MBoneD Working Group
Internet Draft
Document: draft-ietf-mboned-auto-multicast-02.txt
February 9, 2004

Dave Thaler
Mohit Talwar
Amit Aggarwal
Microsoft
Lorenzo Vicisano
Cisco
Dirk Ooms
Alcatel

IPv4 Automatic Multicast Without Explicit Tunnels (AMT)

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

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

Copyright Notice

1. Abstract

Automatic Multicast Tunneling (AMT) allows multicast communication amongst isolated multicast-enabled sites or hosts, attached to a network which has no native multicast support. It also enables them to exchange multicast traffic with the native multicast infrastructure (MBone) and does not require any manual configuration. AMT uses an encapsulation interface so that no changes to a host stack or applications are required, all protocols (not just UDP) are handled, and there is no additional overhead in core routers.

2. Introduction

The primary goal of this document is to foster the deployment of native IP multicast by enabling a potentially large number of nodes

to connect to the already present multicast infrastructure. Therefore, the techniques discussed here should be viewed as an interim solution to help in the various stages of the transition to a native multicast network.

To allow fast deployment, the solution presented here only requires small and concentrated changes to the network infrastructure, and no changes at all to user applications or to the socket API of end-nodes' operating systems. The protocols introduced in this specification are implemented in a few strategically-placed network nodes and in user-installable software modules (pseudo device drivers and/or user-mode daemons) that reside underneath the socket API of end-nodes' operating systems. This mechanism is very similar to that used by "6to4" [6TO4, ANYCAST] to get automatic IPv6 connectivity.

Effectively, AMT treats the unicast-only internetwork as a large non-broadcast multi-access (NBMA) link layer, over which we require the ability to multicast. To do this, multicast packets being sent to or from a site must be encapsulated in unicast packets. If the group has members in multiple sites, AMT encapsulation of the same multicast packet will take place multiple times by necessity.

The following problems are addressed:

1. Allowing isolated sites/hosts to receive the SSM flavor of multicast ([SSM]).

2. Allowing isolated sites/hosts to transmit the SSM flavor of multicast.

3. Allowing isolated sites/hosts to receive general multicast (ISM [RFC1112]).

This document does not address allowing isolated sites/hosts to transmit general multicast. We expect that other solutions (e.g., Tunnel Brokers, a la [BROKER]) will be used for sites that desire this capability.

3. Definitions

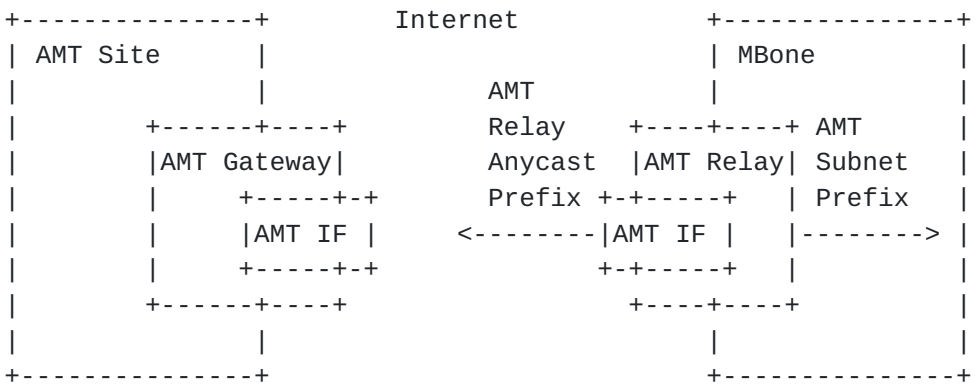


Figure 1: Automatic Multicast Definitions.

AMT Pseudo-Interface

AMT encapsulation of multicast packets inside unicast packets occurs at a point that is logically equivalent to an interface, with the link layer being the unicast-only network. This point is referred to as a pseudo-interface. Some implementations may treat it exactly like any other interface and others may treat it like a tunnel end-point.

AMT Gateway

A host, or a site gateway router, supporting an AMT Pseudo-Interface. It does not have native multicast connectivity to the native multicast backbone infrastructure. It is simply referred to in this document as a "gateway".

AMT Site

A multicast-enabled network not connected to the multicast backbone served by an AMT Gateway. It could also be a stand-alone AMT Gateway.

AMT Relay Router

A multicast router configured to support transit routing between AMT Sites and the native multicast backbone infrastructure. The relay router has one or more interfaces connected to the native multicast infrastructure, zero or more interfaces connected to the non-multicast capable internetwork, and an AMT pseudo-interface. It is simply referred to in this document as a "relay".

As with [6TO4], we assume that normal multicast routers do not want to be tunnel endpoints (especially if this results in high fanout), and similarly that service providers do not want encapsulation to arbitrary routers. Instead, we assume that special-purpose routers will be deployed that are suitable for serving as relays.

AMT Relay Anycast Prefix

A well-known address prefix used to advertise (into the unicast routing infrastructure) a route to an available AMT Relay Router. This could also be private (i.e. not well-known) for a private relay.

The value of this prefix is x.x.x.0/nn [length and value TBD IANA].

AMT Relay Anycast Address

An anycast address which is used to reach the nearest AMT Relay Router.

This address corresponds to host number 1 in the AMT Relay Anycast Prefix, x.x.x.1.

AMT Unicast Autonomous System ID

A 16-bit Autonomous System ID, for use in BGP in accordance to this memo. AS 10888 might be usable for this, but for now we'll assume it's different, to avoid confusion. This number represents a "pseudo-AS" common to all AMT relays using the well known AMT Relay Anycast Prefix (private relays use their own ID).

To protect themselves from erroneous advertisements, managers of border routers often use databases to check the relation between the advertised network and the last hop in the AS path. Associating a specific AS number with the AMT Relay Anycast Address allows us to enter this relationship in the databases

used to check inter-domain routing [ANYCAST].

AMT Subnet Prefix

A well-known address prefix used to advertise (into the M-RIB of the native multicast-enabled infrastructure) a route to AMT Sites. This prefix will be used to enable sourcing SSM traffic from an AMT Gateway.

AMT Gateway Anycast Address

An anycast address in the AMT Subnet Prefix range, which is used by an AMT Gateway to enable sourcing SSM traffic from local applications.

AMT Multicast Autonomous System ID

A 16-bit Autonomous system ID, for use in MBGP in accordance to this memo. This number represents a "pseudo-AS" common to all AMT relays using the well known AMT Subnet Prefix (private relays use their own ID). We assume that the existing AS 10888 is suitable for this purpose. (Note: if this is a problem, a different one would be fine.)

4. Overview

4.1. Receiving Multicast in an AMT Site

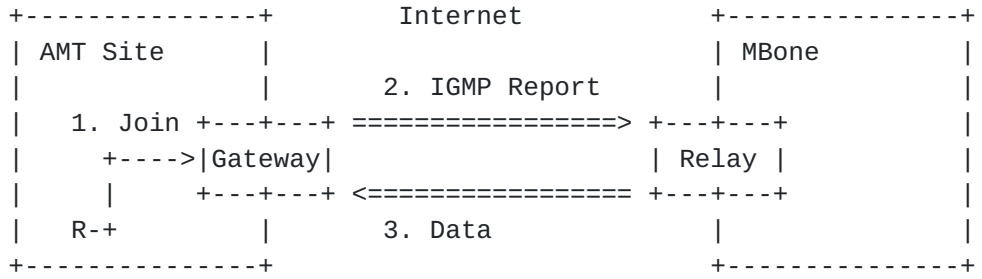


Figure 2: Receiving Multicast in an AMT Site.

AMT relays and gateways cooperate to transmit multicast traffic sourced within the native multicast infrastructure to AMT sites: relays receive the traffic natively and unicast-encapsulate it to gateways; gateways decapsulate the traffic and possibly forward it into the AMT site.

Each gateway has an AMT pseudo-interface that serves as a default multicast route. Requests to join a multicast session are sent to this interface and encapsulated to a particular relay reachable across the unicast-only infrastructure.

Each relay has an AMT pseudo-interface too. Multicast traffic sent on this interface is encapsulated to zero or more gateways that have joined to the relay. The AMT recipient-list is determined for each multicast session. This requires the relay to keep state for each gateway which has joined a particular group or (source, group) pair). Multicast packets from the native infrastructure behind the relay will be sent to each gateway which has requested them.

All multicast packets (data and control) are encapsulated in unicast packets. To work across NAT's, the encapsulation is done over UDP using a well-known port number [TBD IANA].

Each relay, plus the set of all gateways (perhaps unknown to the relay) using the relay, together can be thought of as being on a separate logical NBMA link. This implies that the AMT recipient-list is a list of "link layer" addresses which are (IP address, UDP port) pairs.

Since the number of gateways using a relay can be quite large, and we expect that most sites will not want to receive most groups, an explicit-joining protocol is required for gateways to communicate group membership information to a relay. The two most likely candidates are the IGMP [IGMPv3] protocol, and the PIM-Sparse Mode [PIMSM] protocol. Since an AMT gateway may be a host, and hosts typically do not implement routing protocols, gateways will use IGMP as described in Section 5 below. This allows a host kernel (or a pseudo device driver) to easily implement AMT gateway behavior, and obviates the relay from the need to know whether a given gateway is a host or a router. From the relay's perspective, all gateways are indistinguishable from hosts on an NBMA leaf network.

4.1.1. Scalability Considerations

The requirement that a relay keep group state per gateway that has joined the group introduces potential scalability concerns.

However, scalability of AMT can be achieved by adding more relays, and using an appropriate relay discovery mechanism for gateways to discover relays. The solution we adopt is to assign an anycast address to relays. However, simply sending periodic IGMP Reports to the anycast address can cause duplicates. Specifically, if routing

changes such that a different relay receives a periodic IGMP Report, both the new and old relays will encapsulate data to the AMT site until the old relay's state times out. This is obviously undesirable. Instead, we use the anycast address merely to find a unicast address which can then be used.

Since adding another relay has the result of adding another independent NBMA link, this allows the gateways to be spread out among more relays so as to keep the number of gateways per relay at a reasonable level.

4.1.2 Spoofing Considerations

An attacker could affect the group state in the relay by spoofing the source address in the join or leave reports. This can be used to launch reflection or denial of service attacks on the target. Such attacks can be mitigated by using a three way handshake between the gateway and the relay for each multicast membership report. On receiving an IGMP report, the relay sends a message to the source of the report with the original report as well as a nonce. The state in the relay is updated only on receiving a confirmation for the report with the nonce in it.

4.2. Sourcing Multicast from an AMT site

Two cases are discussed below: multicast traffic sourced in an AMT site and received in the MBone, and multicast traffic sourced in an AMT site and received in another AMT site.

In both cases only SSM sources are supported. Furthermore this specification only deals with the source residing directly in the gateway. To enable a generic node in an AMT site to source multicast, additional coordination between the gateway and the source-node is required.

The general mechanism used to join towards AMT sources is based on the following:

1. Applications residing in the gateway use addresses in the AMT Subnet Prefix to send multicast, as a result of sourcing traffic on the AMT pseudo-interface.

2. The AMT Subnet Prefix is advertised for RPF reachability in the M-RIB by relays and gateways.

3. Relays or gateways that receive a join for a source/group pair

use information encoded in the address pair to rebuild the address of the gateway (source) to which to encapsulate the join (see section 5 for more details). The membership reports use the same three way handshake as outlined in section 4.1.2.

4.2.1. Supporting Site-MBone Multicast

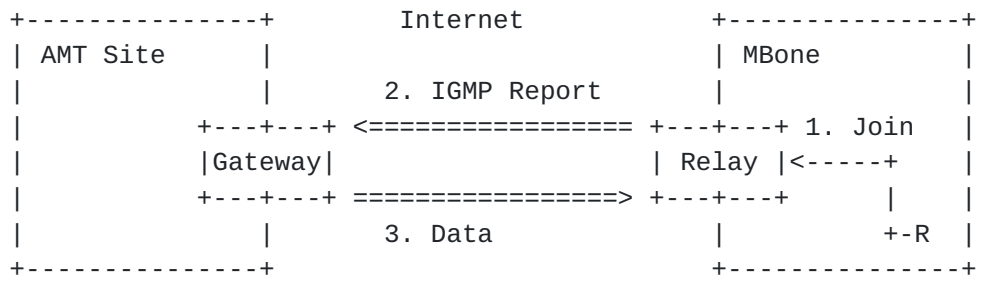


Figure 3: Site-MBone Multicast.

If a relay receives an explicit join from the native infrastructure, for a given (source, group) pair where the source address belongs to the AMT Subnet Prefix, then the relay will periodically (using the rules specified in Section 5) UDP encapsulate an IGMP Report for the group to the gateway. The gateway must keep state per relay from which an IGMP Report has been sent, and forward multicast traffic from the site to all relays from which IGMP Reports have been received. The choice of whether this state and replication is done

at the link-layer (i.e., by the tunnel interface) or at the network-layer is implementation-dependent.

If there are multiple relays present, this ensures that data from the AMT site is received via the closest relay to the receiver. This is necessary when the routers in the native multicast infrastructure employ Reverse-Path Forwarding (RPF) checks against the source address, such as occurs when [PIMSM] is used by the multicast infrastructure.

The solution above will scale to an arbitrary number of relays, as long at the number of relays requiring multicast traffic from a given AMT site remains reasonable enough to not overly burden the site's gateway.

4.2.2. Supporting Site-Site Multicast

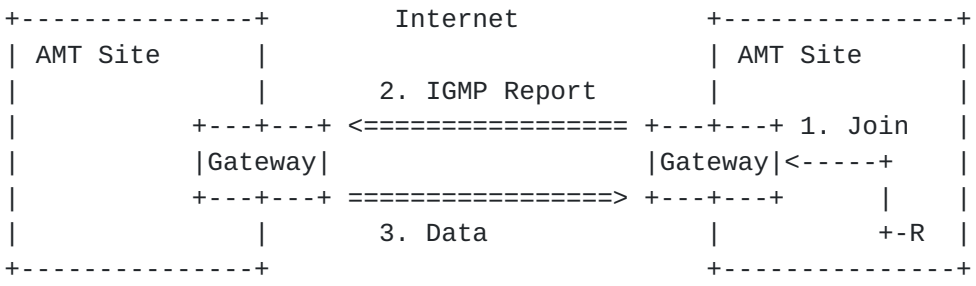


Figure 4: Site-Site Multicast.

Since we require gateways to accept IGMP Reports, as described above, it is also possible to support multicast among AMT sites, without requiring assistance from any relays.

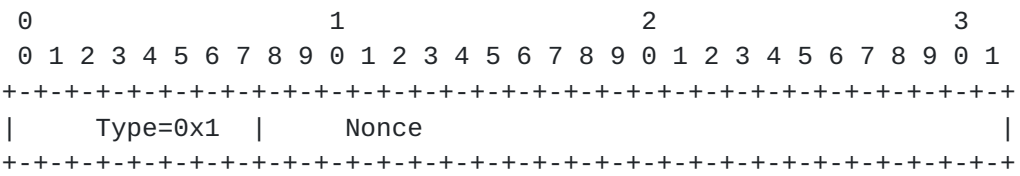
When a gateway wants to join a given (source, group) pair, where the source address belongs to the AMT Subnet Prefix, then the gateway will periodically unicast encapsulate an IGMPv3 [IGMPv3] Report directly to the site gateway for the source.

We note that this can result in a significant amount of state at a site gateway sourcing multicast to a large number of other AMT sites. However, it is expected that this is not unreasonable for two reasons. First, the gateway does not have native multicast connectivity, and as a result is likely doing unicast replication at present. The amount of state is thus the same as what such a site already deals with. Secondly, any site expecting to source traffic to a large number of sites could get a point-to-point tunnel to the native multicast infrastructure, and use that instead of AMT.

5. Message Formats

5.1. AMT Relay Discovery

The AMT Relay Discovery message is a UDP packet sent from the AMT gateway unicast address to the AMT relay anycast address to discover the unicast address of an AMT relay. The payload of the UDP packet contains the following fields.



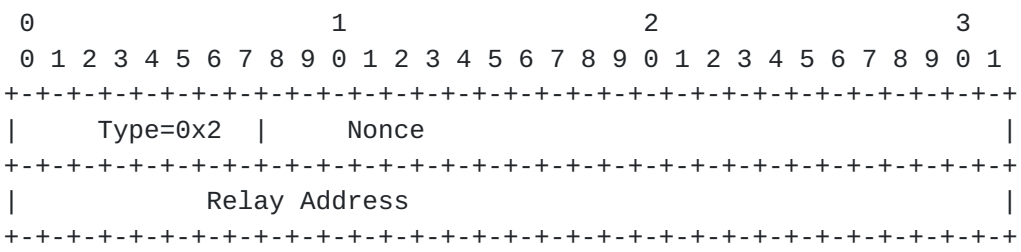
Fields:

Type
: The type of the message.

Nonce
: A 24 bit random value generated by the gateway and replayed by the relay.

5.2. AMT Relay Advertisement

The AMT Relay Advertisement message is a UDP packet sent from the AMT relay anycast address to the source of the discovery message. The payload of the UDP packet contains the following fields.



Fields:

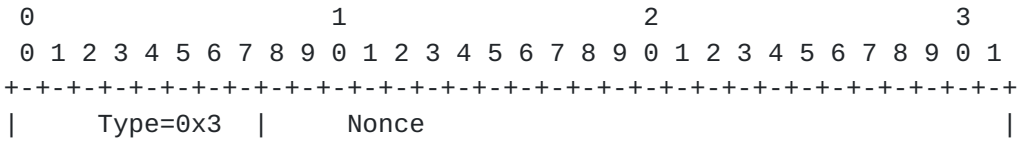
Type
: The type of the message.

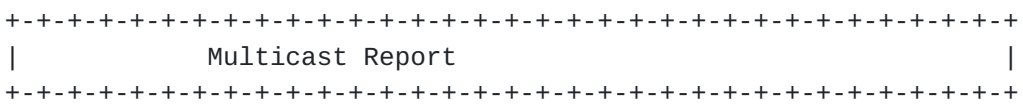
Nonce
: A 24 bit random value replayed from the discovery message.

Relay Address
: The unicast IP address of the AMT relay.

5.3. Membership Report Confirmation

The membership report confirmation is a UDP packet sent by the gateway or relay to the source of a multicast membership report.





Fields:

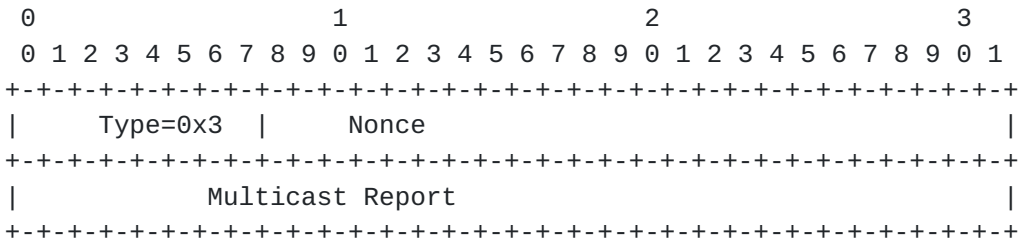
Type
The type of the message.

Nonce
A 24 bit random value generated by the relay or gateway on receiving a multicast report.

Multicast Report
The complete multicast report that the relay or gateway is trying to confirm.

5.4. Membership Report Acknowledgement

The membership report acknowledgement is a UDP packet sent by the source of a membership report to a gateway or relay/



Fields:

Type
The type of the message.

Nonce
A 24 bit random value replayed from the confirmation message.

Multicast Report
The complete multicast report that the relay or gateway is trying to confirm.

6. AMT Gateway Details

This section details the behavior of an AMT Gateway, which may be a router serving an AMT site, or the site may consist of a single host, serving as its own gateway.

6.1. At Startup Time

At startup time, the AMT gateway will bring up an AMT pseudo-interface, to be used for encapsulation. The gateway will then send a AMT Relay Discovery message to the AMT Relay Anycast Address, and note the unicast address (which is treated as a link-layer address to the encapsulation interface) from the AMT Relay Advertisement message. This discovery should be done periodically (e.g., once a day) to re-resolve the unicast address of a close relay. The gateway also initializes a timer used to send periodic IGMP Reports to a random value from the interval [0, [Query Interval]] before sending the first periodic report, in order to prevent startup synchronization (e.g., after a power outage).

If the gateway is serving as a local router, it SHOULD also function as an IGMP Proxy, as described in [IGMPPROXY], with its IGMP host-mode interface being the AMT pseudo-interface. This enables it to translate group memberships on its downstream interfaces into IGMP Reports. The gateway MUST also advertise itself as the default route for multicast in the M-RIB (or it must be the default unicast router if unicast and multicast topologies are congruent). Also, if a shared tree routing protocol is used inside the AMT site, each tree-root must be a gateway, e.g., in PIM-SM each RP must be a gateway.

Finally, to support sourcing traffic to SSM groups by a gateway with a global unicast address, the AMT Subnet Prefix is treated as the subnet prefix of the AMT pseudo-interface, and an anycast address is added on the interface. This anycast address is formed by concatenating the AMT Subnet Prefix followed by the high bits of the gateway's global unicast address. For example, if IANA assigns the prefix x.y/16 as the AMT Subnet Prefix, and the gateway has global unicast address a.b.c.d, then the AMT Gateway's Anycast Address will be x.y.a.b. Note that multiple gateways might end up with the same address anycast assigned to their pseudo-interfaces.

6.2. Joining Groups with MBone Sources

The IGMP protocol usually operates by having the Querier multicast an IGMP Query message on the link. This behavior does not work on NBMA links which do not support multicast. Since the set of gateways is typically unknown to the relay (and potentially quite large), unicasting the queries is also impractical. The following behavior is used instead.

Applications residing in a gateway should join groups on the AMT pseudo-interface, causing IGMP Membership Reports to be sent over that interface. When UDP encapsulating the IGMP Reports (and in fact any other messages, unless specified otherwise in this document), the destination address in the outer IP header is the relay's unicast address. To provide robustness, gateways unicast IGMP Reports to the relay every [Query Interval] (defined as 125 in [IGMPv3]) seconds. The gateway also needs to respond to Membershsip Confirmation messages sent by the relay with a Membership Acknowledgement message.

Generating periodic reports can be done in any implementation-specific manner. Some possibilities include:

1. The AMT pseudo-interface might periodically manufacture IGMPv3 Queries as if they had been received from an IGMP Querier, and deliver them to the IP layer, after which normal IGMP behavior will cause the appropriate reports to be sent.

2. The IGMP module itself might provide an option to operate in periodic mode on specific interfaces.

6.3. Responding to Relay Changes

When a gateway determines that its current relay is unreachable (e.g., upon receipt of a ICMP Unreachable message for the relay's unicast address), it immediately repeats the unicast address resolution step by sending a UDP encapsulated ICMP Echo Request to the AMT Relay Anycast Address, and storing the source address of the UDP encapsulated ICMP Echo Response as the new unicast address to use as a "link-layer" destination.

6.4. Creating SSM groups

When a gateway wants to create an SSM group (i.e., in 232/8) for which it can source traffic, the remaining 24 bits MUST be generated as described below. ([SSM] states that "the policy for allocating these bits is strictly locally determined at the sender's host.")

When the gateway determined its AMT Gateway Anycast Address as described above, it used the high bits of its global unicast address. The remaining bits of its global unicast address are appended to the 232/8 prefix, and any spare bits may be allocated using any policy (again, strictly locally determined at the sender's host).

For example, if the AMT Subnet Prefix is x.y/16, and the device has global unicast address a.b.c.d, then it MUST allocate SSM groups in

the range 232.c.d/24.

6.5. Joining SSM Groups with AMT Sources

An IGMPv3 Report for a given (source, group) pair MAY be encapsulated directly to the source, when the source address belongs to the AMT Subnet Prefix.

The "link-layer" address to use as the destination address in the outer IP header is obtained as follows. The source address in the inclusion list of the IGMPv3 report will be an AMT Gateway Anycast Address with the high bits of the address, and the remaining bits will be in the middle of the group address.

For example, if the AMT Subnet Prefix is x.y/16, and the IGMPv3 Report is for (x.y.a.b, 232.c.d.e), then the "link layer" destination address used for encapsulation is a.b.c.d.

6.6. Receiving IGMPv3 Reports on the AMT Interface

When an IGMPv3 report is received on the AMT pseudo-interface, and the report is a request to join a given (S, G) pair, then the following actions are taken.

If S is not the AMT Gateway Anycast Address of the gateway, the packet is dropped. If G does not contain the low bits of the global unicast address (as described above), the packet is also dropped.

Otherwise, the gateway sends a Membership Confirmation message to the source of the IGMPv3 report. The message contains a random nonce. On receiving a Membership Acknowledgement message, the gateway verifies that the nonce in the acknowledgement is the same as the one in the confirmation message. If the two differ, the message is dropped without any change to the gateway state. If the two nonces are the same, the gateway adds the source address (from the outer IP header) and UDP port of the report to a membership list for G. Maintaining this membership list may be done in any implementation-dependent manner. For example, it might be maintained by the "link-layer" inside the AMT pseudo-interface, making it invisible to the normal IGMP module.

6.7. Sending data to SSM groups

When multicast packets are sent on the AMT pseudo-interface, they

are encapsulated as follows. If the group address is not an SSM group, then the packet is dropped (this memo does not currently provide a way to send to non-SSM groups).

If the group address is an SSM group, then the packet is unicast encapsulated to each remote node from which the gateway has received an IGMPv3 report for the packet's (source, group) pair.

7. Relay Router Details

7.1. At startup time

At startup time, the relay router will bring up an NBMA-style AMT pseudo-interface. It shall also add the AMT Relay Anycast Address on some interface.

The relay router shall then advertise the AMT Relay Anycast Prefix into the unicast-only Internet, as if it were a connection to an external network. When the advertisement is done using BGP, the AS path leading to the AMT Relay Anycast Prefix shall include the identifier of the local AS and the AMT Unicast Autonomous System ID.

The relay router shall also enable IGMPv3 on the AMT pseudo-interface, except that it shall not multicast Queries (this might be done, for example, by having the AMT pseudo-device drop them, or by having the IGMP module not send them in the first place).

Finally, to support sourcing SSM traffic from AMT sites, the AMT Subnet Prefix is assigned to the AMT pseudo-interface, and the AMT Subnet Prefix is injected into the M-RIB of MBGP.

7.2. Receiving Echo Requests to the Anycast Address

When a relay receives a AMT Relay Discovery message directed to the AMT Relay Anycast Address, it should respond with a AMT Relay Advertisement containing its unicast address. The source and destination addresses of the advertisement should be the same as the destination and source addresses of the discovery message respectively. Further, the nonce in the discovery message MUST be copied into the advertisement message.

7.3. Receiving Joins from AMT Gateways

The relay operates passively, sending no Queries but simply tracking

membership information according to Reports and Leave messages, as a router normally would. In addition, the relay must also do explicit membership tracking, as to which gateways on the AMT pseudo-interface have joined which groups. On receiving a membership report, the gateway generates a Membership Confirmation message with a random nonce in it. On receiving a Membership Acknowledgement, it updates the group state if the nonce in the reply matches the one in the confirmation message. When data arrives for that group, the traffic must be encapsulated to each gateway which has joined that group.

The explicit membership tracking and unicast replication may be done in any implementation-specific manner. Some examples are:

1. The AMT pseudo-device driver might track the group information and perform the replication at the "link-layer", with no changes to a pre-existing IGMP module.

2. The IGMP module might have native support for explicit membership tracking, especially if it supports other NBMA-style interfaces.

7.4. Receiving (S,G) Joins from the Native Side, for AMT Sources

The relay encapsulates an IGMPv3 report to the AMT source as described above in Section 5.5.

8. IANA Considerations

The IANA should allocate a prefix dedicated to the public AMT Relays to the native multicast backbone. The prefix length should be determined by the IANA; the prefix should be large enough to guarantee advertisement in the default- free BGP networks; a length of 16 will meet this requirement. This is a one time effort; there is no need for any recurring assignment after this stage.

The IANA should also allocate an Autonomous System ID which can be used as a pseudo-AS when advertising routes to the above prefix. Furthermore, to support sourcing SSM traffic from AMT gateways, the IANA should allocate a subnet prefix dedicated to the AMT link. The prefix length should be determined by the IANA; the prefix should be large enough to guarantee advertisement in the default- free BGP networks; a length of 16 will meet this requirement. This is a one time effort; there is no need for any recurring assignment after

this stage. It should also be noted that this prefix length directly affects the number of groups available to be created by the AMT gateway: a length of 16 gives 256 groups, and a length of 8 gives 65536 groups. For diagnostic purposes, it is helpful to have a prefix length which is a multiple of 8, although this is not required.

An autonomous system number dedicated to a pseudo-AS for multicast is already in use today (AS 10888), and so it is expected that no additional AS number is required for this prefix.

Finally, IANA should reserve a well-known UDP port number for AMT encapsulation.

9. Security Considerations

The anycast technique introduces a risk that a rogue router or a rogue AS could introduce a bogus route to the AMT Relay Anycast Prefix, and thus divert the traffic. Network managers have to guarantee the integrity of their routing to the AMT Relay anycast prefix in much the same way that they guarantee the integrity of all other routes.

Within the native MBGP infrastructure, there is a risk that a rogue router or a rogue AS could introduce a bogus route to the AMT Subnet Prefix, and thus divert joins and cause RPF failures of multicast traffic. Again, network managers have to guarantee the integrity of the MBGP routing to the AMT subnet prefix in much the same way that they guarantee the integrity of all other routes in the M-RIB.

Gateways and relays will accept and decapsulate multicast traffic from any source from which regular unicast traffic is accepted. If this is for any reason felt to be a security risk, then additional source address based packet filtering MUST be applied:

1. To avoid that a rogue sender (that can't do traditional spoofing because of e.g. access lists deployed by its ISP) makes use of AMT to send packets to an SSM tree, a relay that receives an encapsulated multicast packet MUST discard the multicast packet if the IPv4 source address in the outer header is not composed of the last 2 bytes of the source address and the 2 middle bytes of the destination address of the inner header (i.e. a.b.c.d must be composed of the a.b of x.y.a.b and the c.d of 232.c.d.e).

2. A gateway MUST discard encapsulated multicast packets if the

source address in the outer header is not the address to which the encapsulated join message was sent. An AMT Gateway that receives an encapsulated IGMPv3 (S,G)-Join MUST discard the message if the IPv4 destination address in the outer header is not composed of the last 2 bytes of S and the 2 middle bytes of G (i.e. the destination address a.b.c.d must be composed of the a.b of the multicast source x.y.a.b and the c.d of the multicast group 232.c.d.e).

3. A gateway MUST drop an AMT Relay Advertisement if the nonce in the advertisement does not match the nonce in the discovery packet sent by the gateway. This prevents an attacker from acting as an AMT anycast relay even without publishing a route to the AMT Anycast Subnet Prefix.

4. A gateway or relay MUST not update its group state on receiving a membership report. Instead, it MUST generate a Membership Confirmation message to the source of the report. On receiving a Membership Acknowledgement, the group state should be updated only if the nonce in the acknowledgement matches the one in the confirmation message. This prevents an attacker from spoofing the source address of a membership report and causing a denial of service or reflection attack on the target.

10. Acknowledgements

Most of the mechanisms described in this document are based on similar work done by the NGTrans WG for obtaining automatic IPv6 connectivity without explicit tunnels ("6to4"). Tony Ballardie provided helpful discussion that inspired this document.

11. Appendix A: Open Issues

Under the proposed mechanism, a gateway sends its IGMPv3 Reports for MBone sources to the relay closest to itself (discovered using the UDP encapsulated "ping"). This ensures that, as far as possible, multicast traffic flows through the native multicast infrastructure and the automatic multicast encapsulation is short.

However, there might be reasons to create automatic tunnels to the relay closest to the MBone source instead. An ISP, for example, might be willing to provide a relay for only its own customers, those wishing to multicast their transmission to a much wider audience. A mechanism, complementary to the one described in this document, might be used to provide this facility. It uses UDP encapsulated ICMP Redirect messages as described below.

While injecting routes for its sources into the M-RIB, such an ISP might, for example, use a new BGP attribute to convey the address of the preferred relay. This would let other relays redirect any IGMP Reports to the preferred relay by sending a UDP encapsulated ICMP Redirect.

An IGMP Report sent by a gateway to the relay closest to it would consist of the following packet:

OuterIP [UDP [InnerIP [IGMP Report]]]

The relay would respond with:

OuterIP' [UDP' [InnerIP' [ICMP Redirect [InnerIP [IGMP Report]]]]]

An ICMP Redirect contains the first 64 bits of the original packet [ICMP]. Hence the gateway would get 44 bytes (64 - sizeof(Inner IP)) of the IGMP Report, enough to easily extract the (source, group) pair, and redirect its report to the preferred gateway.

Certainly additional complexity is undesirable, so it is an open issue as to whether redirects are needed at all.

12. Authors' Addresses

Dave Thaler

Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
Phone: +1 425 703 8835
EMail: dthaler@microsoft.com

Mohit Talwar
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
Phone: +1 425 705 3131
EMail: mohitt@microsoft.com

Amit Aggarwal
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
Phone: +1 425 706 0593
EMail: amitag@microsoft.com

Lorenzo Vicisano
Cisco Systems
170 West Tasman Dr.
San Jose, CA 95134
Phone: +1 408 525 2530
EMail: lorenzo@cisco.com

Dirk Ooms
Alcatel
F. Wellesplein 1, 2018 Antwerp, Belgium
Phone: +32 3 2404732
EMail: dirk.ooms@alcatel.be

13. Normative References

[ICMP] Postel, J., "Internet Control Message Protocol", RFC 792, September 1981.

[IGMPPROXY] W. Fenner, "IGMP-based Multicast Forwarding (``IGMP Proxying'')", Work in progress, draft-fenner-igmp-proxy-03.txt, July 2000.

[IGMPv3] Cain, B., Deering, S., Fenner, B., Kouvelas, I., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[SSM] Holbrook, H., and B. Cain, "Source-Specific Multicast for IP", Work in progress, draft-holbrook-ssm-arch-04.txt, October 2003.

14. Informative References

[6TO4] Carpenter, B., and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[BROKER] Durand, A., Fasano, P., Guardini, I., and D. Lento, "IPv6 Tunnel Broker", RFC 3053, January 2001.

[ANYCAST] C. Huitema, "An Anycast Prefix for 6to4 Relay Routers", RFC 3068, June 2001.

[PIMSM] Estrin, D. Farinacci, D., Helmy, A., Thaler, D., Deering, S., Handley, M., Jacobson, V., Liu, C., Sharma, P., and L. Wei. "Protocol Independent Multicast-Sparse Mode (PIM-SM):

Protocol Specification", RFC 2362, June 1998.

15. Full Copyright Statement