

IGMP and MLD snooping switches  
<[draft-ietf-idmr-snoop-01.txt](#)>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of [Section 10 of RFC2026](#) [[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/1id-abstracts.txt>

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

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. The memo also covers MLDv2 for IPv6. It is intended as an accompanying document to the IGMPv3 and MLDv2 specifications.

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. MLDv2 for IPv6 is discussed. 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 and best current solutions. We hope that a description of IGMP snooping will

be of aid to the IETF when standardizing new protocols and behaviors within this scope.

- We also hope to bring this work to the attention of switch vendors, typically active within the IEEE community but perhaps not within IETF, in order to minimize protocol 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.

The discussions in this document are based on IGMP which applies to IPv4 only. For IPv6 we must use MLD in stead. Because MLD is based on IGMP with only a few differences these discussions also apply to IPv6.

## 2. IGMP snooping overview

For a full description of IGMP we refer to [[IGMPv3](#)], however, IGMP operation is 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 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 the IEEE MAC bridge standard [[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, and it will be standard for 10Gbps ethernet too. In this case the wasted bandwidth 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 forwarded 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 a 'list' of member ports 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, when sending IGMP datagrams all is done to ensure that the packets are not routed. For example the TTL is set to 1 and the IP header contains the router alert option. This is a hint to developers that there is probably a need to send this packet to the CPU.

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 joined using IGMP but also on ports on which multicast routers are attached. The reason for this is that in IGMP there is only one active querier. This means that all other routers on the network are suppressed and thus not detectable by the switch.

### 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.
- Host H1 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. Host H1 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, host H1 will be unable to receive packets forwarded to the group address until router R sends out another Membership Query.

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 host H1 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 hosts H1 and H2 were to join the group at approximately the same time, they would both start off random timers for the transmission of their first Membership Reports. If host H2 selects a longer interval than H1, it will hear H1's report message and cancel the one it was about to send. Switch A, therefore,

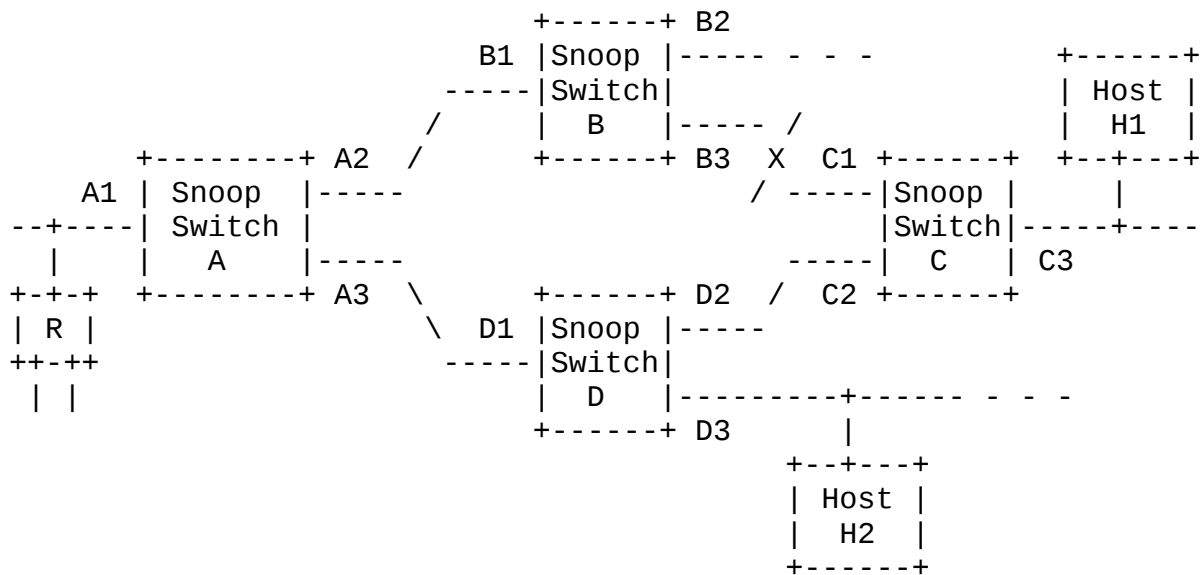


Figure 2

never learns that node H2 has joined the group. When the switch learns that H1 is now accessible through port A2, it has no way of knowing that it should continue forwarding group packets to port A3 as well.

Two recommendations can be made based on the above discussion:

- The switch should play an active role when detecting a topology change; The spanning tree root bridge (which is also a snooping switch) should initiate the transmission of a IGMP General Query, for example through signalling the CPU. This will help to reduce the join latency otherwise introduced.
- IGMP Membership Reports should not be flooded because this will lead to Join suppression.

## 2.2. IGMPv2 snooping and 224.0.0.X

Special attention should be brought to the IP address range from 224.0.0.1 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:

Multicast routers are discouraged from routing packets when 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

```

224.0.0.2  All Routers on this Subnet
224.0.0.4  DVMRP
224.0.0.5  (M)OSPF routers
224.0.0.6  (M)OSPF DRs
224.0.0.9  RIP2 Routers
224.0.0.13 PIM Routers
224.0.0.22 IGMPv3 Membership Reports

```

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.

As an example of this, it is reported in [\[MSOFT\]](#) that a number of switches can be misconfigured to perform IGMP snooping and forwarding for all IP multicast groups:

Figure 3 illustrates the 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. The switch then generates a Membership Query on all ports to determine which ports have devices attached 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.

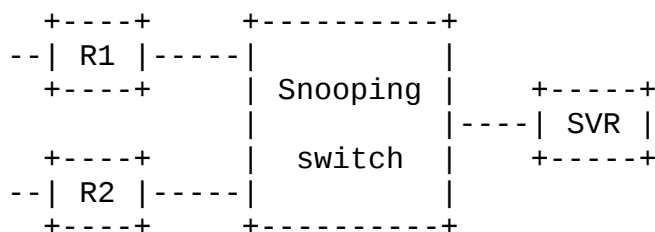


Figure 3.

There are two possible fixes to this problem: One is to require that all routers (also being hosts) which use IP multicast respond 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. This of course requires an examination of the network layer header. Note that these are IP address ranges and that mapping these to MAC address range 01-00-5e-00-00-X is subject to problems discussed in the previous sections.

### 2.3. IGMPv3 and IGMPv2 coexistence

IGMPv3 and IGMPv2 are designed to interoperate with older versions of IGMP. Both hosts and routers are capable of falling back to an earlier version when receiving older IGMP messages, thus enabling a mixed deployment and migration to new versions. While this works fine in a network of hosts and routers an IGMP snooping switch introduces problems.

In figure 4 where hosts H1 and H2 are connected to an IGMP snooping switch on ports P1 and P2 respectively, consider the following sequence of communication:

- Router R sends an IGMPv3 Query
- Host H1 sends an IGMPv2 Report since it has only implemented v2. R notices this and switches to IGMPv2 mode. The report is not received by H2 because of the snooping functionality.
- Switch S puts H1's port P1 in the forwarding table.
- Host H2 sends an IGMPv3 Report in response to R's Query.
- Switch S fails to add H2's port P2 to the forwarding table because it doesn't support IGMPv3.
- H2 does not receive any traffic before R sends its next Query which will put H2 in IGMPv2 mode.

This introduces a Join latency for host H2, which apparently cannot be avoided. The latency is potentially of the order of minutes. It is



possible however to reduce this latency by tuning the Query Interval which defaults to 125 seconds.

When operating in a mixed deployment mode it is suggested that initially the Query Interval is set to "a low value" until the compatibility modes have stabilized both host and routers on the same IGMP version. After stabilization the Query Interval could be increased to its original value.

#### 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.

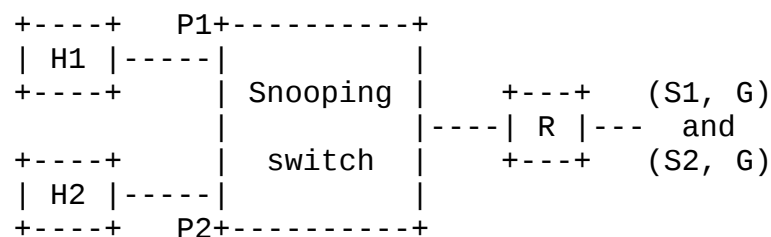


Figure 4.

This is a problem caused by layer 2 based forwarding of a layer 3 flow in conjunction with the difference between the link layer and the network layer information.

The example above means that host implementations cannot rely on the router to perform all source address filtering. Therefore they must still filter out packets that do not match the source address

criterion specified in the Join messages. While this might be seen as an inconvenience, 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 complete solution would be for the switch to further qualify the search process by including the source IP address when determining which ports should forward the packet.

Similar problems occur with the attempt to exclude sources.

### 3. Snooping Requirements

Note that in the following we provide suggestions for good/best practices when designing IGMP snooping devices. Keywords as MUST, SHOULD, MUST NOT etc. are suggestions only.

- 1) All IGMP packets (IP packets with IP-PROTO = 2) SHOULD be redirected to the CPU for IGMP snooping processing and table management. This allows for the most flexible IGMP snooping solution.
- 2) The switch that provides support for IGMP 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 subject to "normal" snooping since this could indicate an incompatible change to the message format.
- 3) Packets with a destination IP address in the 224.0.0.X range which are \*not\* IGMP SHOULD be forwarded on all ports.
- 4) Packets with a destination IP address outside 224.0.0.X which are \*not\* IGMP SHOULD be forwarded according to port membership tables and MUST also be forwarded on router ports.
- 5) If a switch receives a \*non\* IGMP multicast packet without having first processed Membership Reports for the group address, it MUST forward the packet on all ports. 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.

6) A snooping switch SHOULD forward IGMP Membership Reports on router "ports" only.

7) The switch supporting IGMP snooping MUST 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 build this list based on the arrival port for packets destined to 224.0.0.X, when

- The packets are IGMP Queries or
- The packets are \*not\* IGMP or
- The ports are configured (by management) as having multicast routers attached

8) IGMP snooping switches MAY maintain forwarding tables based on either MAC addresses or IP addresses. If a switch supports both types of forwarding tables then the default behavior SHOULD be to use IP addresses.

9) Switches which rely on information in the IP header MAY verify that the IP header checksum is correct.

10) IGMP snooping switches SHOULD inform the CPU (or hardware) when a link layer topology change has been detected. Following a topology change the switch SHOULD initiate the transmission of a General Query on all ports in order to reduce Join latency.

#### 4. IPv6 Considerations

In order to avoid confusion, the previous discussions have been based on IGMPv3 functionality which only applies to IPv4 multicast. In the case of IPv6 most of the above discussions are still valid with a few exceptions which we will describe here.

In IPv6 the protocol for multicast group maintenance is called Multicast Listener Discovery (MLDv2). IPv6 is not widely deployed today and neither is IPv6 multicast. However, it is anticipated that at some time IPv6 switches capable of MLD snooping will appear.

The three main differences between IGMPv3 and MLDv2 are

- MLDv2 uses ICMPv6 message types instead of IGMP message types.
- The ethernet encapsulation is a mapping of 32bits of the 128bit DIP addresses into 48bit DMAC addresses [[IPENCAPS](#)].



addresses in the foreseeable future.

Finally we mention the reserved address range FF0X:0:0:0:0:X:X where X is any value. This range is similar to 224.0.0.X for IPv4 and is reserved to routing protocols and resource discovery [[RFC2375](#)]. In the case of IPv6 it is suggested that packets in this range are forwarded on all ports if they are not MLD packets.

## 5. Security Considerations

Security considerations for IGMPv3 are accounted for in [[IGMPv3](#)]. The introduction of IGMP snooping switches adds the following considerations 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 source 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.

IGMP snooping switches which rely on the IP header of a packet for their operation and which do not validate the header checksum potentially will forward packets on the wrong ports. Even though the IP headers are protected by the ethernet checksum this is a potential vulnerability.

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 least has matured.

## 6. References

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

[CISCO] Cisco Tech Notes, "Multicast In a Campus Network: CGMP and IGMP snooping", <http://www.cisco.com/warp/public/473/22.html>

- [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
- [IPENCAPS] Crawford, M., "Transmission of IPv6 Packets over Ethernet Networks", [RFC2464](#), December 1998.
- [MLDV2] Vida, R., "Multicast Listener Discovery Version 2 (MLDV2) for IPv6", [draft-vida-mld-v2-00.txt](#), February 2001.
- [MRDISC] Biswas, S. "IGMP Multicast Router Discovery", [draft-ietf-idmr-igmp-mrdisc-06.txt](#), May 2001.
- [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", [RFC1112](#), 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.
- [RFC2375] Hinden, R. "IPv6 Multicast Address Assignments", [RFC2375](#), July 1998.

## 7. Acknowledgements

We would like to thank Bill Fenner, Yiqun Cai, Edward Hilquist and Martin Bak for comments and suggestions on this document.

## 8. Author's Addresses:

Morten Jagd Christensen  
Vitesse Semiconductor Corporation  
Hoerkaer 16  
2730 Herlev  
DENMARK  
email: [mjc@vitesse.com](mailto:mjc@vitesse.com)

Frank Solensky  
Gotham Networks  
15 Discovery Way  
Acton, MA 01720  
USA  
email: [fsolensky@GothamNetworks.com](mailto:fsolensky@GothamNetworks.com)  
[solensky@acm.org](mailto:solensky@acm.org)

## Table of Contents

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">2.</a>	IGMP snooping overview . . . . .	<a href="#">2</a>
<a href="#">2.1</a>	Problems in older networks . . . . .	<a href="#">5</a>
<a href="#">2.2</a>	IGMPv2 snooping and 224.0.0.X . . . . .	<a href="#">6</a>
<a href="#">2.3</a>	IGMPv3 and IGMPv2 coexistence . . . . .	<a href="#">8</a>
<a href="#">2.4</a>	Source Specific Joins . . . . .	<a href="#">9</a>
<a href="#">3.</a>	Snooping Requirements . . . . .	<a href="#">10</a>
<a href="#">4.</a>	IPv6 Considerations . . . . .	<a href="#">11</a>
<a href="#">5.</a>	Security Considerations . . . . .	<a href="#">13</a>
<a href="#">6.</a>	References . . . . .	<a href="#">13</a>
<a href="#">7.</a>	Acknowledgements . . . . .	<a href="#">15</a>
<a href="#">8.</a>	Author's Addresses: . . . . .	<a href="#">15</a>