RIVERSTONE NETWORKS
ADVANCED TECHNICAL PAPER SERIES

Internet Group Management Protocol

Michael Gibbs, Riverstone Networks

ABSTRACT

A detailed examination of the IGMP protocol, this paper covers versions 1, 2 and 3 covering protocol mechanics and is designed to give the reader technical knowledge of multicast network design. This paper assumes a fundamental understanding of IP multicast routing but does not provide router configurations. For example configurations please see the Multicast Support Pages in the Riverstone online Advanced Technical Documentation Library.

Riverstone Networks, Inc.
5200 Great America Pkwy, Santa Clara, CA 95054 USA
(877) 778-9595, (408) 878-6500, www.riverstonenet.com

Version 1.2, January 27, 2003

Introduction

Like the Internet itself, Multicast Routing has evolved considerably from its academic roots to current commercial applications. As commercial applications became apparent, resources dedicated to the enhancement and extention of this useful protocol have increased. Early comercial applications included the financial industry which had to stream large amounts of mission critical real time data to large numbers of end users. Currently used many enterprises and carrier environments, IGMP is also seeing use in Video over DSL, a particularly challenging application requiring special handling.

IGMP allows hosts to communicate their interest and desire to receive data destined to a specific multicast group. This desire is communicated to the local upstream router, which then uses this information to build, prune or graft multicast distribution trees. Although there are currently three versions of IGMP, with each successive generation offering incrementally better efficiency and functionality. While the most widely used and deployed one is version 2, this paper will cover all 3 versions.

Internet Group Membership Protocol Version 1

IGMP version 1 was the original verison of IGMP. This early effort was implemented on a variety of operating systems including many versions of Unix as well as Windows 95 and NT4. Although no longer used for new deployments, IGMP V1 has a significant history and an understanding of this protocol can be help troubleshooting complex multicast networks in the field such as Video over DSL. The Riverstone RS platform no longer supports IGMP V1 although it does provide IGMP V1 backwards compatibility for downstream hosts.

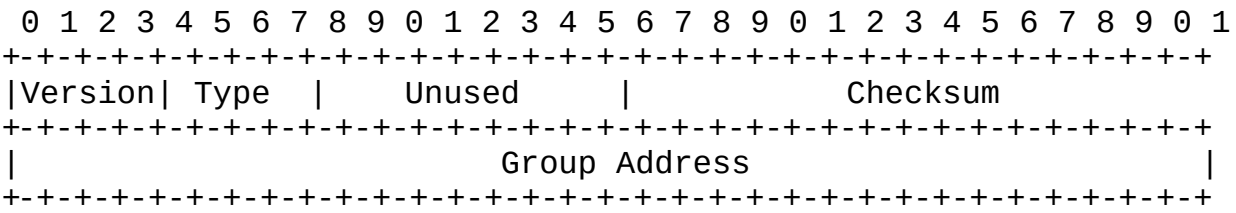
IGMP V1 Message Types

IGMP version 1 is a rather simple protocol. There are essentially 2 major messages communicated between hosts and routers. There are membership queries and membership reports.

IGMP membership queries enable the router to solicit membership information from hosts on directly attached interfaces. IGMP membership reports allow the host to inform its local router of its desire to receive a given multicast group.

The IGMP message format and the meanings of the fields in the messages can be seen below in figure 1.

Figure 1



Version

RFC 1112 specifies version 1 of IGMP. Version 0 was specified in RFC-988 and is obsolete.

Type

There are two types of IGMP message of concern to hosts:

1 Membership Query
2 Membership Report

Unused

This is an unused field. It is zeroed when sent and ignored when received.

Checksum

The checksum is the 16-bit one's complement of the one's complement the 8-octet IGMP message.

Group Address

A Host Membership Query message, the group address field is zeroed when sent.

A Host Membership Report message, the group address field is the IP address of the group being joined.

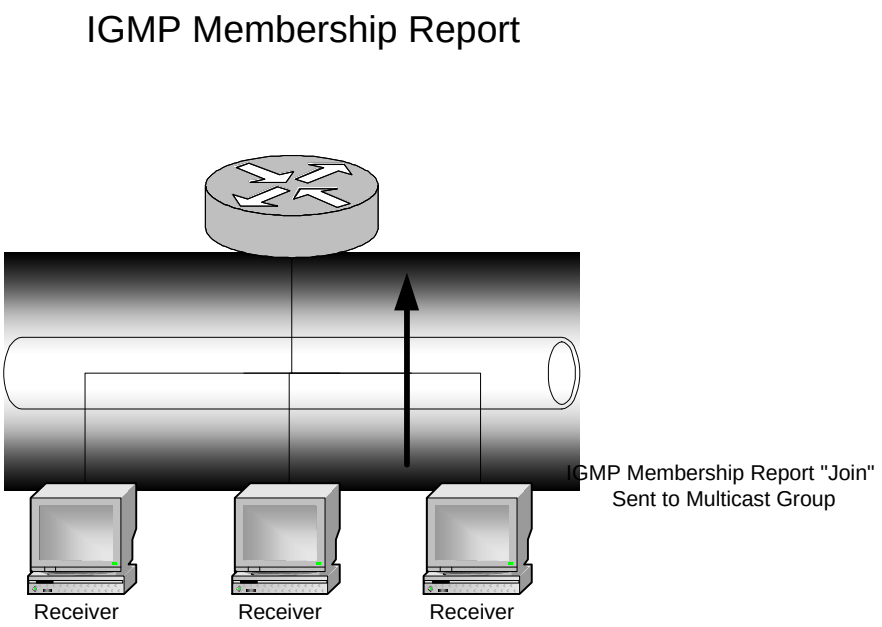
IGMP V1 Join Process

When a host desires to receive a given multicast stream it sends an IGMP membership report message to the multicast group it wishes to receive. This host membership report is frequently referred to as an IGMP join. This message has a TTL of 1 so it will not be forwarded by the local router. Since this message is sent to the multicast group being joined all other members of this multicast

group can hear this message. This is used in order to perform report suppression in response to an IGMP query. This will be discussed in the query section.

Figure 2

Figure 2 shows the IGMP membership report process. In this example a host sends a membership report for the group it wishes to join.



IGMP V1 Query Process

The router issues queries to all the multicast hosts that it is supporting as an audit of the groups the multicast recievers desire.. All hosts see this query and one host per multicast group must respond to the query. It is essential to note that in a true routed environment without IGMP snooping only one host per multicast group needs to respond. When a host receives a query it must respond within the host response time (10 seconds) of its group membership unless it hears a response from another group member in this time. This 10 second host response time was designed to reduce the amount of IGMP burstiness on a LAN segment.

Figure 3
In figure 3 we see how the router sends a general query at every query interval to audit which interfaces have receivers and which multicast groups they are subscribed to.

IGMP Membership Query

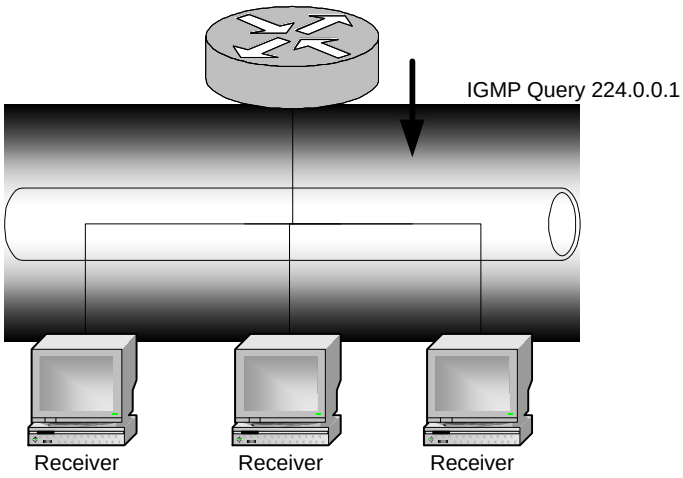
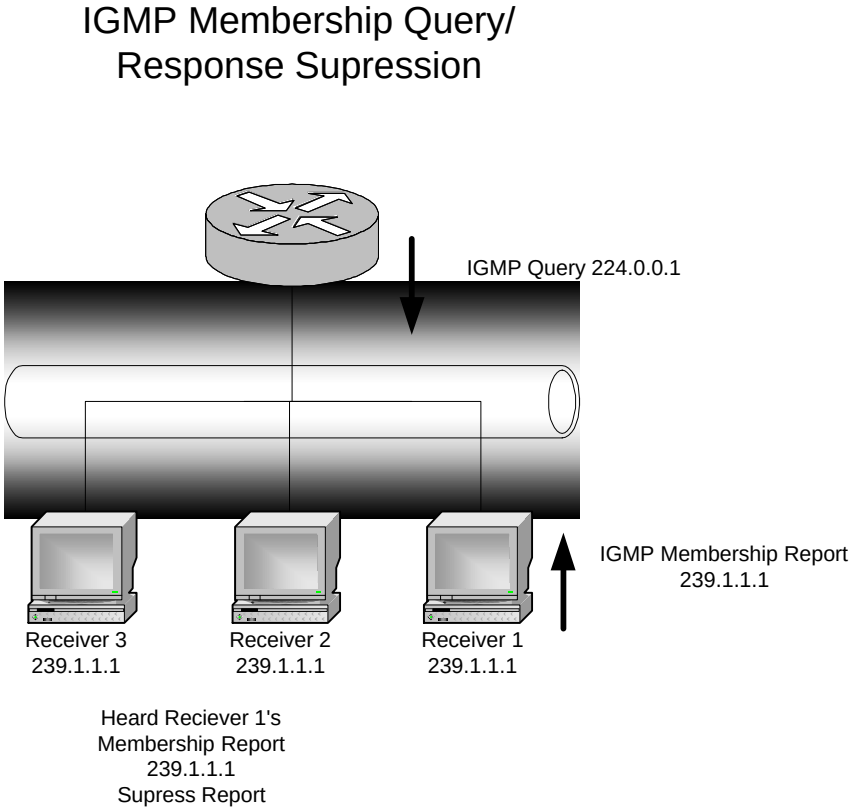


Figure 4

In Figure 4 we see the IGMP report suppression mechanism. This was designed to decrease the IGMP traffic on a subnet. It does not apply to IGMP snooping.



IGMP V1 Leave Process

One very important limitation of IGMP version one is that it did not have the provisioning for a host to explicitly tell the router of its desire to leave a multicast group. Instead the host “informed” the router of its lack of interest in receiving a multicast stream by not responding to IGMP membership queries from the router. Therefore leave latency was high in IGMP version 1. Leave latency was calculated by IGMP query interval * robustness (number of queries) added to host response time.

One other limitation of IGMP version 1 was that it did not have the ability to elect a querier in the protocol itself. On a given LAN segment there can only be 1 IGMP querier. All other IGMP routers must be backup or non-queriers. IGMP version 1 did not have the ability to perform querier election. IGMP version 1 relied on the multicast routing protocol to perform querier election.

Internet Group Membership Protocol Version 2

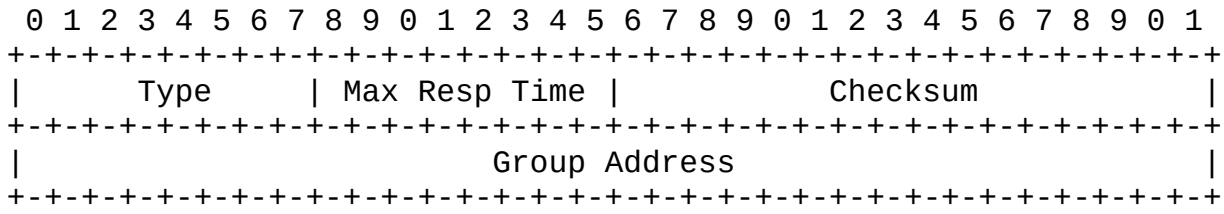
IGMP version 2 is the default standard of IGMP. It is implemented on Microsoft Windows 98, 2000, ME and most modern Linux and Unix versions. Additionally, it is included on most IGMP version 1 hosts that were patched for Y2K upgrades. IGMP version 2 has significant enhancements over its predecessor IGMP V1, the most significant of which is the expedited leave mechanism. The RS platform fully supports the IGMP V2 standard.

IGMP V2 Message Types

IGMP version 2 adds two new messages in addition to the two messages present in IGMP version 1. IGMP V2 adds a leave group message as well as a group specific query. In addition there are subtle changes to the general query and host membership report. This section will cover the message types and the fields inside these messages.

Figure 5

Figure 5 shows an IGMP version 2 message and the respective fields in the message types.



Type

There are three types of IGMP messages. They are defined by the Hex value in the type field.

0x11 = Membership Query

There are two types of Membership Query messages:

General Query. Audit every query interval to find out which hosts are members of which groups.

Group-Specific Query, sent in response to a IGMP leave message. Used to learn if a particular group has any remaining members on the attached network.

Note: Both of these messages have the value 0X11. These two messages are differentiated by the Group Address.

General Query messages are addressed to 224.0.0.1

Group-Specific Query messages are sent to the group for which the router received a leave group message.
0x16 = Version 2 Membership Report

0x17 = Leave Group

There is an additional type of message, for backwards-compatibility with IGMPv1:

0x12 = Version 1 Membership Report

This document refers to Membership Reports simply as "Reports". When no version is specified, the statement applies equally to both versions.

Max Response Time

The Max Response Time field is used in Membership Query messages, and specifies the maximum allowed time a host should wait before sending a report. The unit of measurement for this field is in1/10 second. In all other messages, it is set to zero by the sender and ignored by receivers.

Tuning this setting allows IGMPv2 routers to adjust "leave latency" and IGMP burstiness on a LAN segment.

Checksum

A 16-bit one's complement of the of the whole IGMP message (the entire IP payload).

Group Address

In a Membership Query message, the group address field is set to 0.0.0.0 on general queries and set to the group being queried when sending a Group-Specific Query.

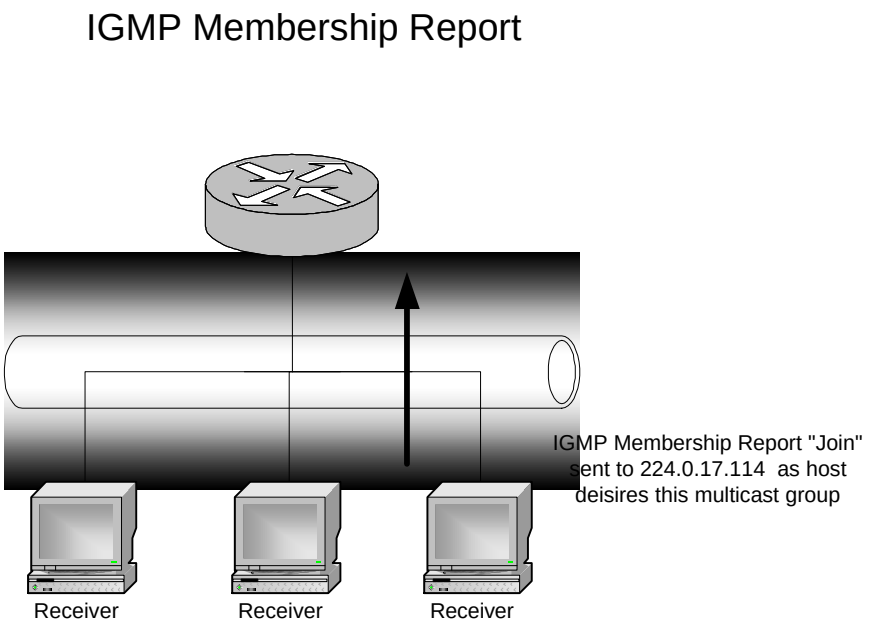
Host Membership Reports or Leave Group messages, the group address field holds the IP address of the group being joined or left.

IGMP V2 Join Process

When a host desires to receive a given multicast stream it sends an IGMP membership report message to the multicast group it wishes to receive. Like in IGMP version 1 this message has a TTL of one so it will not be forwarded by the local router. Since this message is sent to the multicast group being joined all other members of this multicast group can hear this message.

Figure 6

Figure 6 shows the IGMP join process. Below we see a host joining the desired multicast group with an IGMP membership report. Additionally, we see this message in debug format on the as it appears on the RS console.



11-15 18:09:49 IGMP RECV 10.50.0.182 -> 224.0.17.114 New Host Membership Report: group 224.0.17.114

IGMP V2 Query Process

The IGMP query process is similar in IGMP version 2 to version 1 but it has some enhancement. Enhancements typically increase complexity. This is no exception. There can only be one querier per LAN segment. IGMP version 2 has an election mechanism to elect the querier on a LAN segment. Effectively when a router that has IGMP enabled is added to a LAN segment it assumes it is the querier. It begins sending IGMP queries. If a router hears a superior query (a querry from a router at a lower IP address) it will suppress sending queries for a specified until it no longer hears the other querier. At that point it will again assume it is the querier. This is known as the other querier present query interval. This value is calculated from ((the Robustness Variable) times (the Query Interval)) plus (one half of one Query Response Interval).

Figure 7

In figure 7 we see multiple IGMP speaking routers on a subnet. IGMP V2 specifies that only one router can be a querier for a given multicast group. There is a deterministic election based upon IP address. In this example we see how the election occurs.

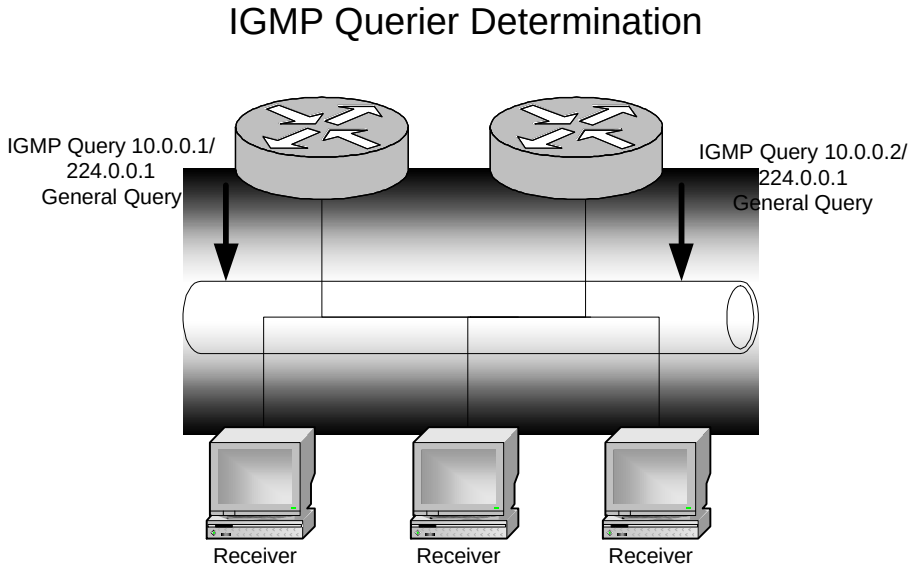
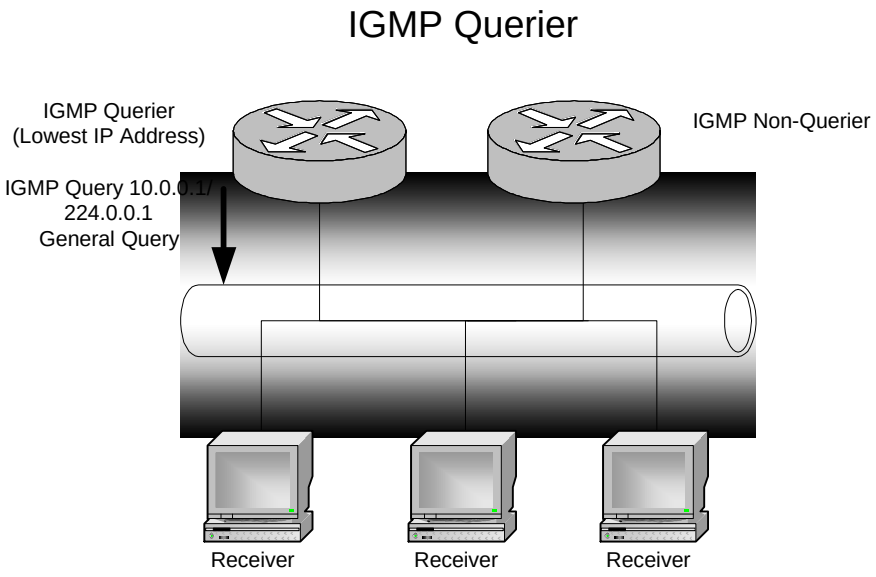


Figure 8

In the below example we see that a router is elected to be the querier based upon its lower IP address. The other router will function in a backup manner and be the non-querier.



Another enhancement in the IGMP version 2 query process is the ability for the router to tell the hosts how long they have until they need to respond to a query. This time is called Max-Response-Time. As mentioned previously this enables the network administrator to tune IGMP burstiness.

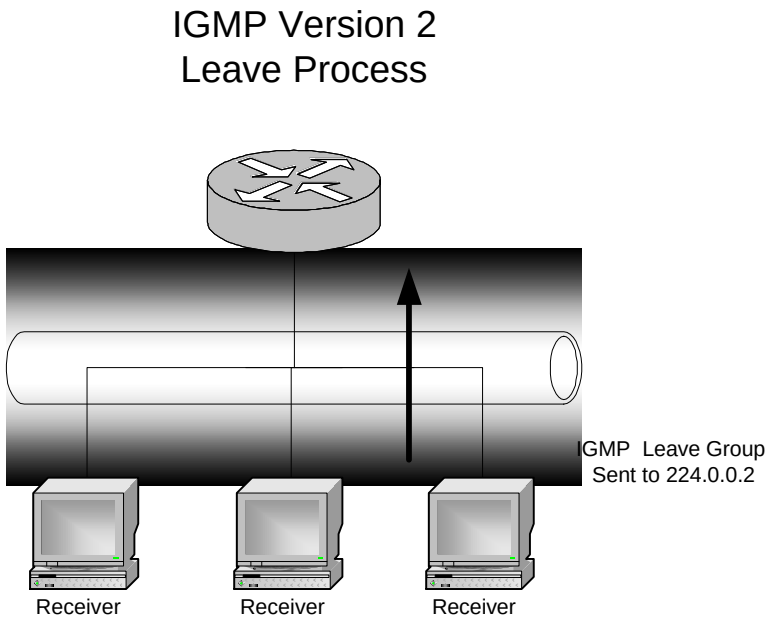
IGMP V2 Leave Process

The IGMP Version 2 leave process has significant enhancements over IGMP version 1. In this implementation the host explicitly tells the router when it no longer desires to receive the multicast data. This significantly reduces leave latency. As we discuss this leave process it is important to remember that with the exclusion of IGMP snooping the router does not know how many hosts are joined to a multicast group. The router knows of only one receiver per multicast group. This is due to the report suppression mechanism described above. The leave process works as follows. The host that does not desire to receive the multicast data anymore sends a leave-group message to the multicast group 224.0.0.2 (All multicast routers). The router performs an audit to determine if there are any other members of this multicast group before cutting offthe flow of data to this group. The router performs this by sending a group specific query to

the group in question. This query has a timeout of one second. If a host replies to this query the router will continue sending data to this group. If no host responds the router will send 1 additional group specific query. If no host replies to this group specific query, the router will stop sending multicast packets to this multicast group on this interface.

Figure 9

In figure 9 we see a host tell a router that it no longer wishes to receive data for a given multicast group. Additionally, we see this message in debug format as it appears on the RS command console.



11-15 18:10:29 IGMP RECV 10.50.0.182 -> 224.0.0.2 Host Membership Leave: group 224.0.17.114

Figure 10

In the figure below we see a group specific query in response to a leave group message. This process is repeated two times by default before determining it is safe to delete a group membership entry.

Group Specific Query

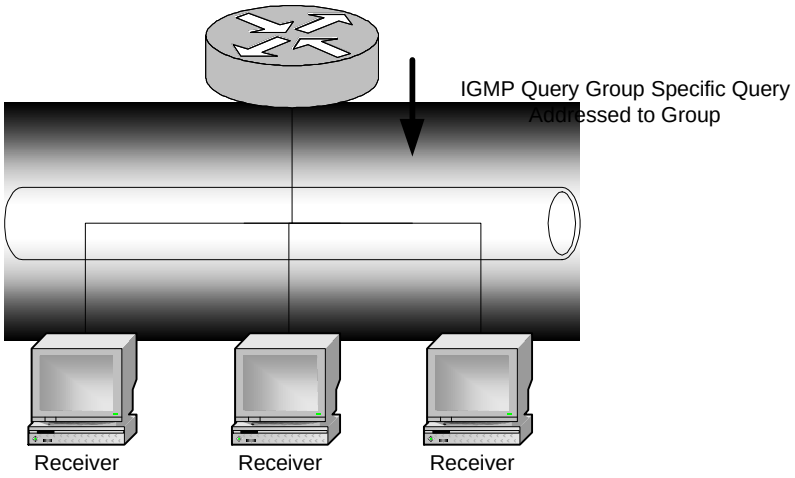
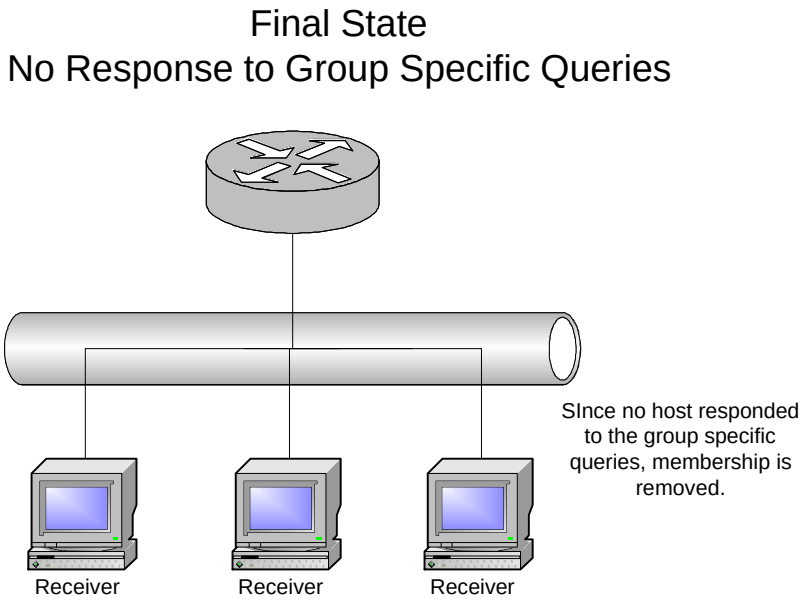


Figure 11

In figure 11 no host has responded to the group specific query. Because no host has responded to the group specific query the router has removed it membership entry on the interface. This will signal the multicast routing protocol to prune if necessary.



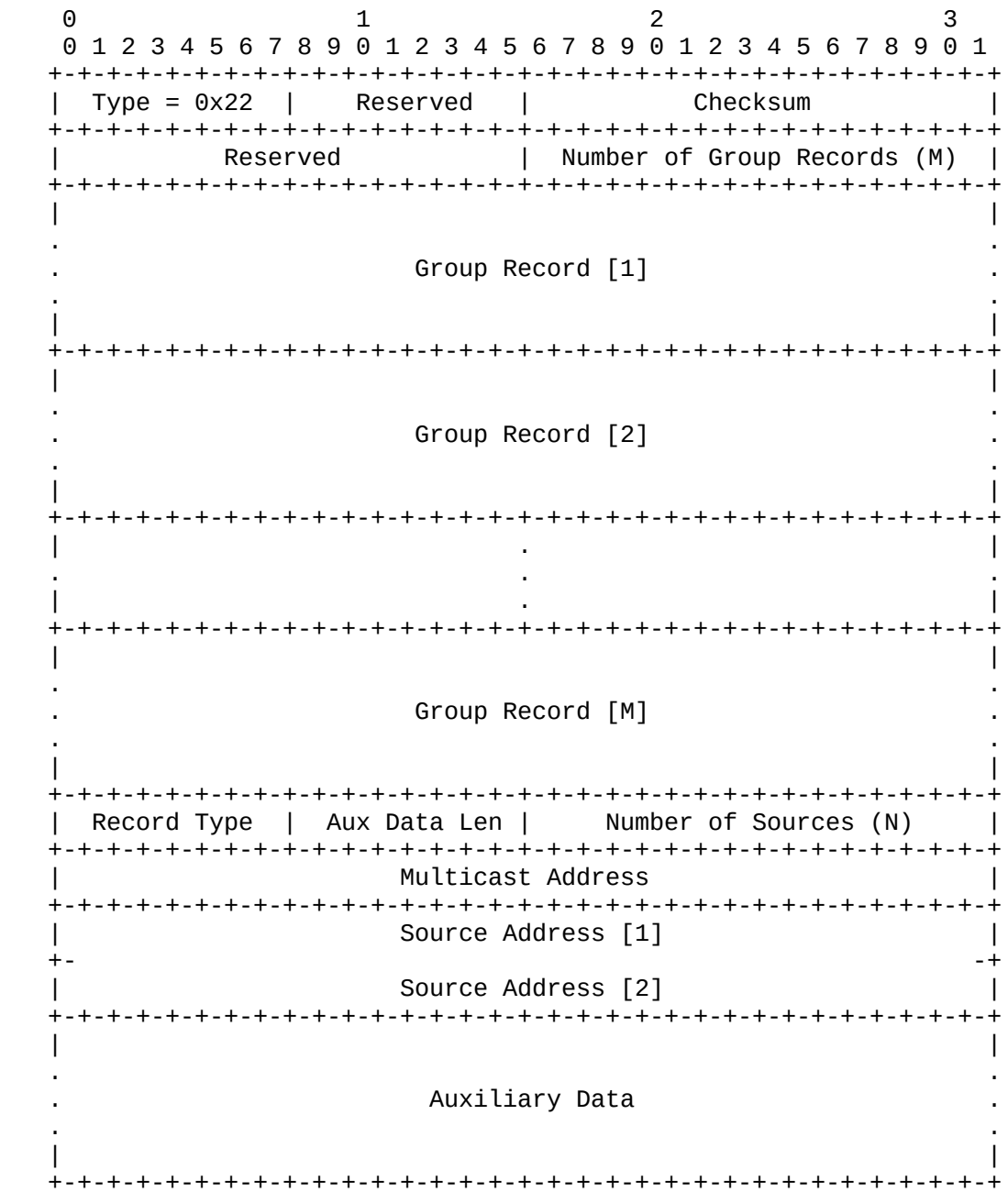
Internet Group Membership Protocol Version 3

IGMP version 3 builds upon prior versions by adding provisions for a host to not only specify a group but also specify the source publishing to the group, which totally changes the paradigm for IP multicast routing. IGMP version 3 enables new protocols like PIM source specific multicast. Additionally, IGMP version 3 increases security on multicast networks because the receiver can specify the source of the multicast group and not only the group it desires to receive. If that alone was not enough IGMP version 3 has the ability to almost eliminate the IP multicast address allocation issues. This is because IP multicast is now not only limited to the 224/4 space but the entire unicast address spectrum. IGMP version 3 is currently implemented on Microsoft Windows XP and it is possible to get an IGMP V3 stack for some newer versions of Linux. Supported on the RS platform in future releases of ROS software,IGMP version 3 was recently standardized in RFC 3376.

IGMP V3 Message Types

This section below diagrams an IGMP V3 packet. Additionally, it explains what the meanings are of each field.

Figure 12


0 1 2 3
0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
Type = 0x22
Reserved
Checksum
Reserved
Number of Group Records (M)
Group Record [1]
Group Record [2]
Group Record [M]
Record Type
Aux Data Len
Number of Sources (N)
Multicast Address
Source Address [1]
Source Address [2]
Auxiliary Data

Type

As in IGMP version 2 the value of the type field determines they kind of IGMP message.

IGMP V3 has 2 message types.

0x11 Membership Query

0x22 Host Membership Report

Additionally, the implementation must recognize version 1 and 2 messages.

Max Resp Code

The Max Resp Code field specifies the maximum amount of time a host has to respond to a general query,

Max response time as in IGMP V2 the time, is represented in units of 1/10 second.

If Max Resp Code < 128, Max Resp Time = Max Resp Code

However, new in IGMP V3,

If Max Resp Code >= 128, Max Resp Code represents a floating-point value. Large values enable IGMP burstiness tuning in the exponential range.

Reserved

The Reserved fields are set to zero.

Checksum

A 16-bit one's complement of the of the whole IGMP message (the entire IP payload).

Number of Group Records

This field specifies how many Group Records are present in this message.

Group Record

Each Group Record is a block of fields containing information regarding the sender's membership in a single multicast group on
the interface where the Report is sent.

Aux Data Len

The Aux Data Len field contains the length of the Auxiliary Data field. It is in units of 32-bit words.

Number of Sources

This field specifies how many source addresses are present in this Record.

Multicast Address

This field contains the IP multicast address group address for which this Group Record pertains.

Source Address

The Source Address [i] fields are a vector of n IP unicast addresses, where n is the value in this record's Number of Sources (N) field.

Auxiliary Data

This field, if present, contains additional information regarding this Group Record. IGMPv3, does not define any auxiliary data.

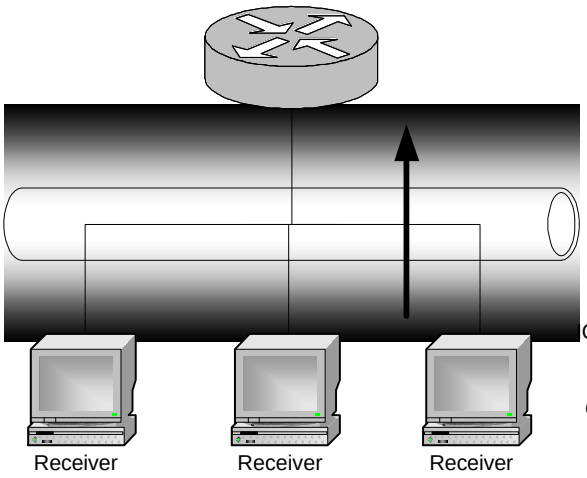
IGMP V3 Join Process

IGMP version 3 is significantly different from any of the other versions previously mentioned. The first significant difference is that IGMP messages are now sent out of band. A host sends its membership report to 224.0.0.22. Since these packets are addressed not to the group joined but an out of band control channel the router will effectively learn all hosts desiring to receive a given multicast group. This means that IGMP version 3 is effectively stateful. This enables new enhancements in terms of expedited leave and simplified IGMP snooping architectures. The IGMP host membership report contains new data in IGMP version 3. This join can specify specific sources in addition to the group being joined. It can specify all sources for group G, specific sources for group G, or all sources but those being excluded from group G.

Figure 13

In the figure below we see an IGMP host membership report. In this case the host only wants to receive a specific source for the multicast group being joined. The host accordingly sends its IGMP V3 join with the include list specifying only source S for Group G.

IGMP Membership Report
Host Desires Only (S 239.1.1.1, G 10.1.1.1)



IGMP V3 Query Process

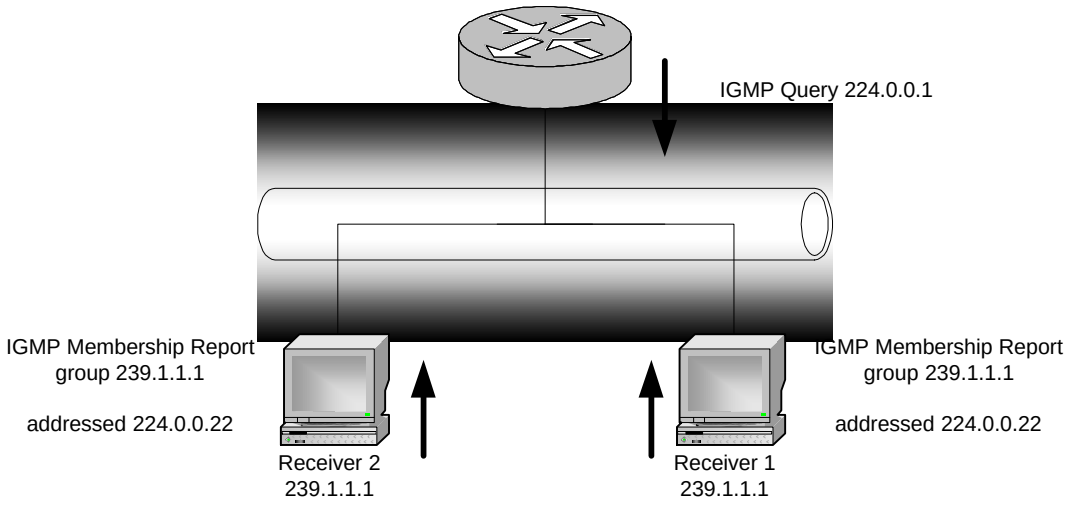
Since IGMP V3 uses an out of band control channel the query process differs from other versions of IGMP. In previous versions of IGMP only one host per multicast group needed to respond to a general query. This was to reduce IGMP traffic. Unfortunately this was stateless. The router per the IGMP spec had no idea who was a receiver only that there was one receiver per multicast group. In IGMP version 3 queries are still sent every query interval. Like other versions of IGMP the IGMP query is sent to 224.0.0.0 (all multicast hosts). All version 3 hosts will respond to a version 3 query. The hosts will respond with their group of interest to the address 224.0.0.22.

Since this could potentially turn into a lot of IGMP traffic the V3 spec allows for significant tuning of max response time to reduce the degree of IGMP burstiness on a subnet.

Figure 14

In the below figure we see a general query. Note that all hosts respond to the query.

IGMP Membership Query
Version 3



IGMP V3 Leave Process

The IGMP leave process can occur in two ways. It can occur almost identically as it does in version 2 with the exception of the address to which the leave group message is sent. In IGMP V 3 the leave group message is sent to 224.0.0.22 unlike version 2 of IGMP where the leave group message is sent to 224.0.0.2.

Therefore the router will send a group specific query in response to a leave group just to confirm that there are no other members that desire a given multicast group. (See IGMP version 2 leave group figures for clarification. Figures 9, 10 and 11.)

Additionally it is possible for a receiver to leave a specific source for group G. In this case the router will send a group and source specific query to make sure that there are no other receivers that desire that specific S,G pair. If there were additional receivers for that S,G pair the leave group message would be overridden and the router would continue to forward traffic from source S destined for group G.

Figure 15

In the below figure we see a host communicate its desire to not receive a specific source on the group that it desires to continue to receive. Therefore the host sends a report of its desired state change,

IGMP Membership Membership
State Change

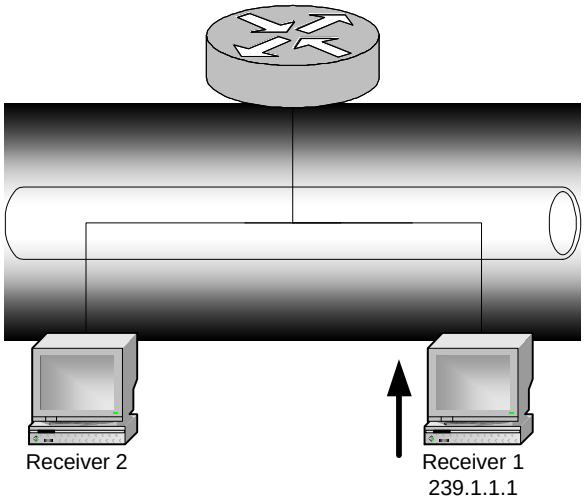
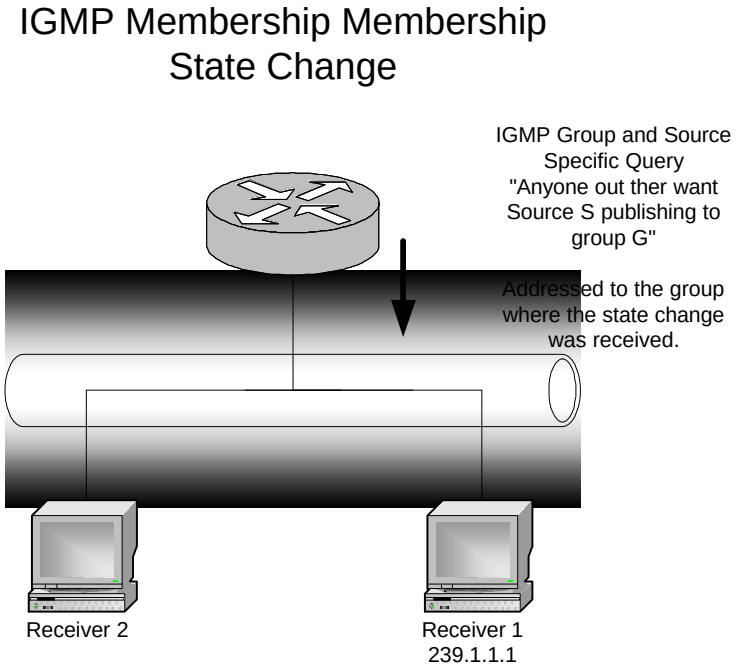


Figure 16

Since the router saw a state change, it will send a Group and Source Specific Query. Like a group specific query it will be sent to the group where the state change was received.



Internet Group Membership Protocol Snooping

This paper has discussed how a host signals a router that it would like to receive a multicast data stream. There is however a significant issue to be overcome in order to make multicasting scaleable. Since a layer 3 address will eventually become a layer 2 MAC address (i.e. 01-00-5e-11-11-11) it is important to remember this is the destination of a multicast group and not a host. (See Riverstone Intro to Multicast Paper for more information on this matter if desired.) A switch by default will forward unknown packets out all ports in a broadcast domain except the source port. Since no port will ever have this MAC address the switch can never learn which port to send this frame. Consequently without some other means the switch will always send multicast frames out all ports except the port for which that the frame was received. This flooding is obviously not desirable and consumes a significant degree of network and system resources. In order to get around this problem most modern switches watch the hosts IGMP membership reports and only forward multicast packets for the groups out of the ports for which the switch snooped an IGMP join message. IGMP snooping has enabled efficient LAN multicasting. The RS platform efficiently performs IGMP snooping in hardware. The RS platform's IGMP snooping does not have any of the layer 2 snooping scalability issues seen in older, lower performance LAN switches. The RS platform even has a version of IGMP snooping for ATM links called port aware. This allows the same efficiency of IGMP snooping on LAN interfaces to be ported to NBMA networks.

Figure 17

In the below figure we see a switch that does not have IGMP snooping. Since there is one receiver on the LAN that is joined to 239.1.1.1 (mac address 01:00:50:01:10:01) the switch has no choice but to forward those packets to all members. This means that all receivers will receive this data regardless on whether or not they desire to receive it.

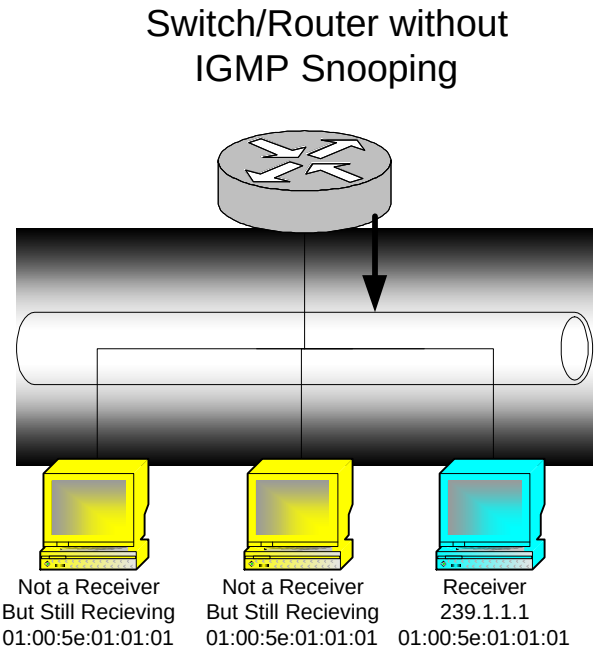


Figure 18

In the below figures we see that the switch performed IGMP snooping and watched for the port on the Switch where the IGMP join was received.

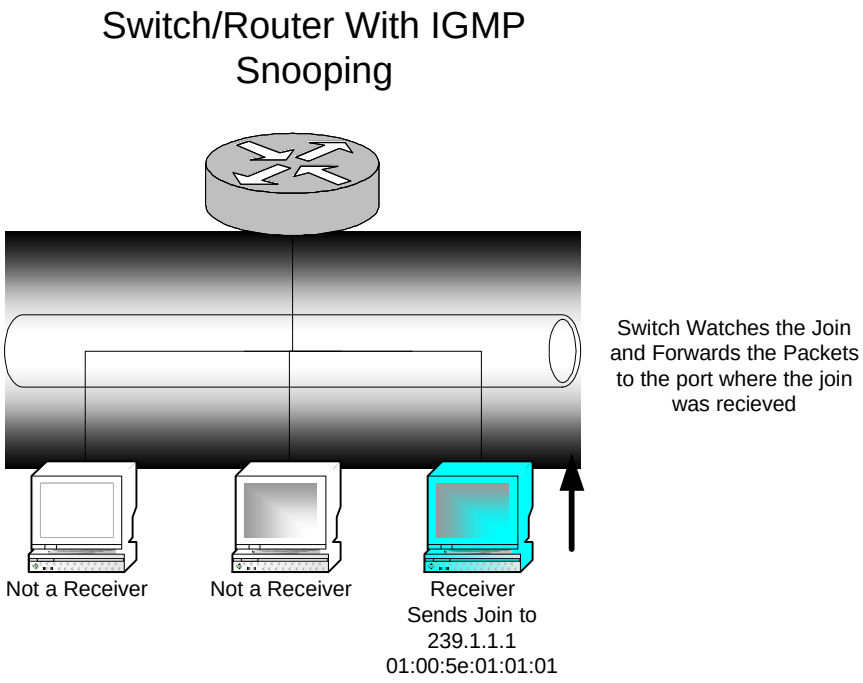
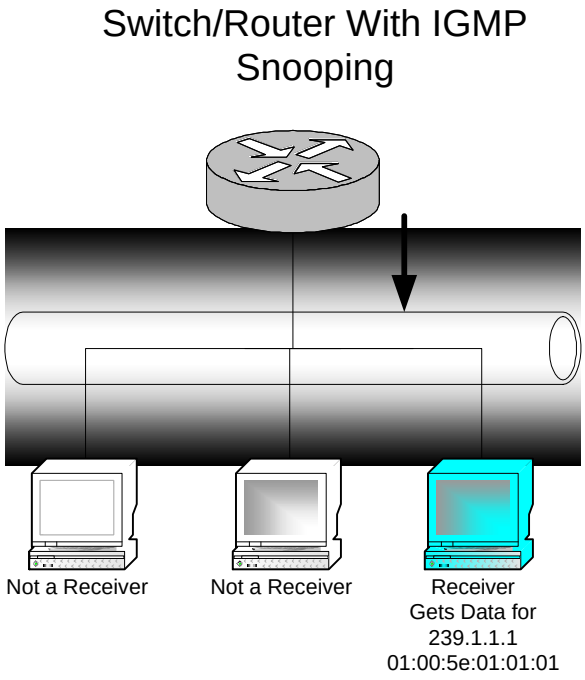


Figure 19

Since the switch knew where the IGMP join was received it knew where to send the multicast packets.



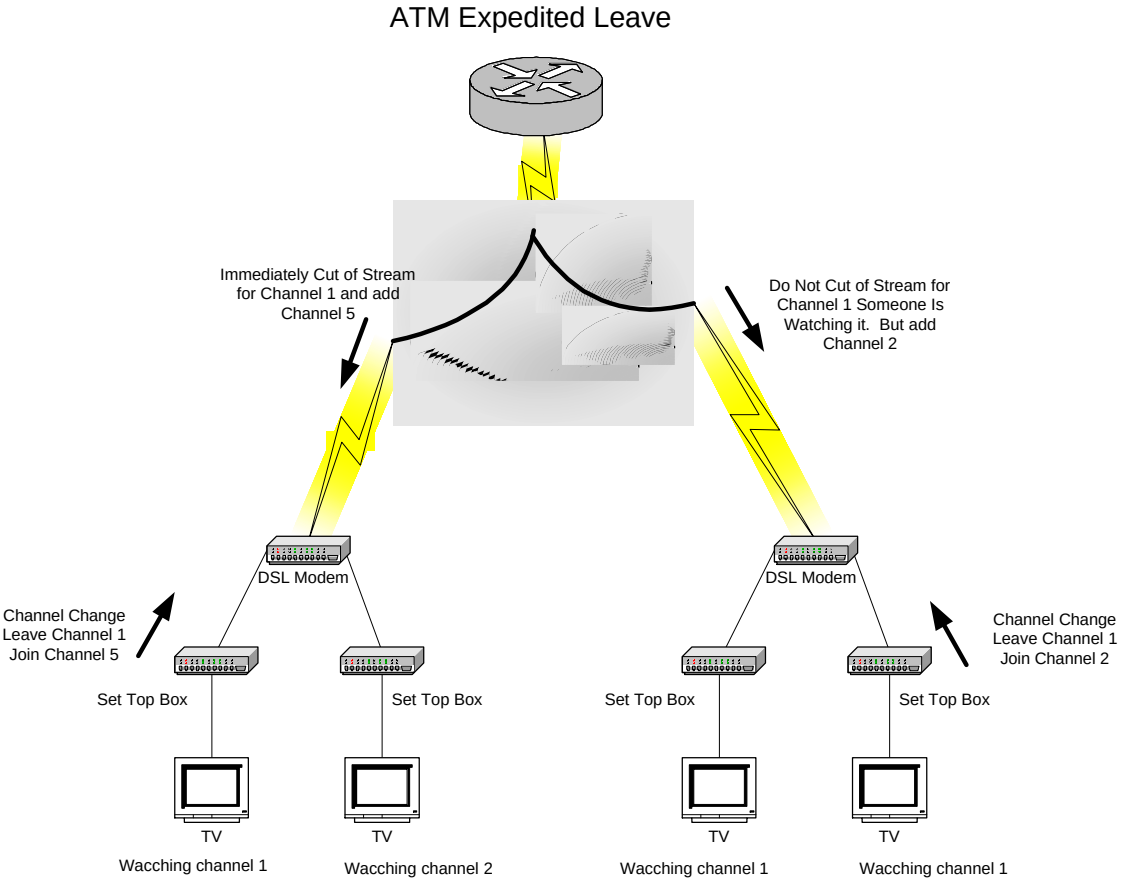
It is necessary to describe one final point about IGMP snooping on the RS. IGMP snooping is technically a layer 2 function as it was designed around which ports a switch would send a multicast packet to. The line between switches and routers is not as clear as it used to be. Accordingly, IGMP snooping is enabled on layer 2 VLANs only on the RS platform. Enabling IGMP on an IP interface on the RS will enable full IGMP functionality as well as IGMP snooping as it is assumed that any time IGMP is added to the RS it is functioning as a Layer 3 switch.

Riverstone Enhanced IGMP functionality on ATM

Since Riverstone has been a pioneer on Video Over DSL it has enhanced IGMP functionality on ATM to deliver high performance video in a bandwidth constrained environment. Recall that a router does not know how many receivers it has for a given multicast group other than there is at least one receiver per multicast group on a given interface. IGMP version 2 is not a stateful protocol. Accordingly there is a two second leave latency with IGMP version 2. In a bandwidth constrained environment there would be two seconds of potential oversubscription every time a user initiated a channel change in a multicast video environment (leave group and join group). The RS will watch the join groups on any receiver in an ATM environment and store them in an IGMP snooping/fast leave table. Since the RS will know exactly which ports have receivers on ATM interfaces the latency associated with group specific queries does not need to occur. For clarification please see the below figures.

Figure 20

In the below example we see a typical Video Over DSL network. The Riverstone IGMP implementation utilizes a proprietary stateful IGMP membership database. Since this implementation knows which users are on which VC's the RS can reduce leave latency from the standard IGMP time of two seconds to approximately 50ms. This enhanced functionality makes the RS ideal for Video Over DSL deployments on ATM.



Summary

IGMP is a protocol used to inform multicast routers of whether or not hosts are interested in receiving traffic bound for a multicast group. The router uses the information obtained from IGMP to determine whether of not to build a multicast tree with its multicast routing protocol. There are currently 3 versions of IGMP, 1 2 and 3, with newwer versions backwards compatible with previous versions. The RS router family currently supports version two of the IGMP protocol and has advanced IGMP over ATM, which supports new and exciting applications like Video Over DSL.

Riverstone Networks, Inc.
5200 Great America Pkwy, Santa Clara, CA 95054 USA
(877) 778-9595 or (408) 878-6500 or www.riverstonenet.com

Version 1.2, January 27, 2003