Network Working Group
Request for Comments: 2226
Category: Standards Track

T. Smith
IBM Corporation
G. Armitage
Lucent Technologies

October 1997

IP Broadcast over ATM Networks

Status of this Memo

Copyright Notice

Abstract

This memo describes how the IP multicast service being developed by the IP over ATM working group may be used to support IP broadcast transmission. The solution revolves around treating the broadcast problem as a special case of multicast, where every host in the subnet or cluster is a member of the group.

An understanding of the services provided by RFC 2022 is assumed.

1. Introduction.

The IETF's first step in solving the problems of running IP over Asynchronous Transfer Mode (ATM) technology is described in RFC 1577 [1]. It provides for unicast communication between hosts and routers within Logical IP Subnets (LISs), and proposes a centralized ATM ARP Server which provides IP to ATM address resolution services to LIS members.

Two classes of IP service were omitted - multicast and broadcast transmissions. Multicasting allows a single transmit operation to cause a packet to be received by multiple remote destinations.

Broadcasting typically allows a single transmit operation to cause a packet to be received by all IP hosts that are members of a particular 'subnet'.

To address the need for multicast support (represented by transmission to IP addresses in the Class D space), RFC 2022 ("Support for Multicast over UNI 3.0/3.1 based ATM Networks") [2] was created. This memo creates an analog of the RFC 1577 ARP Server - a new entity known as the MARS (Multicast Address Resolution Server). The MARS operates as a centralized registry and distribution mechanism for mappings between IP multicast addresses and groups of ATM unicast addresses. Host behavior is also defined for establishing and managing point to multipoint VCs, based on the information returned by the MARS, when hosts wish to transmit packets to a multicast group.

This memo aims to show how RFC 2022 may be used to emulate IP broadcast within Logical IP Subnets. While the broadcast technique does not align itself well with the underlying point-to-point nature of ATM, clearly, some applications will still wish to use IP broadcasts. Client-server applications where the client searches for a server by sending out a broadcast is one scenario. Routing protocols, most notably RIP, are other examples.

2. Review of Unicast and Multicast.

Both the unicast and multicast cases take advantage of the point-to-point and point-to-multipoint capabilities defined in the ATM Forum UNI 3.1 document [4]. A unicast IP address has a single ATM level destination. Unicast transmissions occur over point to point Virtual Channels (VCs) between the source and destination. The ARP Server holds mappings between IP destination addresses and their associated ATM destination address. Hosts issue an ARP_REQUEST to the ARP Server when they wish to ascertain a particular mapping. The ARP Server replies with either an ARP_REPLY containing the ATM address of the destination, or an ARP_NAK when the ARP Server is unable to resolve the address. If the request is successful the host establishes a VC to the destination interface. This VC is then used to forward the first (and subsequent) packets to that particular IP destination. RFC 1577 describes in further detail how hosts are administratively grouped in to Logical IP Subnets (LISs), and how the ARP Server establishes the initial mappings for members of the LIS it serves.

The basic host behavior for multicasting is similar - the sender must establish and manage a point to multipoint VC whose leaf nodes are the group's actual members. Under UNI 3.1 these VCs can only be established and altered by the source (root) interface.

The MARS is an evolution of the ARP Server model, and performs two key functions. The first function is the maintenance of a list of ATM addresses corresponding to the members for each group. This list is created by a host registration process which involves two messages - a MARS_JOIN which declares that a host wishes to join the specified group(s), and a MARS_LEAVE which indicates that a host wishes to leave the specified group(s).

MARS_JOIN and MARS_LEAVE messages are also redistributed to all members of the group so that active senders may dynamically adjust their point to multipoint VCs accordingly.

The other major function is the retrieval of group membership from MARS (analogous to the ARP Server providing unicast address mappings). When faced with the need to transmit an IP packet with a Class D destination address, a host issues a MARS_REQUEST to the MARS. If the group has members the MARS returns a MARS_MULTI (possibly in multiple segments) carrying a set of ATM addresses. The host then establishes an initial point to multipoint VC using these ATM addresses as the leaf nodes. If the MARS had no mapping it would return a MARS_NAK.

(RFC 2022 also discusses how the MARS can arrange for Class D groups to be supported by either multicast servers, or meshes of point to multipoint VCs from host to host. However, from the host's perspective this is transparent, and is not central to this discussion of IP broadcast support.)

This memo describes how a host may utilize the registration and group management functions in an existing MARS based IP/ATM network to emulate IP broadcasts.

3. Broadcast as a special case of Multicast.

Many of the problems that occur when implementing a broadcast solution also occur in when implementing a multicast solution. In fact, broadcast may be considered a special case of multicast. That is, broadcast is a multicast group whose members include all members in the LIS.

There are two broadcast groups which this memo addresses:

1) 255.255.255.255 - "All ones" broadcast

2) x.z - CIDR-prefix (subnet) directed broadcast

Broadcast (1) is sometimes referred to as a limited broadcast to this physical network. Broadcast (2) can be thought of as the the broadcast for subnets or networks in the old paradigm. As described in [6] and [7], the notion of subnets and networks is being replaced with a more efficient utilization of the routing address space known as Classless Inter-Domain Routing. The CIDR-prefix (x) is the combination of IP address and subnet mask that denotes the subnet number. The host portion of the address (z) is all ones. One should note that while these broadcasts have different scopes at the IP or network layer, they have precisely the same scope at the link layer -- namely that all members of the LIS will receive a copy.

These addresses may be used in two environments:

- Broadcasting to all members of a given LIS where a priori knowledge of a host's IP address and subnet mask are known (e.g. the CIDR-prefix directed broadcast).

- Broadcasting to all members of a physical network without knowledge of a host's IP address and subnet mask (e.g. the all ones broadcast).

On a broadcast medium like Ethernet, these two environments result in the same physical destination. That is, all stations on that network will receive the broadcast even if they are on different logical subnets, or are non-IP stations. With ATM, this may not be the case. Because ATM is non-broadcast, a registration process must take place. And if there are stations that register to some broadcast groups, but not others, then the different broadcast groups will have different memberships. The notion of broadcast becomes inconsistent.

One case that requires the use of the all ones broadcast is that of the diskless boot, or bootp client, where the host boots up, and does not know its own IP address or subnet mask. Clearly, the host does not know which subnet it belongs to. So, to send a broadcast to its bootp server, the diskless workstation must use the group which contains no subnet information, i.e. the 255.255.255.255 broadcast group. Carrying the example a little further, the bootp server, after receiving the broadcast, can not send either a directed frame nor a subnet directed broadcast to respond to the diskless workstation. Instead, the bootp server must also use the 255.255.255.255 group to communicate with the client.

While the all ones broadcast is required at the IP layer, it also has relevance at the link layer when deciding which broadcast group to register with in MARS. In other words, a bootp client wishing to register for a link layer broadcast, can only register for

255.255.255.255 in the MARS address space because the client's subnet is unknown at the time. Given that some applications must use the all ones address in MARS for their broadcast group, and that we wish to minimize the number of broadcast groups used by LIS members, the all ones group in MARS MUST be used by all members of the LIS when registering to receive broadcast transmissions. The VCC used for transmitting any broadcast packet will be based on the members registered in the MARS under the 255.255.255.255 address position. This VCC will be referred to as the "broadcast channel" through the remainder of this memo.

4. The MARS role in broadcast.

Many solutions have been proposed, some of which are listed in Appendix A. This memo addresses a MARS solution which appears to do the best job of solving the broadcast problem.

There are a number of characteristics of the MARS architecture that should be kept intact. They include:

- MARS contains no knowledge of subnet prefixes and subnet masks. Each group address registered with MARS is managed independently.
- A MARS may only serve one LIS. This insures that the broadcast group 255.255.255.255 is joined by hosts from one LIS, keeping its scope bound to conventional interpretation.
- The Multicast Server (MCS) described in [2] may be used to service the broadcast groups defined in this memo without modification. The MCS will reduce the number of channels used by the network.

The MARS needs no additional code or special algorithms to handle the resolution of IP broadcast addresses. It is simply a general database that holds {Protocol address, ATM.1, ATM.2, ... ATM.n} mappings, and imposes no constraints on the type and length of the 'Protocol address'. Whether the hosts view it as Class D or 'broadcast' (or even IP) is purely a host side issue.

It is likely that end points will want to use the IP broadcast emulation described here in order to support boot time location of the end point's IP address. This leads to the observation that the MARS should NOT expect to see both the IP source and ATM source address fields of the MARS_JOIN filled in. This is reasonable, since only the ATM source address is used when registering the end point as a group member.

The MARS architecture is sufficient to insure the integrity of the broadcast group list without any modification.

5. Host Requirements for Broadcast.

The following list of bullets describes additional characteristics of a MARS-compliant host. These characteristics are required to take advantage of the broadcast function.

- A host must register as a MARS client.
- A host, soon after registration MUST issue a MARS_JOIN to the all ones broadcast address (i.e. 255.255.255.255) with the mar$flags.layer3grp reset.
- When transmitting packets, the host should map all IP layer broadcasts to the VCC (broadcast channel) created and maintained based on the all ones entry in MARS.
- A host MUST monitor the MARS_JOIN/MARS_LEAVE messages for 255.255.255.255 to keep the broadcast channel current.
- A broadcast channel should be torn down after a period of inactivity. The corresponding timeout period MAY be specified with a minimum value of one minute, and a RECOMMENDED default value of 20 minutes.

One should note that while every member participating in the broadcast MUST be a member of the all ones group, not all members will choose to transmit broadcast information. Some members will only elect to receive broadcast information passively. Therefore, in a LIS with n stations, there may be less than n channels terminated at each station for broadcast information. Further reductions may be gained by adding a Multicast Server (MCS) to the broadcast environment which could reduce the number of VCs to two (one incoming, one outgoing), or one for a station that only wishes to listen.

It is well understood that broadcasting in this environment may tax the resources of the network and of the hosts that use it. Therefore, an implementer MAY choose to provide a mechanism for retracting the host’s entry in the broadcast group after it has been established or prior to joining the group. The MARS_LEAVE is used to request withdrawal from the group if the host wishes to disable broadcast reception after it has joined the group. The default behavior SHALL be to join the all ones broadcast group in MARS.

6. Implications of IP broadcast on ATM level resources.

RFC 2022 discusses some of the implications of large multicast groups on the allocation of ATM level resources, both within the network and within end station ATM interfaces.

The default mechanism is for IP multicasting to be achieved using meshes of point to multipoint VCs, direct from source host to group members. Under certain circumstances system administrators may, in a manner completely transparent to end hosts, redirect multicast traffic through ATM level Multicast Servers (MCSs). This may be performed on an individual group basis.

It is sufficient to note here that the IP broadcast 'multicast group' will constitute the largest consumer of VCs within your ATM network when it is active. For this reason it will probably be the first multicast group to have one or more ATM MCSs assigned to support it. However, there is nothing unique about an MCS assigned to support IP broadcast traffic, so this will not be dealt with further in this memo. RFC 2022 contains further discussion on the possible application of multiple MCSs to provide fault-tolerant architectures.

7. Further discussion.

A point of discussion on the ip-atm forum revolved around "auto configuration" and "diskless boot". This memo describes a broadcast solution that requires the use of the MARS. Therefore, at a minimum, the ATM address of the MARS must be manually configured into a diskless workstation. Suggestions such as universal channel numbers, and universal ATM addresses have been proposed, however, no agreement has been reached.

Another topic for discussion is multiprotocol support. MARS is designed for protocol independence. This memo specifically addresses the IP broadcast case, identifying which addresses are most effective in the IP address space. However, the principles apply to any layer 3 protocol. Further work should be performed to identify suitable addresses for other layer 3 protocols.

Finally, there has been support voiced for a link layer broadcast that would be independent of the layer 3 protocol. Such a solution may provide a simpler set of rules through which broadcast applications may be used. In addition, some solutions also provide for more efficient use of VCCs.

Security Considerations

This memo addresses a specific use of the MARS architecture and components to provide the broadcast function. As such, the security implications are no greater or less than the implications of using any of the other multicast groups available in the multicast address range. Should enhancements to security be required, they would need to be added as an extension to the base architecture in RFC 2022.

Acknowledgments

The apparent simplicity of this memo owes a lot to the services provided in [2], which itself is the product of much discussion on the IETF's IP-ATM working group mailing list. Grenville Armitage worked on this document while at Bellcore.

References

[1] Laubach, M., "Classical IP and ARP over ATM", RFC 1577, December 1993.

[2] Armitage, G., "Support for Multicast over UNI 3.0/3.1 based ATM Networks", RFC 2022, November 1995.

[3] Deering, S., "Host Extensions for IP Multicasting", STD 5, RFC 1112, August 1989.

[4] ATM Forum, "ATM User-Network Interface Specification Version 3.0", Englewood Cliffs, NJ: Prentice Hall, September 1993.

[5] Perez, M., Liaw, F., Grossman, D., Mankin, A., Hoffman, E. and A. Malis, "ATM Signaling Support for IP over ATM", RFC 1755, February 1995.

[6] Fuller, V., Li, T., Yu, J., and K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, September 1993.

[7] Baker, F., "Requirements for IP Version 4 Routers", RFC 1812, June 1995.

[8] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels, BCP 14, RFC 2119, March 1997.

Authors' Addresses

Timothy J. Smith
Network Routing Systems,
International Business Machines Corporation.
N21/664
P.O.Box 12195
Research Triangle Park, NC 27709

Phone: (919) 254-4723
EMail: tjsmith@vnet.ibm.com

Grenville Armitage
Bell Labs, Lucent Technologies.
101 Crawfords Corner Rd,
Holmdel, NJ, 07733

EMail: gja@lucent.com

Appendix A. Broadcast alternatives

Throughout the development of this memo, there have been a number of alternatives explored and discarded for one reason or another. This appendix documents these alternatives and the reason that they were not chosen.

A.1 ARP Server Broadcast Solutions.

The ARP Server is a good candidate to support broadcasting. There is an ARP Server for every LIS. The ARP Server contains the entire LIS membership. These are fundamental ingredients for the broadcast function.

A.1.1 Base Solution without modifications to ARP Server.

One may choose as an existing starting point to use only what is available in RFC 1577. That is, a host can easily calculate the range of members in its LIS based on its own IP address and subnet mask. The host can then issue an ARP Request for every member of the LIS. With this information, the host can then set up point-to-point connections with all members, or can set up a point-to-multipoint connection to all members. There you have it, the poor man's broadcast.

While this solution is very straight forward, it suffers from a number of problems.

- The load on the ARP Server is very large. If all stations on a LIS choose to implement broadcasting, the initial surge of ARP Requests will be huge. Some sort of slow start sequence would be needed.

- The amount of resource required makes this a non-scalable solution. The authors believe that broadcasting will require an MCS to reduce the number of channel resources required to support each broadcast 'group'. Using the ARP Server in this manner does not allow an MCS to be transparently introduced. (Basic RFC1577 interfaces also do not implement the extended LLC/SNAP encapsulation required to safely use more than one MCS).

- The diskless boot solution can not function in this environment because it may be unable to determine which subnet to which it belongs.

A.1.2 Enhanced ARP Server solution.

This solution is similar to the base solution except that it takes some of the (MARS) multicast solution and embeds it in the ARP Server. The first enhancement is to add the MARS_MULTI command to the set of opcodes that the ARP Server supports. This would allow a host to issue a single request, and to get back the list of members in one or more MARS_REPLY packets. Rather than have a registration mechanism, the ARP Server could simply use the list of members that have already been registered. When a request comes in for the subnet broadcast address, the ARP Server would aggregate the list, and send the results to the requester.

This suffers from two drawbacks.

1) Scalability with regard to number of VCs is still an issue. One would eventually need to add in some sort of multicast server solution to the ARP Server.

2) The diskless boot scenario is still broken. There is no way for a station to perform a MARS_MULTI without first knowing its IP address and subnet mask.

The diskless boot problem could be solved by adding to the ARP Server a registration process where anyone could register to the 255.255.255.255 address. These changes would make the ARP Server look more and more like MARS.

A.2 MARS Solutions.

If we wish to keep the ARP Server constant as described in RFC 1577, the alternative is to use the Multicast Address Resolution Server (MARS) described in [2].

MARS has three nice features for broadcasting.

1) It has a generalized registration approach which allows for any address to have a group of entities registered. So, if the subnet address is not known, a host can register for an address that is known (e.g. 255.255.255.255).

2) The command set allows for lists of members to be passed in a single MARS_MULTI packet. This reduces traffic.

3) MARS contains an architecture for dealing with the scalability issues. That is, Multicast Servers (MCSs) may be used to set up the point-to-multipoint channels

and reduce the number of channels that a host needs to set up to one. Hosts wishing to broadcast will instead send the packet to the MCS who will then forward it to all members of the LIS.

A.2.1. CIDR-prefix (Subnet) Broadcast solution.

One of the earliest solutions was to simply state that broadcast support would be implemented by using a single multicast group in the class D address space -- namely, the CIDR-prefix (subnet) broadcast address group. All members of a LIS would be required to register to this address, and use it as required. A host wishing to use either the 255.255.255.255 broadcast, or the network broadcast addresses would internally map the VC to the subnet broadcast VC. The all ones and network broadcast addresses would exist on MARS, but would be unused.

The problem with this approach goes back to the diskless workstation problem. Because the workstation may not know which subnet it belongs to, it doesn't know which group to register with.

A.2.2. All one's first, subnet broadcast second

This solution acknowledges that the diskless boot problem requires a generic address (one that does not contain CIDR-prefix (subnet) information) to register with and to use until subnet knowledge is known. In essence, all stations first register to the 255.255.255.255 group, then as they know their subnet information, they could optionally de-register from the all one's group and register to the CIDR-prefix (subnet) broadcast group.

This solution would appear to solve a couple of problems:

1) The bootp client can function if the server remains registered to the all one's group continuously.

2) There will be less traffic using the all ones group because the preferred transactions will be on the subnet broadcast channel.

Unfortunately the first bullet contains a flaw. The server must continually be registered to two groups -- the all ones group and the subnet broadcast group. If this server has multiple processes that are running different IP applications, it may be difficult for the link layer to know which broadcast VC to use. If it always uses the all ones, then it will be missing members that have removed themselves from the all ones and have registered to the subnet broadcast. If it always uses the subnet broadcast group, the

diskless boot scenario gets broken. While making the decision at the link layer may require additional control flows be built into the path, it may also require the rewriting of application software.

In some implementations, a simple constant is used to indicate to the link layer that this packet is to be transmitted to the broadcast "MAC" address. The assumption is that the physical network broadcast and the logical protocol broadcast are one and the same. As pointed out earlier, this is not the case with ATM. Therefore applications would need to specifically identify the subnet broadcast group address to take advantage of the smaller group.

These problems could be solved in a number of ways, but it was thought that they added unnecessarily to the complexity of the broadcast solution.

Appendix B. Should MARS Be Limited to a Single LIS?

RFC 2022 explicitly states that a network administrator MUST ensure that each LIS is served by a separate MARS, creating a one-to-one mapping between cluster and a unicast LIS. But, it also mentions that relaxation of this restriction MAY occur after future research warrants it. This appendix discusses some to the potential implications to broadcast should this restriction be removed.

The most obvious change would be that the notion of a cluster would span more than one LIS. Therefore, the broadcast group of 255.255.255.255 would contain members from more than one LIS.

It also should be emphasized that the one LIS limitation is not a restriction of the MARS architecture. Rather, it is only enforced if an administrator chooses to do so.

Full Copyright Statement