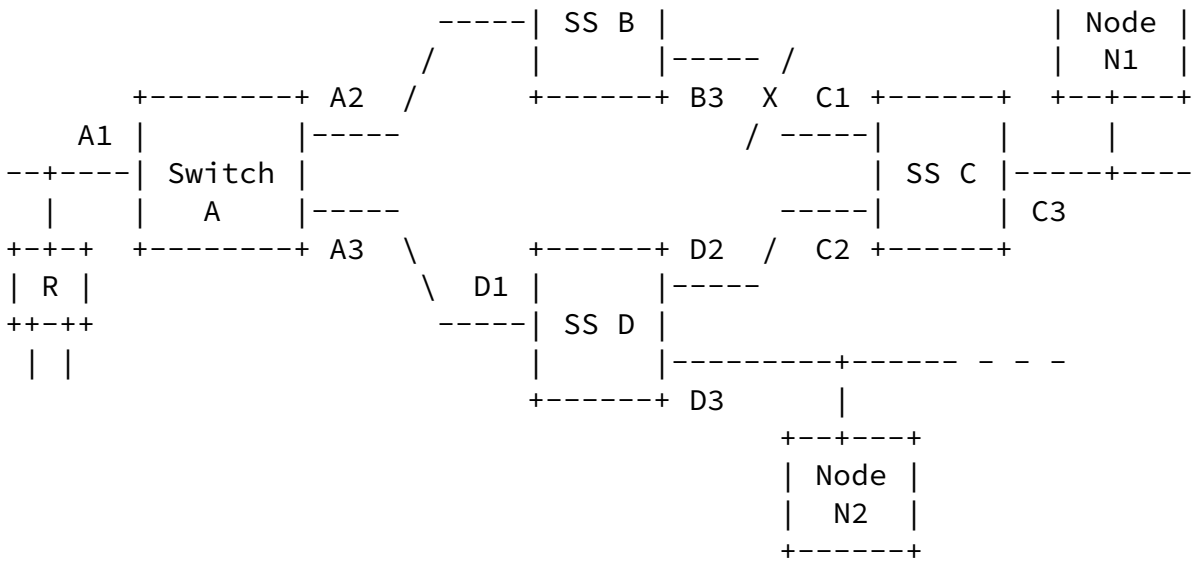


Figure 2

One possible approach to work around this limitation would be for the switch to keep track of which nodes belong to the group, altering the forwarding tables whenever a member becomes visible through a different port. When switch A sees that node N1 has moved from port A3 to A2, the group membership table would be updated. This does not work, however, when more than one node joins the same group address when at least one of them has not yet been upgraded to IGMPv3: if nodes N1 and N2 were to join the group at approximately the same time, they would both start off random timers for the transmission of their first membership report

messages. If node N2 selects a longer interval than N1, it will hear N1's report message and cancel the one it was about to send. Switch A, therefore, never learns that node N2 has joined the group. When the switch learns that N1 is now accessible through port A2, it has no way of knowing that it should continue forwarding group packets to port A3 as well.

2.2. IGMPv2 snooping and 224.0.0.X

Special attention should be brought to the address range from 224.0.0.0 through 224.0.0.255 which is reserved for routing protocols and other low-level topology discovery or maintenance protocols [IANA]. Examples of reserved multicast addresses are:

224.0.0.2 All Routers on this Subnet
224.0.0.4 DVMRP
224.0.0.5 MOSPF
224.0.0.6 MOSPF
224.0.0.9 RIP2 Routers
224.0.0.13 PIM Routers
224.0.0.22 IGMPv3 Membership Reports

Multicast routers are discouraged from routing packets with a destination address falls within this range, regardless of the TTL value. The router will be the originator or consumer of these messages so it has less of a motivation to maintain forwarding path information for these addresses. As a result, it becomes less critical for the router to send out periodic Query messages for these groups. If the router chooses not to the group would be unable to recover from topology changes as described above. Note that the only difference between the 'all hosts' address (224.0.0.1) and the remainder of this range is that the router has no discretion in the former case: it MUST NOT send Queries.

To avoid this situation, IGMP snooping switches should be less conservative when forwarding packets to these addresses and flood them to all ports.

It is reported in [MSOFT] that a number of switches can be mis- configured to perform IGMP snooping and forwarding for all IP multicast groups.

Figure 3 illustrates one scenario where two routers R1 and R2 are communicating using for example 224.0.0.6. The routers never send IGMP joins for this address. The switch floods the (unknown) multicast traffic on all ports.

Now the server SVR is started and it sends an IGMP join for 224.0.0.6, which is snooped by the switch. It then generates a membership query on all ports to determine which ports have devices that also belong to this group.

The routers R1 and R2 do not respond and the switch builds a forwarding port list with only SVR in it. Now R1 and R2 are not able to communicate using this address.

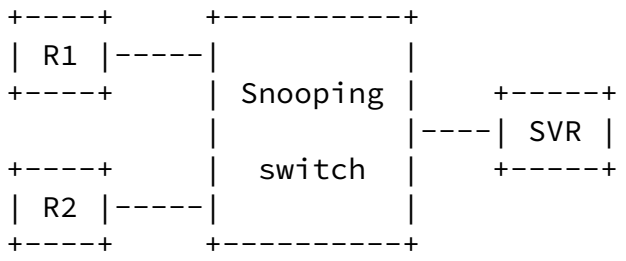


Figure 3.

There are two possible fixes to this problem: One is to require that all routers (also being hosts) which use IP multicast responds to IGMP queries in the range 224.0.0.X. This seems unnecessary as discussed above because of the inherent link local scope of these messages.

Another solution to this problem, which is also discussed above, is that the switch is configured to forward all packets for a range of IP multicast addresses to all ports (flooding).

It is suggested that all multicast packets in the range 224.0.0.1 through 224.0.0.255 are forwarded on all ports.

2.3. IGMPv2 and IGMPv3 coexistence

Consider the following sequence of communication (figure 4.):

- Router R sends IGMPv3 Query
- Host H1 sends IGMPv2 Report (since it has only implemented v2).
- switch S puts H1's port P1 in the flooding list.
- Host H2 sends IGMPv3 Report.
- Switch S fails to put H2's port P2 in the flooding list because it doesn't support IGMPv3.

- H2 never sees any traffic.

{{need to provide description of solution that allows this. Step 4 sounds wrong}}

2.4. Source Specific Joins

Even for IGMPv3 snooping capable switches there can be limitations caused by link layer based forwarding. This is illustrated in figure 4.

Assume that host H1 sends a Join(S1, G) to R and that host H2 sends a Join(S2, G) to R.

The switch adds both hosts to the forwarding list for group G.

Frames originating from sources S1 and S2 for the same multicast address G are routed via R. These are sent from R with the router's MAC address as source.

The switch is unable to distinguish the two different types of flow and forwards both flows to both hosts. This effectively disables the Join source functionality in this network configuration.

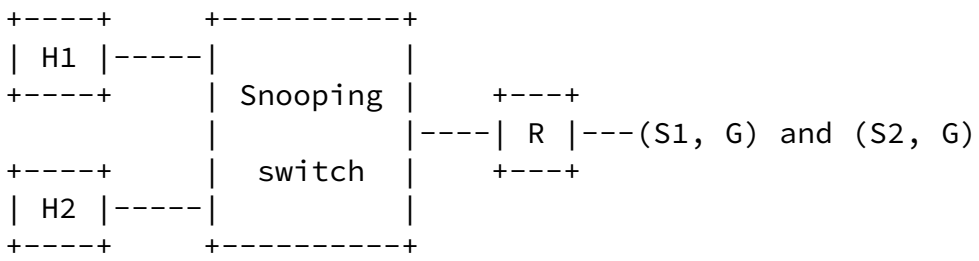


Figure 4.

This is a problem without an obvious solution because of the difference between the link layer and the network layer information.

One approach would be for the switch to simply flood the packets to both ports. This requires that the host implementations do not rely on the router to perform all of the source address filtering for the group address: they must still filter out packets that do not match the source address criteria specified in the join messages. While this might be seen as an inconvience, this is no different than the case where the router is directly connected to both hosts on a shared LAN and no snooping switch is present.

An alternative approach would be for the switch to further qualify the search process by including source address when determining which

ports should forward the packet. While this could work for very simple cases, it is unlikely that this approach could scale into more complex topologies or provide satisfactory performance in even the simple cases.

Similar problems occur with the attempt to exclude sources.

3. Snooping Requirements

The switch that provides support for IGMP packet snooping MUST forward all unrecognized IGMP messages and MUST NOT attempt to make use of any information beyond the end of the network layer header. In particular, messages where any Reserved fields are non-zero MUST NOT be snooped since this could indicate an incompatible change to the message format.

If a switch receives a multicast packet without having first processed Membership Reports for the group address, it MUST forward the packet into all active network segments. In other words, the switch must allow for the possibility that connected hosts and routers have been upgraded to support future versions or extensions of IGMP that the switch does not yet recognize. A switch MAY have a configuration option that suppresses this operation, but default behavior MUST be to allow flooding of unregistered packets.

In order to operate correctly, the switch supporting IGMP snooping MUST also maintain a list of multicast routers. This list SHOULD be built using IGMP Multicast Router Discovery [MRDISC] which is currently going through IETF Last Call. IGMP snooping switches MAY in addition use information about which ports packets for the address 224.0.0.X range arrive, when

- The packets are IGMP Queries or
- The packets are anything but IGMP or
- The ports are manually configured as having multicast routers attached

4. Security Considerations

Security considerations for IGMPv3 are accounted for in [IGMPv3]. The introduction of IGMP snooping switches adds the following consid-

erations with regard to IP multicast.

The exclude source failure which could cause traffic from sources that are 'black listed' to reach hosts that have requested otherwise. This can also occur in certain network topologies without IGMP snooping.

It is possible to generate packets which make the switch wrongly believe that there is a multicast router on the segment on which the sender is attached. This will potentially lead to excessive flooding on that segment. The authentication methods discussed in [IGMPv3] will also provide protection in this case.

Generally though, it is worth to stress that IP multicast must so far be considered insecure until the work of for example the suggested Multicast Security (MSEC) working group or similar is completed or at lease has matured.

5. References

[BRIDGE] IEEE 802.1D, "Media Access Control (MAC) Bridges"

[CISCO] Cisco Tech Notes, "Multicast In a Campus Network: CGMP and IGMP snooping"

[IANA] Internet Assigned Numbers Authority, "Internet Multicast Addresses", http://www.isi.edu/in-notes/iana/assignments/multicast-addresses

[IGMPv3] Cain, B., "Internet Group Management Protocol, Version 3", draft-ietf-idmr-igmp-v3-06.txt, November 2000

[MRDISC] Biswas, S. "IGMP Multicast Router Discovery", draft-ietf-idmr-igmp-mrdisc-05.txt, October 2000.

[MSOFT] Microsoft support article Q223136, "Some LAN Switches with IGMP Snooping Stop Forwarding Multicast Packets on RRAS Startup", http://support.microsoft.com/support/kb/articles/Q223/1/36.ASP

[RFC1112] Deering, S., "Host Extensions for IP Multicasting", RFC 1112, August 1989.

[RFC2026] Bradner, S. "The Internet Standards Process -- Revision 3", RFC2026, October 1996.

[RFC2236] Fenner, W., "Internet Group Management Protocol, Version 2", RFC2236, November 1997.

6. Author's Addresses:

Morten Jagd Christensen
Exbit Technology
Hoerkaer 18
2730 Herlev
DENMARK
email: mjc@exbit.dk

Frank Solensky
Gotham Networks
15 Discovery Way
Acton, MA 01720
USA
email: fsolensky@GothamNetworks.com (effective 09 March 2001)
solensky@acm.org

Table of Contents