Network Working Group
Internet Draft
Expiration Date: July 2003
Category: Informational

M. Christensen
Thrane & Thrane
K. Kimball
Hewlett-Packard
F. Solensky
Bluejavelin

January 2003

Considerations for IGMP and MLD snooping switches

<draft-ietf-magma-snoop-05.txt>

Status of this Memo

Abstract

This memo describes the requirements for IGMP- and MLD-snooping switches. These are based on best current practices for IGMPv2, with further considerations for IGMPv3- and MLDv2-snooping. Additional areas of relevance, such as link layer topology changes and Ethernet-specific encapsulation issues, are also considered.

Interoperability issues that arise between different versions of IGMP are not the focus of this document. Interested readers are directed to [IGMPv3] for a thorough description of problem areas.

This document is intended to accompany the IGMPv3 and MLDv2

specifications.

1. Introduction

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 multi- cast 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. In general, signifi- cant bandwidth can be wasted by flooding.

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 communica- tions layers in the ISO model, and instead utilize information in the upper- level protocol headers as factors to be considered in the processing at the lower levels. This is analogous to the man- ner in which a router can act as a firewall by looking into the transport protocol's header before allowing a packet to be for- warded 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 net- work where no node has expressed interest in receiving packets addressed to the group address. This is in contrast to normal switch behavior where multicast traffic is typically forwarded on all interfaces.

Many switch datasheets state support for IGMP snooping, but no requirements for this exist today. It is the authors' hope that the information presented in this draft will supply this founda-

tion.

The requirements presented here are based on the following information sources: The IGMP specifications [RFC112][RFC2236][IGMPv3], vendor-supplied technical documents [CISCO], bug reports [MSOFT],

discussions with people involved in the design of IGMP snooping switches, MAGMA mailinglist discussions, and on replies by switch vendors to an implementation questionnaire.

The discussions in this document are based on IGMP, which applies only to IPv4. For IPv6, MLD must be used instead. Because MLD is based on IGMP, we do not repeat the whole discussion and requirements for MLD snooping switches. Instead, we point out the few cases where there are differences from IGMP.

Note that the IGMP snooping function should apply only to IPv4 multicasts. Other multicast packets, such as IPv6, might be suppressed by IGMP snooping if additional care is not taken in the implementation. It is desired not to restrict the flow of non-IPv4 multicasts other than to the degree which would happen as a result of regular bridging functions. The same note can be made of MLD snooping switches with respect to suppression of IPv4.

2. IGMP Snooping Requirements

The following sections list the requirements for an IGMP snooping switch. The requirement is stated and is supplemented by a discussion. All implementation discussions are examples only and there may well be other ways to achieve the same functionality.

2.1. Forwarding rules

The IGMP snooping functionality is here separated into a control section (IGMP forwarding) and a data section (Data forwarding).

2.1.1. IGMP Forwarding Rules

1) A snooping switch should forward IGMP Membership Reports only to those ports where multicast routers are attached. Alternatively stated: a snooping switch should not forward IGMP Membership Reports to ports on which only hosts are attached. An administrative control may be provided to override this restriction, allowing the report messages to be flooded to other ports.

This is the main IGMP snooping functionality. Sending membership reports (as described in IGMP versions 1 and 2) to other hosts can result in unintentionally preventing a host from joining a specific multicast group. This is not a problem in

an IGMPv3-only network because there is no cancellation of IGMP Membership reports.

The administrative control allows IGMP Membership Report messages to be processed by network monitoring equipment such as packet analyzers or port replicators.

The switch supporting IGMP snooping must maintain a list of multicast routers and the ports on which they are attached. This list can be constructed in any combination of the following ways:

a) This list should be built by the snooping switch sending Multicast Router Solicitation messages as described in IGMP Multicast Router Discovery [MRDISC]. It may also snoop Multicast Router Advertisement messages sent by and to other nodes.

b) The arrival port for IGMP Queries (sent by multicast routers) where the source address is not 0.0.0.0.

c) Ports explicitly configured by management to be IGMP-forwarding ports, in addition to or instead of any of the above methods to detect router ports.

2) IGMP snooping switches may also implement "proxy-reporting" in which reports received from downstream hosts are summarized and used to build internal membership states as described in [PROXY]. The IGMP proxy-reporting switch would then report its

own state in response to upstream queriers. If the switch does not already have an IP address assigned to it, the source address for these reports should be set to all-zeros.

An IGMP proxy-reporting switch may act as Querier for the downstream hosts while proxy reporting to the 'real' upstream queriers.

It should be noted that there may be multiple IGMP proxy-reporting switches in the network all using the 0.0.0.0 source IP address. In this case the switches can be uniquely identified through their link layer source MAC address.

IGMP membership reports must not be rejected because of a source IP address of 0.0.0.0.

3) The switch that supports IGMP snooping must flood all unrecognized IGMP messages to all other ports and must not attempt to make use of any information beyond the end of the network layer

header.

In addition, earlier versions of IGMP should interpret IGMP fields as defined for their versions and must not alter these fields when forwarding the message. When generating new messages, a given IGMP version should set fields to the appropriate values for its own version. If any fields are reserved or otherwise undefined for a given IGMP version, the fields should be ignored when parsing the message and must be set to zeroes when new messages are generated by implementations of that IGMP version.

4) An IGMP snooping switch should be aware of link layer topology changes. Following a topology change the switch should initiate the transmission of a General Query on all ports in order to reduce network convergence time. If the switch is not the Querier, it should use the 'all-zeros' IP Source Address in these proxy queries. When such proxy queries are received, they must not be included in the Querier election process.

5) An IGMP snooping switch must not make use of information in IGMP packets where the IP or IGMP headers have checksum or

integrity errors. The switch should not flood such packets but if it does, it should take some note of the event (i.e., increment a counter). These errors and their processing are further discussed in [IGMPv3], [MLD] and [MLDv2].

6) The snooping switch must not rely exclusively on IGMP group leave announcements to determine when entries should be removed from the forwarding table. The snooping switch should implement a membership timeout feature to ensure unneeded forwarding table entries will be appropriately removed if downstream members silently leave the group or become unavailable for any reason. If the switch implements this timeout behavior, it must have a feature to override it if the switch is also configured to forward unregistered multicast packets on all ports. Additionally, if timeout is implemented, a group's forwarding table entry should be removed from a port when no IGMP report has been received for [(Query Interval x Number of Queries) + Query Response Time] seconds. These variables may be learned dynamically but IGMP snooping switches implementing timeout should have a configuration option that allows these variables to be set manually.

2.1.2. Data Forwarding Rules

1) Packets with a destination IP (DIP) address in the 224.0.0.X range which are not IGMP must be forwarded on all ports.

   This requirement is based on fact that many hosts exist today which do not Join IP multicast addresses in this range before sending or listening to IP multicasts. Furthermore since the 224.0.0.X address range is defined as link local (not to be routed) it seems unnecessary to keep state for each address in this range. Additionally, some vendors' applications, which are not IGMP, use this 224.0.0.X address range, and these applications would break if the switch were to prune them due to not seeing a Join.

2) Packets with a destination IP address outside 224.0.0.X which are not IGMP should be forwarded according to group-based port membership tables and must also be forwarded on router ports.

   This is the core IGMP snooping requirement for the data path.

   Discussion: An implementation could maintain separate membership and multicast router tables in software and then "merge" these tables into a current forwarding cache.

3) If a switch receives a non-IGMP IPV4 multicast packet without having first processed Membership Reports for the group address, it may forward the packet on all ports, but it must forward the packet on router ports. A switch may forward an unregistered packet only on router ports, but the switch must have a configuration option that suppresses this restrictive operation and forces flooding of unregistered packets on all ports. In environments with v3 hosts where the snooping switch does not support v3, failure to flood unregistered streams could prevent v3 hosts from receiving their traffic. Alternatively, in environments where the snooping switch supports all of the IGMP versions that are present, flooding unregistered streams may cause IGMP hosts to be overwhelmed by multicast traffic, even to the point of not receiving Queries and failing to issue new membership reports for their own groups.

4) All non-IPv4 multicast packets should be flooded, except where normal IEEE bridging operation would result in filtering multicast packets. Discussion: This ensures that enabling IGMP snooping does not break, for example, IPv6 multicast.

5) 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.

   Discussion: Forwarding based on MAC addresses is subject to the problem associated with the 32-fold IP address to 1 MAC address mapping.

6) Switches which rely on information in the IP header should ver-

ify that the IP header checksum is correct. If the checksum fails, the information in the packet must not be incorporated into the forwarding table. Further, the packet should be discarded.

7) When IGMPv3 "include source" and "exclude source" membership reports are received on shared segments, the switch needs to forward the superset of all received membership reports onto the shared segment. Forwarding of traffic from a particular source S to a group G must happen if at least one host on the shared segment reports an IGMPv3 membership of the type INCLUDE(G, Slist1) or EXCLUDE(G, Slist2) where S is an element of Slist1 and not an element of Slist2.

2.2. IGMP snooping related problems

A special problem arises in networks consisting of IGMPv3 routers as well as IGMPv2 and IGMPv3 hosts interconnected by an IGMPv2 snooping switch. The router will continue to maintain IGMPv3 even in the presence of IGMPv2 hosts, and thus the network will not likely converge on IGMPv2. But it is likely that the IGMPv2 snooping switch will not recognize or process the IGMPv3 membership reports. Groups for these unrecognized reports will then either be flooded (with all of the problems that may create for hosts in a network with a heavy multicast load) or pruned by the snooping switch.

Therefore it is recommended that in such a network, the multicast router be configured to use IGMPv2.

3. IPv6 Considerations

In order to avoid confusion, the previous discussions have been based on the IGMP protocol 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.

The control and data forwarding rules in the IGMP section can, with

a few considerations, also be applied to MLD. This means that the basic functionality of intercepting MLD packets, and building membership lists and multicast router lists, is the same as for IGMP.

In IPv6, the data forwarding rules are more straight forward because MLD is mandated for addresses with scope 2 (link-scope) or greater. The only exception is the address FF02::1 which is the all hosts link-scope address for which MLD messages are never sent. Packets with the all hosts link-scope address should be forwarded on all ports.

MLD messages are also not sent to packets in the address range FF0X::/16 when X is 0 or 1 (which are reserved and node-local, respectively), and these addresses should never appear in packets on the link.

The three main differences between IPv4 and IPv6 in relation to multicast are:

- The IPv6 protocol for multicast group maintenance is called Multicast Listener Discovery (MLDv2). MLDv2 uses ICMPv6 message types instead of IGMP message types.

- The ethernet encapsulation is a mapping of 32 bits of the 128 bit DIP addresses into 48 bit DMAC addresses [IPENCAPS].

- Multicast router discovery is done using Neighbor Discovery Protocol (NDP) for IPv6. NDP uses ICMPv6 message types.

The IPv6 packet header does not include a checksum field. Nevertheless, the switch should detect other packet integrity issues. When the snooping switch detects such an error, it must not include information from the corresponding packet in the MLD forwarding table. The forwarding code should drop the packet and take further reasonable actions as advocated above.

The fact that MLDv2 is using ICMPv6 adds new requirements to a snooping switch because ICMPv6 has multiple uses aside from MLD. This means that it is no longer sufficient to detect that the next-header field of the IP header is ICMPv6 in order to identify packets relevant for MLD snooping.

Discussion: If an implementation was software-based, wrongly identifying non-MLD packets as candidates for MLD snooping would potentially fill the CPU queue with irrelevant packets thus preventing the snooping functionality. Furthermore, ICMPv6 packets destined for other hosts would not reach their destinations.

A solution is either to require that the snooping switch looks further into the packets, or to be able to detect a multicast DMAC address in conjunction with ICMPv6. The first solution is desirable only if it is configurable which message types should trigger a CPU redirect and which should not. The reason is that a hardcoding of message types is inflexible for the introduction of new message types. The second solution introduces the risk of new protocols which use ICMPv6 and multicast DMAC addresses but which are not related to MLD, wrongly being identified as MLD. It is suggested that solution one is preferred if the switch is capable of triggering CPU redirects on individual ICMPv6 message types. If this is not the case, then use solution two.

The mapping from IP multicast addresses to multicast DMAC addresses introduces a potentially enormous overlap. The structure of an IPv6 multicast address is shown in the figure below. Theoretically 2**80, two to the power of 80 (128 - 8 - 4 - 4 - 32) unique DIP addresses could map to one DMAC address. This should be compared to 2**5 in the case of IPv4.

Initial allocation of IPv6 multicast addresses, however, uses only the lower 32 bits of group ID. This eliminates the address ambiguity for the time being, but it should be noted that the allocation policy may change in the future. Because of the potential overlap it is recommended that IPv6 address based forwarding is preferred to MAC address based forwarding.

```
|   8    |  4 |  4 |             112 bits                  |
+--------+----+----+---------------------------------------+
|11111111|flgs|scop|             group ID                  |
+--------+----+----+---------------------------------------+
```

4. Normative References

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

[IGMPv3] Cain, B., "Internet Group Management Protocol, Version 3", RFC3376, October 2002.

[IPENCAPS] Crawford, M., "Transmission of IPv6 Packets over Ethernet Networks", RFC2464, December 1998.

[MLD] Deering, S., Fenner, B., and Haberman, B. "Multicast

Listener Discovery (MLD) for IPv6", RFC2710, October 1999.

[MLDv2] Vida, R., "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", draft-vida-mld-v2-06.txt, November 2002.

[MRDISC] Biswas, S. "IGMP Multicast Router Discovery", draft-ietf-idmr-igmp-mrdisc-10.txt, January 2003.

[PROXY] Fenner, B. et al, "IGMP-based Multicast Forwarding (IGMP Proxying)", draft-ietf-magma-igmp-proxy-01.txt, November 2002.

[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.

[RFC2375] Hinden, R. "IPv6 Multicast Address Assignments", RFC2375, July 1998.

5. Informative References

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

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

[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

6. 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.

1) 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.

2) 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.

3) 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.

4) In IGMP, there is no mechanism for denying recipients access to groups (i.e. no "exclude receiver" functionality). Hence, apart from IP-level security configuration outside the scope of IGMP, any multicast stream may be received by any host without restriction.

Generally, IGMP snooping must be considered insecure due to the issues above. However, none of the these issues are any worse for IGMP snooping than for IGMP implementations in general.

7. IGMP Questionnaire

As part of this work the following questions were asked both on the MAGMA discussion list and sent to known switch vendors implementing IGMP snooping. The individual contributions have been anonymized

upon request and do not necessarily apply to all of the vendors' products.

The questions were:

Q1 Does your switches perform IGMP Join aggregation? In other words, are IGMP joins intercepted, absorbed by the hardware/software so that only one Join is forwarded to the querier?

Q2 Is multicast forwarding based on MAC addresses? Would datagrams addressed to multicast IP addresses 224.1.2.3 and 239.129.2.3 be forwarded on the same ports-groups?

Q3 Is it possible to forward multicast datagrams based on IP addresses (not routed)? In other words, could 224.1.2.3 and

239.129.2.3 be forwarded on different port-groups with unaltered TTL?

Q4 Are multicast datagrams within the range 224.0.0.1 to 224.0.0.255 forwarded on all ports whether or not IGMP Joins have been sent?

Q5 Are multicast frames within the MAC address range 01:00:5E:00:00:01 to 01:00:5E:00:00:FF forwarded on all ports whether or not IGMP joins have been sent?

Q6 Does your switch support forwarding to ports on which IP multicast routers are attached in addition to the ports where IGMP Joins have been received?

Q7 Is your IGMP snooping functionality fully implemented in hardware?

Q8 Is your IGMP snooping functionality partly software implemented?

Q9 Can topology changes (for example spanning tree configuration changes) be detected by the IGMP snooping functionality so that for example new queries can be sent or tables can be updated to ensure robustness?

The answers were:

	Switch Vendor					
	1	2	3	4	5	6
Q1 Join aggregation	x	x	x		x	x
Q2 Layer-2 forwarding	x	x	x	x	(1)	
Q3 Layer-3 forwarding	(1)		(1)		(1)	x
Q4 224.0.0.X aware	(1)	x	(1)	(2)	x	x
Q5 01:00:5e:00:00:XX aware	x	x	x	(2)	x	x
Q6 Mcast router list	x	x	x	x	x	x
Q7 Hardware implemented						
Q8 Software assisted	x	x	x	x	x	x
Q9 Topology change aware	x	x	x	x		(2)

x Means that the answer was Yes.
(1) In some products (typically high-end) Yes, in others No.
(2) Currently no, but will be really soon.

8. IETF IPR Statement

9. Acknowledgements

We would like to thank Martin Bak, Les Bell, Yiqun Cai, Ben Carter, Paul Congdon, Toerless Eckert, Bill Fenner, Brian Haberman, Edward Hilquist, Hugh Holbrook, Kevin Humphries, Suzuki Shinsuke, Jaff Thomas and Rolland Vida for comments and suggestions on this document.

Furthermore, the following companies are acknowledged for their contributions: 3Com, Alcatel, Cisco Systems, Enterasys Networks, Hewlett-Packard, Vitesse Semiconductor Corporation. The ordering of these names do not correspond to the column numbers in the response table.

10. Revision History

This section, while incomplete, is provided as a convenience to the working group members. It will be removed when the document is released in its final form.

draft-ietf-magma-snoop-05.txt: January 2003 Changes in wording of IGMP forwarding rule 6) and Data forwarding rule 7). Corrections in the references section.

Apart from above, no substantial changes has occured in the document. Several editorial changes, however, have been made to comply with the rfc editors requirements:

References splitted in normative and informative sections.

Abstract shortened.

Changed all occurances of MUST, MAY etc. to lowercase to reflect that this is not a standards track document.

Sections moved around so they appear in the required order.

draft-ietf-magma-snoop-04.txt: November 2002 Editorial changes only.

draft-ietf-magma-snoop-03.txt: October 2002

IGMP Forwarding rules:
Add references to and become consistant with the current IGMP proxy draft,

Unrecognized IGMP packets should not be ignored because "mbz" fields are not zero since packets from future versions are expected to maintain consistency.

Corrections related to IGMP Querier election process.

Add clarification to how lists of router ports may be assembled.

Data Forwarding rules:
Added discussion of the problems for different IGMP environments in choosing whether to flood or to prune unregistered multicasts.

Added refinements for how to handle NON-IPv4 multicasts, to keep IGMP-snooping functionality from interfering with IPv6 and other multicast traffic. Any filtering for non-IPv4 multicasts should be based on bridge behavior and not IGMP snooping behavior.

IGMP snooping related problems:
Fixed description of interoperability issues in environments with v3 routers and hosts, and v2 snooping switches.

Added discussion of the IGMPv3 "include source" and "exclude source" options, and the inability to support them on shared

segments.

IPv6 Considerations:
Clarifications regarding address ranges FF00::, FF01:: and all

hosts FF02::1 in relation to data forwarding.

draft-ietf-magma-snoop-02.txt: June 2002

Status section removes document history; moved into this section instead.

Introduction restores text from the -00 revision that describes snooping and its goals

IGMP flooding rules eased, allowing management option to broaden beyond "routers only".

Removed a should/MAY inconsistancy between IPv4 Forwarding and IPv6 processing of checksums.

IGMP Forwarding Rules: clarify text describing processing of non-zero reserved fields.

Data Forwarding Rules, item 3 is changed from "MUST forward to all ports" to "MAY"; item 4 default changes from "MUST" to "should use network addresses".

Added two sets of additional responses to the questionnaire and text indicating that responses don't cover all products.

Removed (commented out) description of IPR issues: IESG is aware of them.

draft-ietf-magma-snoop-01.txt: January 2002

Extensive restructuring of the original text.

draft-ietf-idmr-snoop-01.txt: 2001

Added several descriptions of cases where IGMP snooping implementations face problems. Also added several network topology figures.

draft-ietf-idmr-snoop-00.txt: 2001

Initial snooping draft. An overview of IGMP snooping and the

problems to be solved.

11. Author's Addresses

Morten Jagd Christensen
Thrane & Thrane
Lundtoftegaardsvej 93 D
2800 Lyngby
email: mjc@tt.dk

Karen Kimball
Hewlett-Packard
8000 Foothills Blvd.
Roseville, CA 95747
email: karen.kimball@hp.com

Frank Solensky
Bluejavelin, Inc.
3 Dundee Park
Andover, MA 01810
email: fsolensky@bluejavelin.net

Table of Contents