MAGMA Working Group
Internet Draft
June 2002
Expiration Date: December 2002

M. Christensen
morten@jagd-christensen.com
F. Solensky
Premonitia

IGMP and MLD snooping switches
<draft-ietf-magma-snoop-02.txt>

Status of this Memo

Abstract

This memo describes the requirements for IGMP and MLD snooping switches. The requirements for IGMPv2 snooping switches are based on best current practices. IGMPv3 and MLDv2 snooping are also cov¡ered in this draft although we are not aware of any such implemen¡tations at the time of writing. Areas which are of relevance to IGMP and MLD snooping switches, such as link layer topology changes and Ethernet specific encapsulation issues are also considered.

Interoperability issues that arise betweed different versions of IGMP are not discussed in this document. Interested readers are directed to [IGMPv3] for a thorough description of problem area.

This document is intended as an accompanying document to the IGMPv3 and MLDv2 specifications.

11.. IInnttrroodduuccttiioonn

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.

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 num¡ ber of attached nodes.

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 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 net¡ work where no node has expressed interest in receiving packets addressed to the group address. This is in contrast to normal switch behaviour 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 information.

The requirements presented here is based on the following informa¡tion sources: The IGMP specifications [RFC112][RFC2236][IGMPv3],

vendor supplied technical documents [CISCO], bug reports [MSOFT], discussions with people invovled in 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 to IPv4 only. For IPv6 we must use MLD 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 is a difference compared to IGMP.

22.. IIGGMMPP SSnnoooopppiinngg RReeqquuiirreemmeennttss

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

22..11.. FFoorrwwaarrddiinngg rruulleess

The IGMP snooping functionality is here separated in a control sec¡tion (IGMP forwarding) and data section (Data forwarding).

22..11..11.. IIGGMMPP FFoorrwwaarrddiinngg RRuulleess

1) A snooping switch SHOULD forward IGMP Membership Reports only to those ports where multicast routers are attached. Alterna¡tively stated: A snooping switch SHOULD NOT forward IGMP Mem¡bership 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 member¡ ship 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 mes¡ sages to be processed by network monitoring equipment such as packet analysers 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 follow¡ ing ways:

a) This list SHOULD be built using IGMP Multicast Router Dis¡ covery [MRDISC] by the snooping switch sending Multicast Router Solicitation messages on its own. 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 where the address is not 0.0.0.0.

c) A list of ports configured by management as described in the previous step.

2) IGMP snooping switches MAY implement "proxy-reporting" in which reports received from downstream hosts are summarized and used to build internal membership states as described in [PROXY]. An IGMP proxy-reporting switch would then report its own state in response to upstream queriers. If the switch does not alreay have an IP address it SHOULD use the address 0.0.0.0 as source in these reports.

An IGMP proxy-reporting switch may act as Querier for the down¡ stream 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 identi¡fied 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; however, these messages MUST NOT be included the election process so that a snooping switch does not elected over an active router.

3) The switch that supports IGMP snooping MUST flood all unrecog¡nized 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 particular, messages where any reserved fields in the IGMP header are non-zero MUST NOT be subject to "normal" snooping since this could indicate an incompatible change to the IGMP message format.

4) An IGMP snooping switch SHOULD be aware of link layer topology changes. Following a topology change the switch SHOULD initi¡ate the transmission of a General Query on all ports in order to reduce network convergence time.

5) An IGMP snooping switch MUST NOT make use of information in IGMP packets where the IP or IGMP headers have checksum or intregity errors. The switch SHOULD NOT flood such packets but if it does, it SHOULD take some note of the event (i.e.: incre¡ment a counter). These errors and their processing are further discussed in [IGMPv3], [MLD] and [MLDv2].

22..11..22.. DDaattaa FFoorrwwaarrddiinngg RRuulleess

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 does not Join IP multicast addresses in this range before sending or listening to IP multicast. 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.

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 member¡ ship and multicast router tables in software and then "merge" these tables into a current forwarding cache.

3) If a switch receives a non-IGMP 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 config¡ uration option that suppresses this restrictive operation and forces flooding of unregistered packets on all ports.

4) 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.

5) 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 dis¡ carded.

22..22.. IIGGMMPP ssnnooooppiinngg rreellaatteedd pprroobblleemmss

A special problem arise in the network consisting of IGMPv3 routers as well as IGMPv2 and IGMPv3 hosts interconnected by a IGMPv2

snooping switch. IGMPv3 has a mechanism to fall back to IGMPv2 when receiving IGMPv2 membership reports. This means that the net¡ work will converged on IGMPv2 eventually. However, the convergence time will lead to supression of v3 Hosts for several minutes.

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

33.. IIPPvv66 CCoonnssiiddeerraattiioonnss

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 Mul¡ ticast 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 32 bits of the 128 bit DIP addresses into 48 bit DMAC addresses [IPENCAPS].

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

The IPv6 packet header does not include a checksum field. Never¡ theless, 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 IGMP 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 redirect pack¡ ets to the CPU. If this was the case the CPU queue assigned for MLD would potentially be filled with non-MLD related packets. Fur¡ thermore ICMPv6 packets destined for other hosts would not reach their destination. A solution is either to require that the snoop¡ ing switch looks further into the packets or to be able to detect a multicast DMAC address in conjunction with ICMPv6. The first solu¡ tion 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 unflexible for the introduc¡ tion of new message types. The second solution introduces the risk of new protocols, which are not related to MLD but uses ICMPv6 and multicast DMAC addresses wrongly being identified as MLD. It is suggested that solution one is the preferred if the switch is capa¡ ble 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 figure 5. 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 ambigu¡ ity for the time being but it should be noted that the allocation policy may change in the future.

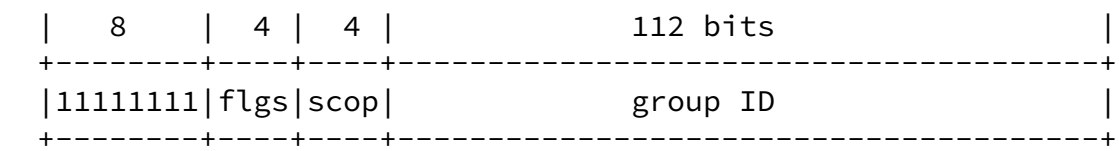

Figure 5

In the case of IPv6 forwarding can be made on the basis of DMAC

addresses in the forseable future.

Finally, we mention the reserved address range FF02::/96. 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.

44.. SSeeccuurriittyy CCoonnssiiddeerraattiioonnss

Security considerations for IGMPv3 are accounted for in [IGMPv3]. The introduction of IGMP snooping switches adds the following con¡ siderations 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 other¡ wise. 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 sug¡ gested Multicast Security (MSEC) working group or similar is com¡ pleted or at least has matured.

55.. IIGGMMPP QQuueessttiioonnnnaaiirree

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 hard¡ ware/software so that only one Join is forwarded to the querier?

Q2 Is multicast forwarding based on MAC addresses? Would data¡ grams 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 unal¡ tered 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 multi¡ cast routers are attached in addition to the ports where IGMP Joins have been received?

Q7 Is your IGMP snooping functionality fully implemented in hard¡ ware?

Q8 Is your IGMP snooping functionality partly software imple¡ mented?

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 real soon.

66.. RReeffeerreennccees

[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/pub¡ lic/473/22.html

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

[IGMPv3] Cain, B., "Internet Group Management Protocol, Version 3", draft-ietf-idmr-igmp-v3-11.txt, May 2002.

[IPENCAPS] Crawford, M., "Transmission of IPv6 Packets over Ether¡

net 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-v2-03.txt, June 2002.

[MRDISC] Biswas, S. "IGMP Multicast Router Discovery", draft-ietf-idmr-igmp-mrdisc-08.txt, January 2002.

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

[PROXY] Fenner, B. et al, "IGMP-based Multicast Forwarding (IGMP Proxying)", draft-ietf-magma-proxy-02(?).txt.

[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.

77.. AAcckknnoowwlleeddggeemmeennttss

We would like to thank Martin Bak, Les Bell, Yiqun Cai, Paul Cong¡ don, Toerless Eckert, Bill Fenner, Brian Haberman, Edward Hilquist, Hugh Holbrook, Kevin Humphries, Karen Kimball and Jaff Thomas 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 necessarily correspond to the column numbers in the response table.

88.. RReevviissiioonn HHiissttoorryy

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-01.txt: January 2002

Extensive restructuring of the original text.

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

Status section removes document history; moved into this sec¡ tion 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.

The next revision:

In the interest of getting this version of the draft released before the deadline (less than seven hours from the moment this paragraph is being typed), we briefly summarize some of the comments on the previous version that need to be included in the next one. We believe that other comments have been addressed in this draft; please let the authors know if this they have either not been included or need to be corrected.

IGMP Forwarding rules:

Add a reference to and become consistant with the next revision of the IGMP proxy draft,

In item 'b': include a description on how the switch determines that a Query came from the router and not another switch. Is there some way to make this distinc¡ tion beyond the source address?

Proxy reporting: further analysis of the impact on the

election process when using 0.0.0.0 as the source address in membership report messages. Also consider the case where the switch assumes the role of Querier when no routers are detected and forfeits the role as soon as one is announced.

Include some discussion about how entries are to be aged from the list, perhaps similar to spanning tree algorithm for unicast MAC address processing.

Data Forwarding rules:

Link-local range to mention the problem is due to routing protocols not sending Report Messages for their respec¡ tive multicast addresses.

Expand discussion of non-IGMP packet forwarding for data that matches an IGMPv3 record. Do snooping switches add

intelligence to recognize SSM versus ASM groups?

IPv6 Considerations:

Is having MLD a subset of ICMPv6 an issue? Should MLDv2 be a separate protocol?

Add reference to ICMPv6 specification for message pro¡ cessing rules.

99.. AAuutthhoorr''ss AAddddrreessssseess

Morten Jagd Christensen
email: morten@jagd-christensen.com

Frank Solensky
Premonitia, Inc.
1 Nanog Park
Acton, MA 01720
email: fsolensky@premonitia.com

TTaabbllee ooff CCoonnttteennttss