Use of Multicast across Inter-domain Peering Points

Abstract

   This document examines the use of Source-Specific Multicast (SSM)
   across inter-domain peering points for a specified set of deployment
   scenarios. The objectives are to (1) describe the setup process for
   multicast-based delivery across administrative domains for these
   scenarios and (2) document supporting functionality to enable this
   process.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   Content and data from several types of applications (e.g., live video
   streaming, software downloads) are well suited for delivery via
   multicast means. The use of multicast for delivering such content or
   other data offers significant savings in terms of utilization of
   resources in any given administrative domain. End User (EU) demand
   for such content or other data is growing. Often, this requires
   transporting the content or other data across administrative domains
   via inter-domain peering points.

   The objectives of this document are twofold:

   o Describe the technical process and establish guidelines for
      setting up multicast-based delivery of application content or
      other data across inter-domain peering points via a set of
      use cases (where "Use Case 3.1" corresponds to Section 3.1,
      "Use Case 3.2" corresponds to Section 3.2, etc.).

   o Catalog all required exchanges of information between the
      administrative domains to support multicast-based delivery. This
      enables operators to initiate necessary processes to support
      inter-domain peering with multicast.

   The scope and assumptions for this document are as follows:

   o Administrative Domain 1 (AD-1) sources content to one or more EUs
      in one or more Administrative Domain 2 (AD-2) entities. AD-1 and
      AD-2 want to use IP multicast to allow support for large and
      growing EU populations, with a minimum amount of duplicated
      traffic to send across network links.

      * This document does not detail the case where EUs are
         originating content. To support that additional service, it is
         recommended that some method (outside the scope of this
         document) be used by which the content from EUs is transmitted
         to the application in AD-1 and AD-1 can send out the traffic as
         IP multicast. From that point on, the descriptions in this
         document apply, except that they are not complete because they
         do not cover the transport or operational aspects of the leg
         from the EU to AD-1.

      * This document does not detail the case where AD-1 and AD-2 are
         not directly connected to each other and are instead connected
         via one or more other ADs (as opposed to a peering point) that
         serve as transit providers. The cases described in this
         document where tunnels are used between AD-1 and AD-2 can be
         applied to such scenarios, but SLA ("Service Level Agreement")

control, for example, would be different. Additional issues
will likely exist as well in such scenarios. This topic is
left for further study.

o For the purposes of this document, the term "peering point" refers
   to a network connection ("link") between two administrative
   network domains over which traffic is exchanged between them.
   This is also referred to as a Network-to-Network Interface (NNI).
   Unless otherwise noted, it is assumed that the peering point is a
   private peering point, where the network connection is a
   physically or virtually isolated network connection solely between
   AD-1 and AD-2. The other case is that of a broadcast peering
   point, which is a common option in public Internet Exchange Points
   (IXPs). See Section 4.2.4 for more details.

o AD-1 is enabled with native multicast. A peering point exists
   between AD-1 and AD-2.

o It is understood that several protocols are available for this
   purpose, including Protocol-Independent Multicast - Sparse Mode
   (PIM-SM) and Protocol-Independent Multicast - Source-Specific
   Multicast (PIM-SSM) [RFC7761], the Internet Group Management
   Protocol (IGMP) [RFC3376], and Multicast Listener Discovery (MLD)
   [RFC3810].

o As described in Section 2, the source IP address of the (so-called
   "(S,G)") multicast stream in the originating AD (AD-1) is known.
   Under this condition, using PIM-SSM is beneficial, as it allows
   the receiver's upstream router to send a join message directly to
   the source without the need to invoke an intermediate Rendezvous
   Point (RP). The use of SSM also presents an improved threat
   mitigation profile against attack, as described in [RFC4609].
   Hence, in the case of inter-domain peering, it is recommended that
   only SSM protocols be used; the setup of inter-domain peering for
   ASM (Any-Source Multicast) is out of scope for this document.

o The rest of this document assumes that PIM-SSM and BGP are used
   across the peering point, plus Automatic Multicast Tunneling (AMT)
   [RFC7450] and/or Generic Routing Encapsulation (GRE), according to
   the scenario in question. The use of other protocols is beyond
   the scope of this document.

o AMT is set up at the peering point if either the peering point or
   AD-2 is not multicast enabled. It is assumed that an AMT relay
   will be available to a client for multicast delivery. The
   selection of an optimal AMT relay by a client is out of scope for

this document. Note that using AMT is necessary only when native
multicast is unavailable in the peering point (Use Case 3.3) or in
the downstream administrative domain (Use Cases 3.4 and 3.5).

o It is assumed that the collection of billing data is done at the
   application level and is not considered to be a networking issue.
   The settlements process for EU billing and/or inter-provider
   billing is out of scope for this document.

o Inter-domain network connectivity troubleshooting is only
   considered within the context of a cooperative process between the
   two domains.

This document also attempts to identify ways by which the peering
process can be improved. Development of new methods for improvement
is beyond the scope of this document.

2. Overview of Inter-domain Multicast Application Transport

A multicast-based application delivery scenario is as follows:

o Two independent administrative domains are interconnected via a
   peering point.

o The peering point is either multicast enabled (end-to-end native
   multicast across the two domains) or connected by one of two
   possible tunnel types:

   * A GRE tunnel [RFC2784] allowing multicast tunneling across the
      peering point, or

   * AMT [RFC7450].

o A service provider controls one or more application sources in
   AD-1 that will send multicast IP packets via one or more (S,G)s
   (multicast traffic flows; see Section 4.2.1 if you are unfamiliar
   with IP multicast). It is assumed that the service being provided
   is suitable for delivery via multicast (e.g., live video streaming
   of popular events, software downloads to many devices) and that
   the packet streams will be carried by a suitable multicast
   transport protocol.

o An EU controls a device connected to AD-2, which runs an
   application client compatible with the service provider's
   application source.

   o The application client joins appropriate (S,G)s in order to
      receive the data necessary to provide the service to the EU. The
      mechanisms by which the application client learns the appropriate
      (S,G)s are an implementation detail of the application and are out
      of scope for this document.

   The assumption here is that AD-1 has ultimate responsibility for
   delivering the multicast-based service on behalf of the content
   source(s). All relevant interactions between the two domains
   described in this document are based on this assumption.

   Note that AD-2 may be an independent network domain (e.g., a Tier 1
   network operator domain). Alternately, AD-2 could also be an
   enterprise network domain operated by a single customer of AD-1. The
   peering point architecture and requirements may have some unique
   aspects associated with enterprise networks; see Section 3.

   The use cases describing various architectural configurations for
   multicast distribution, along with associated requirements, are
   described in Section 3. Section 4 contains a comprehensive list of
   pertinent information that needs to be exchanged between the two
   domains in order to support functions to enable application
   transport.

3. Inter-domain Peering Point Requirements for Multicast

   The transport of applications using multicast requires that the
   inter-domain peering point be enabled to support such a process.
   This section presents five use cases for consideration.

3.1. Native Multicast

   This use case involves end-to-end native multicast between the two
   administrative domains, and the peering point is also native
   multicast enabled. See Figure 1.
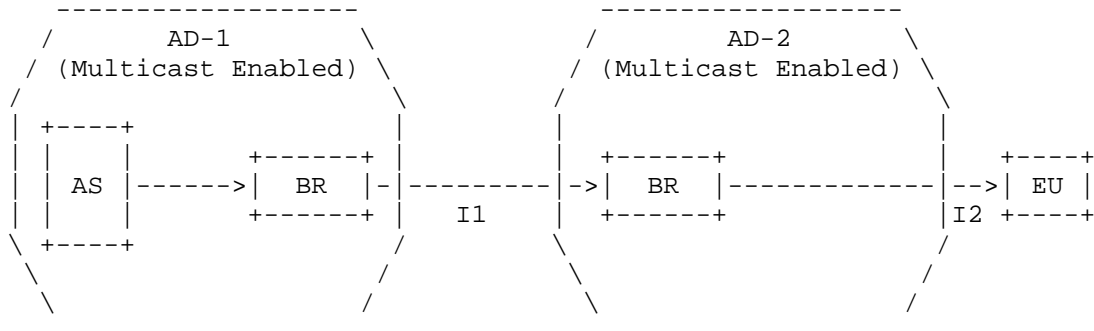
```
     ------------------                    ------------------
    /       AD-1       \                  /       AD-2       \
   / (Multicast Enabled) \               / (Multicast Enabled) \
  /                    \                /                    \
 | +----+               |              |                    |
 | |    |      +------+  |              |  +------+          |   +----+
 | | AS |------>|  BR  |-|----------|-->|  BR  |-------------|-->| EU |
 | |    |      +------+  |   I1     |  +------+          |I2 +----+
 \ +----+             /              \                    /
  \                  /                \                  /
   \                /                  \                /
     ------------------                    ------------------
```

   AD = Administrative Domain (independent autonomous system)
   AS = multicast (e.g., content) Application Source
   BR = Border Router
   I1 = AD-1 and AD-2 multicast interconnection (e.g., MP-BGP)
   I2 = AD-2 and EU multicast connection

        Figure 1: Content Distribution via End-to-End Native Multicast

   Advantages of this configuration:

   o Most efficient use of bandwidth in both domains.

   o Fewer devices in the path traversed by the multicast stream when
      compared to an AMT-enabled peering point.

   From the perspective of AD-1, the one disadvantage associated with
   native multicast to AD-2 instead of individual unicast to every EU in
   AD-2 is that it does not have the ability to count the number of EUs
   as well as the transmitted bytes delivered to them. This information
   is relevant from the perspective of customer billing and operational
   logs. It is assumed that such data will be collected by the
   application layer. The application-layer mechanisms for generating
   this information need to be robust enough so that all pertinent
   requirements for the source provider and the AD operator are
   satisfactorily met. The specifics of these methods are beyond the
   scope of this document.

Architectural guidelines for this configuration are as follows:

a. Dual homing for peering points between domains is recommended as
    a way to ensure reliability with full BGP table visibility.

b. If the peering point between AD-1 and AD-2 is a controlled
    network environment, then bandwidth can be allocated accordingly
    by the two domains to permit the transit of non-rate-adaptive
    multicast traffic. If this is not the case, then the multicast
    traffic must support congestion control via any of the mechanisms
    described in Section 4.1 of [BCP145].

c. The sending and receiving of multicast traffic between two
    domains is typically determined by local policies associated with
    each domain. For example, if AD-1 is a service provider and AD-2
    is an enterprise, then AD-1 may support local policies for
    traffic delivery to, but not traffic reception from, AD-2.
    Another example is the use of a policy by which AD-1 delivers
    specified content to AD-2 only if such delivery has been accepted
    by contract.

d. It is assumed that relevant information on multicast streams
    delivered to EUs in AD-2 is collected by available capabilities
    in the application layer. The precise nature and formats of the
    collected information will be determined by directives from the
    source owner and the domain operators.

3.2. Peering Point Enabled with GRE Tunnel

   The peering point is not native multicast enabled in this use case.
   There is a GRE tunnel provisioned over the peering point. See
   Figure 2.
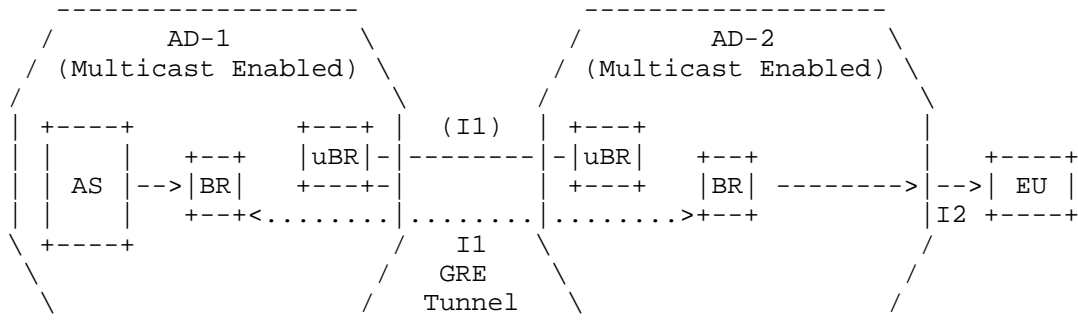
```
     ------------------                  ------------------
    /      AD-1        \                /      AD-2        \
   / (Multicast Enabled) \            / (Multicast Enabled) \
  /                     \            /                     \
 | +----+           +---+ | (I1)  | +---+                  |
 | |    |      +--+  |uBR|-|-------|-|uBR|     +--+         |    +----+
 | | AS |-->|BR|  +---+-|       | +---+     |BR| -------->|-->| EU |
 | |    |      +--+<........|........|       |........>+--+         |I2 +----+
  \ +----+         /  I1   \                              /
   \             /   GRE     \                          /
    \           /   Tunnel     \                      /
     ------------------                  ------------------
```

   AD = Administrative Domain (independent autonomous system)
   AS = multicast (e.g., content) Application Source
   uBR = unicast Border Router - not necessarily multicast enabled;
        may be the same router as BR
   BR = Border Router - for multicast
   I1 = AD-1 and AD-2 multicast interconnection (e.g., MP-BGP)
   I2 = AD-2 and EU multicast connection

              Figure 2: Content Distribution via GRE Tunnel

   In this case, interconnection I1 between AD-1 and AD-2 in Figure 2 is
   multicast enabled via a GRE tunnel [RFC2784] between the two BRs and
   encapsulating the multicast protocols across it.

   Normally, this approach is chosen if the uBR physically connected to
   the peering link cannot or should not be enabled for IP multicast.
   This approach may also be beneficial if the BR and uBR are the same
   device but the peering link is a broadcast domain (IXP); see
   Section 4.2.4.

   The routing configuration is basically unchanged: instead of running
   BGP (SAFI-2) ("SAFI" stands for "Subsequent Address Family
   Identifier") across the native IP multicast link between AD-1 and
   AD-2, BGP (SAFI-2) is now run across the GRE tunnel.

Advantages of this configuration:

o Highly efficient use of bandwidth in both domains, although not as
   efficient as the fully native multicast use case (Section 3.1).

o Fewer devices in the path traversed by the multicast stream when
   compared to an AMT-enabled peering point.

o Ability to support partial and/or incremental IP multicast
   deployments in AD-1 and/or AD-2: only the path or paths between
   the AS/BR (AD-1) and the BR/EU (AD-2) need to be multicast
   enabled. The uBRs may not support IP multicast or enabling it
   could be seen as operationally risky on that important edge node,
   whereas dedicated BR nodes for IP multicast may (at least
   initially) be more acceptable. The BR can also be located such
   that only parts of the domain may need to support native IP
   multicast (e.g., only the core in AD-1 but not edge networks
   towards the uBR).

o GRE is an existing technology and is relatively simple to
   implement.

Disadvantages of this configuration:

o Per Use Case 3.1, current router technology cannot count the
   number of EUs or the number of bytes transmitted.

o The GRE tunnel requires manual configuration.

o The GRE tunnel must be established prior to starting the stream.

o The GRE tunnel is often left pinned up.

Architectural guidelines for this configuration include the
following:

Guidelines (a) through (d) are the same as those described in
Use Case 3.1. Two additional guidelines are as follows:

e. GRE tunnels are typically configured manually between peering
    points to support multicast delivery between domains.

f. It is recommended that the GRE tunnel (tunnel server)
    configuration in the source network be such that it only
    advertises the routes to the application sources and not to the
    entire network. This practice will prevent unauthorized delivery

      of applications through the tunnel (for example, if the
      application (e.g., content) is not part of an agreed-upon
      inter-domain partnership).

3.3. Peering Point Enabled with AMT - Both Domains Multicast Enabled

   It is assumed that both administrative domains in this use case are
   native multicast enabled here; however, the peering point is not.

   The peering point is enabled with AMT. The basic configuration is
   depicted in Figure 3.
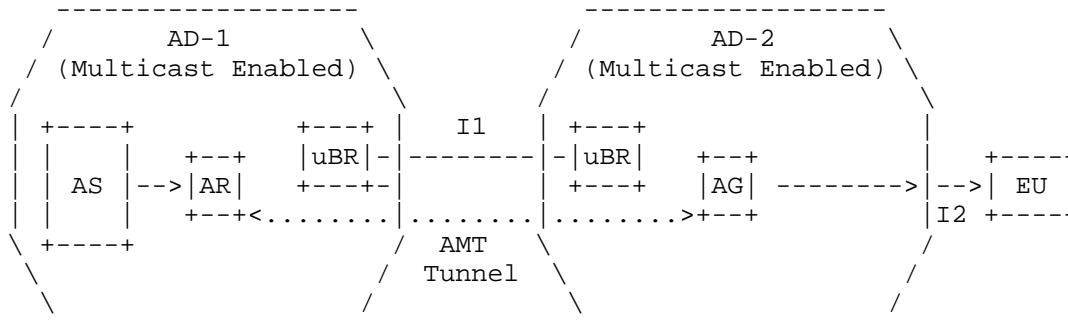
```
        ------------------              ------------------
      /       AD-1       \            /       AD-2       \
     / (Multicast Enabled) \        / (Multicast Enabled) \
    /                     \        /                     \
   | +----+              +---+ |   I1  | +---+            |
   | |    |    +--+    |uBR|-|--------|-|uBR|    +--+     |   +----+
   | | AS |-->|AR|    +---+-|        | +---+   |AG| -------->|-->| EU |
   | |    |   +--+<........|........|........>+--+          |I2 +----+
    \ +----+            /  AMT   \                        /
     \                 /  Tunnel  \                      /
      \               /            \                    /
        ------------------              ------------------
```

   AD = Administrative Domain (independent autonomous system)
   AS = multicast (e.g., content) Application Source
   AR = AMT Relay
   AG = AMT Gateway
   uBR = unicast Border Router - not multicast enabled;
         also, either AR = uBR (AD-1) or uBR = AG (AD-2)
   I1 = AMT interconnection between AD-1 and AD-2
   I2 = AD-2 and EU multicast connection

           Figure 3: AMT Interconnection between AD-1 and AD-2

   Advantages of this configuration:

   o Highly efficient use of bandwidth in AD-1.

   o AMT is an existing technology and is relatively simple to
      implement. Attractive properties of AMT include the following:

      * Dynamic interconnection between the gateway-relay pair across
         the peering point.

      * Ability to serve clients and servers with differing policies.

Disadvantages of this configuration:

o Per Use Case 3.1 (AD-2 is native multicast), current router
   technology cannot count the number of EUs or the number of bytes
   transmitted to all EUs.

o Additional devices (AMT gateway and relay pairs) may be introduced
   into the path if these services are not incorporated into the
   existing routing nodes.

o Currently undefined mechanisms for the AG to automatically select
   the optimal AR.

Architectural guidelines for this configuration are as follows:

Guidelines (a) through (d) are the same as those described in
Use Case 3.1. In addition,

e. It is recommended that AMT relay and gateway pairs be configured
    at the peering points to support multicast delivery between
    domains. AMT tunnels will then configure dynamically across the
    peering points once the gateway in AD-2 receives the (S,G)
    information from the EU.

3.4. Peering Point Enabled with AMT - AD-2 Not Multicast Enabled

   In this AMT use case, AD-2 is not multicast enabled. Hence, the
   interconnection between AD-2 and the EU is also not multicast
   enabled. This use case is depicted in Figure 4.
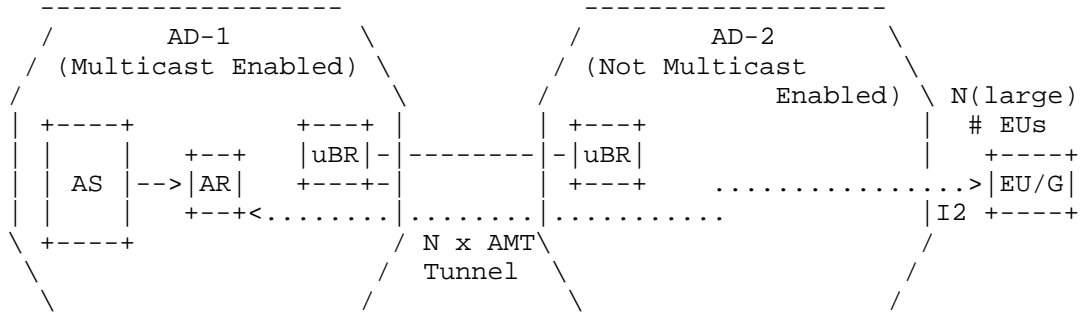
```
        ------------------          ------------------
       /      AD-1        \        /      AD-2         \
      / (Multicast Enabled) \     / (Not Multicast      \
     /                     \    /            Enabled) \ N(large)
    | +----+          +---+ |   | +---+               | # EUs
    | |    |   +--+   |uBR|-|---------|-|uBR|          |   +---+
    | | AS |-->|AR|   +---+-|        | +--+   .............>|EU/G|
    | |    |   +--+<........|........|...........      |I2 +---+
    \ +----+         / N x AMT\               /
     \              / Tunnel \              /
      \            /          \            /
        ------------------          ------------------
```

   AS = multicast (e.g., content) Application Source
   uBR = unicast Border Router - not multicast enabled;
        otherwise, AR = uBR (in AD-1)
   AR = AMT Relay
   EU/G = Gateway client embedded in EU device
   I2 = AMT tunnel connecting EU/G to AR in AD-1 through
        non-multicast-enabled AD-2

       Figure 4: AMT Tunnel Connecting AD-1 AMT Relay and EU Gateway

   This use case is equivalent to having unicast distribution of the
   application through AD-2. The total number of AMT tunnels would be
   equal to the total number of EUs requesting the application. The
   peering point thus needs to accommodate the total number of AMT
   tunnels between the two domains. Each AMT tunnel can provide the
   data usage associated with each EU.

   Advantages of this configuration:

   o Efficient use of bandwidth in AD-1 (the closer the AR is to the
      uBR, the more efficient).

   o Ability of AD-1 to introduce content delivery based on IP
      multicast, without any support by network devices in AD-2: only
      the application side in the EU device needs to perform AMT gateway
      library functionality to receive traffic from the AMT relay.

   o Allows AD-2 to "upgrade" to Use Case 3.5 (see Section 3.5) at a
      later time, without any change in AD-1 at that time.

    o AMT is an existing technology and is relatively simple to
       implement. Attractive properties of AMT include the following:

       * Dynamic interconnection between the AMT gateway-relay pair
          across the peering point.

       * Ability to serve clients and servers with differing policies.

    o Each AMT tunnel serves as a count for each EU and is also able to
       track data usage (bytes) delivered to the EU.

   Disadvantages of this configuration:

   o Additional devices (AMT gateway and relay pairs) are introduced
      into the transport path.

   o Assuming multiple peering points between the domains, the EU
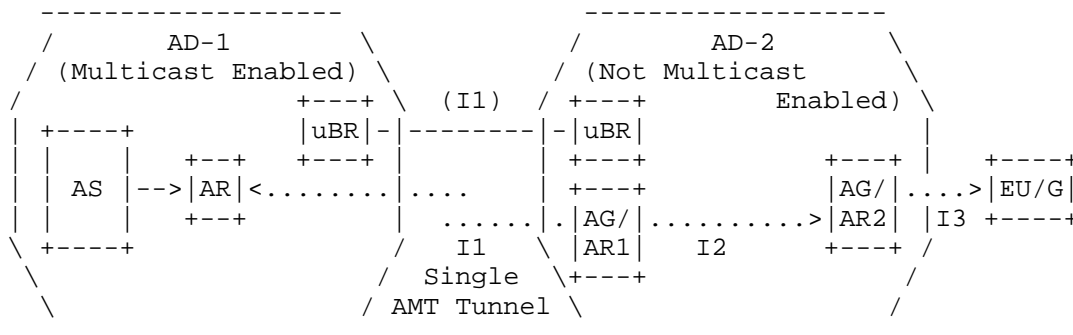   gateway needs to be able to find the "correct" AMT relay in AD-1.

   Architectural guidelines for this configuration are as follows:

   Guidelines (a) through (c) are the same as those described in
   Use Case 3.1. In addition,

   d. It is necessary that proper procedures be implemented such that
      the AMT gateway at the EU device is able to find the correct AMT
      relay for each (S,G) content stream. Standard mechanisms for
      that selection are still subject to ongoing work. This includes
      the use of anycast gateway addresses, anycast DNS names, or
      explicit configuration that maps (S,G) to a relay address; or
      letting the application in the EU/G provide the relay address to
      the embedded AMT gateway function.

   e. The AMT tunnel's capabilities are expected to be sufficient for
      the purpose of collecting relevant information on the multicast
      streams delivered to EUs in AD-2.

3.5. AD-2 Not Multicast Enabled - Multiple AMT Tunnels through AD-2

   Figure 5 illustrates a variation of Use Case 3.4:

```
       ------------------             ------------------
      /      AD-1        \           /      AD-2         \
     / (Multicast Enabled) \       / (Not Multicast      \
    /              +---+ \  (I1)  / +---+       Enabled)   \
   | +----+        |uBR|-|--------|-|uBR|                   |
   | |    |  +--+  +---+ |        | +---+        +---+ |  +----+
   | | AS |-->|AR|<........|....   | +---+       |AG/|....>|EU/G|
   | |    |  +--+      |   ......|.|AG/|..........>|AR2| |I3 +----+
   \ +----+          /   I1  \  |AR1|   I2      +---+ /
    \               /  Single \+---+                 /
     \             / AMT Tunnel \                   /
      ------------------             ------------------
```

   uBR = unicast Border Router - not multicast enabled;
         also, either AR = uBR (AD-1) or uBR = AGAR1 (AD-2)
   AS = multicast (e.g., content) Application Source
   AR = AMT Relay in AD-1
   AGAR1 = AMT Gateway/Relay node in AD-2 across peering point
   I1 = AMT tunnel connecting AR in AD-1 to gateway in AGAR1 in AD-2
   AGAR2 = AMT Gateway/Relay node at AD-2 network edge
   I2 = AMT tunnel connecting relay in AGAR1 to gateway in AGAR2
   EU/G = Gateway client embedded in EU device
   I3 = AMT tunnel connecting EU/G to AR in AGAR2

           Figure 5: AMT Tunnel Connecting AMT Gateways and Relays

   Use Case 3.4 results in several long AMT tunnels crossing the entire
   network of AD-2 linking the EU device and the AMT relay in AD-1
   through the peering point. Depending on the number of EUs, there is
   a likelihood of an unacceptably high amount of traffic due to the
   large number of AMT tunnels -- and unicast streams -- through the
   peering point. This situation can be alleviated as follows:

   o Provisioning of strategically located AMT nodes in AD-2. An
      AMT node comprises co-location of an AMT gateway and an AMT relay.
      No change is required by AD-1, as compared to Use Case 3.4. This
      can be done whenever AD-2 sees fit (e.g., too much traffic across
      the peering point).

   o One such node is on the AD-2 side of the peering point (AMT node
      AGAR1 in Figure 5).

   o A single AMT tunnel established across the peering point linking
      the AMT relay in AD-1 to the AMT gateway in AMT node AGAR1
      in AD-2.

   o AMT tunnels linking AMT node AGAR1 at the peering point in AD-2 to
      other AMT nodes located at the edges of AD-2: e.g., AMT tunnel I2
      linking the AMT relay in AGAR1 to the AMT gateway in AMT
      node AGAR2 (Figure 5).

   o AMT tunnels linking an EU device (via a gateway client embedded in
      the device) and an AMT relay in an appropriate AMT node at the
      edge of AD-2: e.g., I3 linking the EU gateway in the device to the
      AMT relay in AMT node AGAR2.

   o In the simplest option (not shown), AD-2 only deploys a single
      AGAR1 node and lets the EU/G build AMT tunnels directly to it.
      This setup already solves the problem of replicated traffic across
      the peering point. As soon as there is a need to support more AMT
      tunnels to the EU/G, then additional AGAR2 nodes can be deployed
      by AD-2.

   The advantage of such a chained set of AMT tunnels is that the total
   number of unicast streams across AD-2 is significantly reduced, thus
   freeing up bandwidth. Additionally, there will be a single unicast
   stream across the peering point instead of, possibly, an unacceptably
   large number of such streams per Use Case 3.4. However, this implies
   that several AMT tunnels will need to be dynamically configured by
   the various AMT gateways, based solely on the (S,G) information
   received from the application client at the EU device. A suitable
   mechanism for such dynamic configurations is therefore critical.

   Architectural guidelines for this configuration are as follows:

   Guidelines (a) through (c) are the same as those described in
   Use Case 3.1. In addition,

   d. It is necessary that proper procedures be implemented such that
       the various AMT gateways (at the EU devices and the AMT nodes in
       AD-2) are able to find the correct AMT relay in other AMT nodes
       as appropriate. Standard mechanisms for that selection are still
       subject to ongoing work. This includes the use of anycast
       gateway addresses, anycast DNS names, or explicit configuration
       that maps (S,G) to a relay address. On the EU/G, this mapping
       information may come from the application.

   e. The AMT tunnel's capabilities are expected to be sufficient for
       the purpose of collecting relevant information on the multicast
       streams delivered to EUs in AD-2.

4. Functional Guidelines

   Supporting functions and related interfaces over the peering point
   that enable the multicast transport of the application are listed in
   this section. Critical information parameters that need to be
   exchanged in support of these functions are enumerated, along with
   guidelines as appropriate. Specific interface functions for
   consideration are as follows.

4.1. Network Interconnection Transport Guidelines

   The term "network interconnection transport" refers to the
   interconnection points between the two administrative domains. The
   following is a representative set of attributes that the two
   administrative domains will need to agree on to support multicast
   delivery.

   o Number of peering points.

   o Peering point addresses and locations.

   o Connection type - Dedicated for multicast delivery or shared with
      other services.

   o Connection mode - Direct connectivity between the two ADs or via
      another ISP.

   o Peering point protocol support - Multicast protocols that will be
      used for multicast delivery will need to be supported at these
      points. Examples of such protocols include External BGP (EBGP)
      [RFC4760] peering via MP-BGP (Multiprotocol BGP) SAFI-2 [RFC4760].

   o Bandwidth allocation - If shared with other services, then there
      needs to be a determination of the share of bandwidth reserved for
      multicast delivery. See Section 4.1.1 below for more details.

   o QoS requirements - Delay and/or latency specifications that need
      to be specified in an SLA.

   o AD roles and responsibilities - The role played by each AD for
      provisioning and maintaining the set of peering points to support
      multicast delivery.

4.1.1. Bandwidth Management

   Like IP unicast traffic, IP multicast traffic carried across
   non-controlled networks must comply with congestion control
   principles as described in [BCP41] and as explained in detail for UDP
   IP multicast in [BCP145].

   Non-controlled networks (such as the Internet) are networks where
   there is no policy for managing bandwidth other than best effort with
   a fair share of bandwidth under congestion. As a simplified rule of
   thumb, complying with congestion control principles means reducing
   bandwidth under congestion in a way that is fair to competing
   (typically TCP) flows ("rate adaptive").

   In many instances, multicast content delivery evolves from
   intra-domain deployments where it is handled as a controlled network
   service and does not comply with congestion control principles. It
   was given a reserved amount of bandwidth and admitted to the network
   so that congestion never occurs. Therefore, the congestion control
   issue should be given specific attention when evolving to an
   inter-domain peering deployment.

   In the case where end-to-end IP multicast traffic passes across the
   network of two ADs (and their subsidiaries/customers), both ADs must
   agree on a consistent traffic-management policy. If, for example,
   AD-1 sources non-congestion-aware IP multicast traffic and AD-2
   carries it as best-effort traffic across links shared with other
   Internet traffic (subject to congestion), this will not work: under
   congestion, some amount of that traffic will be dropped, often
   rendering the remaining packets as undecodable garbage clogging up
   the network in AD-2; because this traffic is not congestion aware,
   the loss does not reduce this rate. Competing traffic will not get
   their fair share under congestion, and EUs will be frustrated by the
   extremely bad quality of both their IP multicast traffic and other
   (e.g., TCP) traffic. Note that this is not an IP multicast
   technology issue but is solely a transport-layer / application-layer
   issue: the problem would just as likely happen if AD-1 were to send
   non-rate-adaptive unicast traffic -- for example, legacy IPTV
   video-on-demand traffic, which is typically also non-congestion
   aware. Note that because rate adaptation in IP unicast video is
   commonplace today due to the availability of ABR (Adaptive Bitrate)
   video, it is very unlikely that this will happen in reality with IP
   unicast.

   While the rules for traffic management apply whether IP multicast is
   tunneled or not, the one feature that can make AMT tunnels more
   difficult is the unpredictability of bandwidth requirements across
   underlying links because of the way they can be used: with native IP

multicast or GRE tunnels, the amount of bandwidth depends on the
amount of content -- not the number of EUs -- and is therefore easier
to plan for. AMT tunnels terminating in the EU/G, on the other hand,
scale with the number of EUs. In the vicinity of the AMT relay, they
can introduce a very large amount of replicated traffic, and it is
not always feasible to provision enough bandwidth for all possible
EUs to get the highest quality for all their content during peak
utilization in such setups -- unless the AMT relays are very close to
the EU edge. Therefore, it is also recommended that IP multicast
rate adaptation be used, even inside controlled networks, when using
AMT tunnels directly to the EU/G.

Note that rate-adaptive IP multicast traffic in general does not mean
that the sender is reducing the bitrate but rather that the EUs that
experience congestion are joining to a lower-bitrate (S,G) stream of
the content, similar to ABR streaming over TCP. Therefore, migration
from a non-rate-adaptive bitrate to a rate-adaptive bitrate in IP
multicast will also change the dynamic (S,G) join behavior in the
network, resulting in potentially higher performance requirements for
IP multicast protocols (IGMP/PIM), especially on the last hops where
dynamic changes occur (including AMT gateways/relays): in non-rate-
adaptive IP multicast, only "channel change" causes state change, but
in rate-adaptive multicast, congestion also causes state change.

Even though not fully specified in this document, peerings that rely
on GRE/AMT tunnels may be across one or more transit ADs instead of
an exclusive (non-shared, L1/L2) path. Unless those transit ADs are
explicitly contracted to provide other than "best effort" transit for
the tunneled traffic, the tunneled IP multicast traffic must be
rate adaptive in order to not violate BCP 41 across those
transit ADs.

4.2. Routing Aspects and Related Guidelines

The main objective for multicast delivery routing is to ensure that
the EU receives the multicast stream from the "most optimal" source
[INF_ATIS_10], which typically:

o Maximizes the multicast portion of the transport and minimizes any
   unicast portion of the delivery, and

o Minimizes the overall combined route distance of the network(s).

This routing objective applies to both native multicast and AMT; the actual methodology of the solution will be different for each. Regardless, the routing solution is expected to:

o Be scalable,

o Avoid or minimize new protocol development or modifications, and

o Be robust enough to achieve high reliability and to automatically adjust to changes and problems in the multicast infrastructure.

For both native and AMT environments, having a source as close as possible to the EU network is most desirable; therefore, in some cases, an AD may prefer to have multiple sources near different peering points. However, that is entirely an implementation issue.

4.2.1. Native Multicast Routing Aspects

Native multicast simply requires that the administrative domains coordinate and advertise the correct source address(es) at their network interconnection peering points (i.e., BRs). An example of multicast delivery via a native multicast process across two administrative domains is as follows, assuming that the interconnecting peering points are also multicast enabled:

o Appropriate information is obtained by the EU client, who is a subscriber to AD-2 (see Use Case 3.1). This information is in the form of metadata, and it contains instructions directing the EU client to launch an appropriate application if necessary, as well as additional information for the application about the source location and the group (or stream) ID in the form of (S,G) data. The "S" portion provides the name or IP address of the source of the multicast stream. The metadata may also contain alternate delivery information, such as specifying the unicast address of the stream.

o The client uses the join message with (S,G) to join the multicast stream [RFC4604]. To facilitate this process, the two ADs need to do the following:

   * Advertise the source ID(s) over the peering points.

   * Exchange such relevant peering point information as capacity and utilization.

   * Implement compatible multicast protocols to ensure proper multicast delivery across the peering points.

4.2.2. GRE Tunnel over Interconnecting Peering Point

   If the interconnecting peering point is not multicast enabled and
   both ADs are multicast enabled, then a simple solution is to
   provision a GRE tunnel between the two ADs; see Use Case 3.2
   (Section 3.2). The termination points of the tunnel will usually be
   a network engineering decision but generally will be between the BRs
   or even between the AD-2 BR and the AD-1 source (or source access
   router). The GRE tunnel would allow end-to-end native multicast or
   AMT multicast to traverse the interface. Coordination and
   advertisement of the source IP are still required.

   The two ADs need to follow the same process as the process described
   in Section 4.2.1 to facilitate multicast delivery across the peering
   points.

4.2.3. Routing Aspects with AMT Tunnels

   Unlike native multicast (with or without GRE), an AMT multicast
   environment is more complex. It presents a two-layered problem
   in that there are two criteria that should be simultaneously met:

   o Find the closest AMT relay to the EU that also has multicast
      connectivity to the content source, and

   o Minimize the AMT unicast tunnel distance.

   There are essentially two components in the AMT specification:

   AMT relays: These serve the purpose of tunneling UDP multicast
      traffic to the receivers (i.e., endpoints). The AMT relay will
      receive the traffic natively from the multicast media source and
      will replicate the stream on behalf of the downstream AMT
      gateways, encapsulating the multicast packets into unicast packets
      and sending them over the tunnel toward the AMT gateways. In
      addition, the AMT relay may collect various usage and activity
      statistics. This results in moving the replication point closer
      to the EU and cuts down on traffic across the network. Thus, the
      linear costs of adding unicast subscribers can be avoided.
      However, unicast replication is still required for each requesting
      endpoint within the unicast-only network.

   AMT gateway: The gateway will reside on an endpoint; this could be
      any type of IP host, such as a Personal Computer (PC), mobile
      phone, Set-Top Box (STB), or appliances. The AMT gateway receives
      join and leave requests from the application via an Application
      Programming Interface (API). In this manner, the gateway allows
      the endpoint to conduct itself as a true multicast endpoint. The

AMT gateway will encapsulate AMT messages into UDP packets and
send them through a tunnel (across the unicast-only
infrastructure) to the AMT relay.

The simplest AMT use case (Section 3.3) involves peering points that
are not multicast enabled between two multicast-enabled ADs. An
AMT tunnel is deployed between an AMT relay on the AD-1 side of the
peering point and an AMT gateway on the AD-2 side of the peering
point. One advantage of this arrangement is that the tunnel is
established on an as-needed basis and need not be a provisioned
element. The two ADs can coordinate and advertise special AMT relay
anycast addresses with, and to, each other. Alternately, they may
decide to simply provision relay addresses, though this would not be
an optimal solution in terms of scalability.

Use Cases 3.4 and 3.5 describe AMT situations that are more
complicated, as AD-2 is not multicast enabled in these two cases.
For these cases, the EU device needs to be able to set up an AMT
tunnel in the most optimal manner. There are many methods by which
relay selection can be done, including the use of DNS-based queries
and static lookup tables [RFC7450]. The choice of the method is
implementation dependent and is up to the network operators.
Comparison of various methods is out of scope for this document and
is left for further study.

An illustrative example of a relay selection based on DNS queries as
part of an anycast IP address process is described here for Use
Cases 3.4 and 3.5 (Sections 3.4 and 3.5). Using an anycast
IP address for AMT relays allows all AMT gateways to find the
"closest" AMT relay -- the nearest edge of the multicast topology of
the source. Note that this is strictly illustrative; the choice of
the method is up to the network operators. The basic process is as
follows:

o Appropriate metadata is obtained by the EU client application.
   The metadata contains instructions directing the EU client to an
   ordered list of particular destinations to seek the requested
   stream and, for multicast, specifies the source location and the
   group (or stream) ID in the form of (S,G) data. The "S" portion
   provides the URI (name or IP address) of the source of the
   multicast stream, and the "G" identifies the particular stream
   originated by that source. The metadata may also contain
   alternate delivery information such as the address of the unicast
   form of the content to be used -- for example, if the multicast
   stream becomes unavailable.

o Using the information from the metadata and, possibly, information
   provisioned directly in the EU client, a DNS query is initiated in
   order to connect the EU client / AMT gateway to an AMT relay.

o Query results are obtained and may return an anycast address or a
   specific unicast address of a relay. Multiple relays will
   typically exist. The anycast address is a routable
   "pseudo-address" shared among the relays that can gain multicast
   access to the source.

o If a specific IP address unique to a relay was not obtained, the
   AMT gateway then sends a message (e.g., the discovery message) to
   the anycast address such that the network is making the routing
   choice of a particular relay, e.g., the relay that is closest to
   the EU. Details are outside the scope of this document. See
   [RFC4786].

o The contacted AMT relay then returns its specific unicast IP
   address (after which the anycast address is no longer required).
   Variations may exist as well.

o The AMT gateway uses that unicast IP address to initiate a
   three-way handshake with the AMT relay.

o The AMT gateway provides the (S,G) information to the AMT relay
   (embedded in AMT protocol messages).

o The AMT relay receives the (S,G) information and uses it to join
   the appropriate multicast stream, if it has not already subscribed
   to that stream.

o The AMT relay encapsulates the multicast stream into the tunnel
   between the relay and the gateway, providing the requested content
   to the EU.

4.2.4. Public Peering Routing Aspects

   Figure 6 shows an example of a broadcast peering point.

```
         AD-1a                AD-1b
          BR                   BR
           |                    |
        --+-+---------------+-+-- broadcast peering point LAN
           |                    |
          BR                   BR
         AD-2a                AD-2b
```
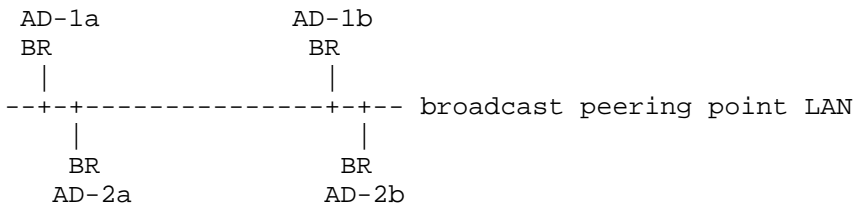
                  Figure 6: Broadcast Peering Point

A broadcast peering point is an L2 subnet connecting three or more
ADs. It is common in IXPs and usually consists of Ethernet
switch(es) operated by the IXP connecting to BRs operated by the ADs.

In an example setup domain, AD-2a peers with AD-1a and wants to
receive IP multicast from it. Likewise, AD-2b peers with AD-1b and
wants to receive IP multicast from it.

Assume that one or more IP multicast (S,G) traffic streams can be
served by both AD-1a and AD-1b -- for example, because both AD-1a and
AD-1b contact this content from the same content source.

In this case, AD-2a and AD-2b can no longer control which upstream
domain -- AD-1a or AD-1b -- will forward this (S,G) into the LAN.
The AD-2a BR requests the (S,G) from the AD-1a BR, and the AD-2b BR
requests the same (S,G) from the AD-1b BR. To avoid duplicate
packets, an (S,G) can be forwarded by only one router onto the LAN;
PIM-SM / PIM-SSM detects requests for duplicate transmissions and
resolves them via the so-called "assert" protocol operation, which
results in only one BR forwarding the traffic. Assume that this is
the AD-1a BR. AD-2b will then receive unexpected multicast traffic
from a provider with whom it does not have a mutual agreement for
that traffic. Quality issues in EUs behind AD-2b caused by AD-1a
will cause a lot of issues related to responsibility and
troubleshooting.

In light of these technical issues, we describe, via the following
options, how IP multicast can be carried across broadcast peering
point LANs:

1. IP multicast is tunneled across the LAN. Any of the GRE/AMT
    tunneling solutions mentioned in this document are applicable.
    This is the one case where a GRE tunnel between the upstream BR
    (e.g., AD-1a) and downstream BR (e.g., AD-2a) is specifically
    recommended, as opposed to tunneling across uBRs (which are not
    the actual BRs).

2. The LAN has only one upstream AD that is sourcing IP multicast,
    and native IP multicast is used. This is an efficient way to
    distribute the same IP multicast content to multiple downstream
    ADs. Misbehaving downstream BRs can still disrupt the delivery
    of IP multicast from the upstream BR to other downstream BRs;
    therefore, strict rules must be followed to prohibit such a case.
    The downstream BRs must ensure that they will always consider
    only the upstream BR as a source for multicast traffic: e.g., no
    BGP SAFI-2 peerings between the downstream ADs across the peering
    point LAN, so that the upstream BR is the only possible next hop
    reachable across this LAN. Also, routing policies can be

      configured to avoid falling back to using SAFI-1 (unicast) routes
      for IP multicast if unicast BGP peering is not limited in the
      same way.

   3. The LAN has multiple upstream ADs, but they are federated and
       agree on a consistent policy for IP multicast traffic across the
       LAN. One policy is that each possible source is only announced
       by one upstream BR. Another policy is that sources are
       redundantly announced (the problematic case mentioned in the
       example in Figure 6 above), but the upstream domains also provide
       mutual operational insight to help with troubleshooting (outside
       the scope of this document).

4.3. Back-Office Functions - Provisioning and Logging Guidelines

   "Back office" refers to the following:

   o Servers and content-management systems that support the delivery
      of applications via multicast and interactions between ADs.

   o Functionality associated with logging, reporting, ordering,
      provisioning, maintenance, service assurance, settlement, etc.

4.3.1. Provisioning Guidelines

   Resources for basic connectivity between ADs' providers need to be
   provisioned as follows:

   o Sufficient capacity must be provisioned to support multicast-based
      delivery across ADs.

   o Sufficient capacity must be provisioned for connectivity between
      all supporting back offices of the ADs as appropriate. This
      includes activating proper security treatment for these
      back-office connections (gateways, firewalls, etc.) as
      appropriate.

   Provisioning aspects related to multicast-based inter-domain delivery
   are as follows.

   The ability to receive a requested application via multicast is
   triggered via receipt of the necessary metadata. Hence, this
   metadata must be provided to the EU regarding the multicast URL --
   and unicast fallback if applicable. AD-2 must enable the delivery of
   this metadata to the EU and provision appropriate resources for this
   purpose.

It is assumed that native multicast functionality is available across
many ISP backbones, peering points, and access networks. If,
however, native multicast is not an option (Use Cases 3.4 and 3.5),
then:

o The EU must have a multicast client to use AMT multicast obtained
   from either (1) the application source (per agreement with AD-1)
   or (2) AD-1 or AD-2 (if delegated by the application source).

o If provided by AD-1 or AD-2, then the EU could be redirected to a
   client download site. (Note: This could be an application source
   site.) If provided by the application source, then this source
   would have to coordinate with AD-1 to ensure that the proper
   client is provided (assuming multiple possible clients).

o Where AMT gateways support different application sets, all AD-2
   AMT relays need to be provisioned with all source and group
   addresses for streams it is allowed to join.

o DNS across each AD must be provisioned to enable a client gateway
   to locate the optimal AMT relay (i.e., longest multicast path and
   shortest unicast tunnel) with connectivity to the content's
   multicast source.

Provisioning aspects related to operations and customer care are as
follows.

It is assumed that each AD provider will provision operations and
customer care access to their own systems.

AD-1's operations and customer care functions must be able to see
enough of what is happening in AD-2's network or in the service
provided by AD-2 to verify their mutual goals and operations, e.g.,
to know how the EUs are being served. This can be done in two ways:

o Automated interfaces are built between AD-1 and AD-2 such that
   operations and customer care continue using their own systems.
   This requires coordination between the two ADs, with appropriate
   provisioning of necessary resources.

o AD-1's operations and customer care personnel are provided direct
   access to AD-2's systems. In this scenario, additional
   provisioning in these systems will be needed to provide necessary
   access. The two ADs must agree on additional provisioning to
   support this option.

4.3.2. Inter-domain Authentication Guidelines

   All interactions between pairs of ADs can be discovered and/or
   associated with the account(s) utilized for delivered applications.
   Supporting guidelines are as follows:

   o A unique identifier is recommended to designate each master
      account.

   o AD-2 is expected to set up "accounts" (a logical facility
      generally protected by credentials such as login passwords) for
      use by AD-1. Multiple accounts, and multiple types or partitions
      of accounts, can apply, e.g., customer accounts, security
      accounts.

   The reason to specifically mention the need for AD-1 to initiate
   interactions with AD-2 (and use some account for that), as opposed to
   the opposite, is based on the recommended workflow initiated by
   customers (see Section 4.4): the customer contacts the content
   source, which is part of AD-1. Consequently, if AD-1 sees the need
   to escalate the issue to AD-2, it will interact with AD-2 using the
   aforementioned guidelines.

4.3.3. Log-Management Guidelines

   Successful delivery (in terms of user experience) of applications or
   content via multicast between pairs of interconnecting ADs can be
   improved through the ability to exchange appropriate logs for various
   workflows -- troubleshooting, accounting and billing, optimization of
   traffic and content transmission, optimization of content and
   application development, and so on.

   Specifically, AD-1 take over primary responsibility for customer
   experience on behalf of the content source, with support from AD-2 as
   needed. The application/content owner is the only participant who
   has, and needs, full insight into the application level and can map
   the customer application experience to the network traffic flows --
   which, with the help of AD-2 or logs from AD-2, it can then analyze
   and interpret.

   The main difference between unicast delivery and multicast delivery
   is that the content source can infer a lot more about downstream
   network problems from a unicast stream than from a multicast stream:
   the multicast stream is not per EU, except after the last
   replication, which is in most cases not in AD-1. Logs from the
   application, including the receiver side at the EU, can provide
   insight but cannot help to fully isolate network problems because of

the IP multicast per-application operational state built across AD-1
and AD-2 (aka the (S,G) state and any other operational-state
features, such as Diffserv QoS).

See Section 7 for more discussion regarding the privacy
considerations of the model described here.

Different types of logs are known to help support operations in AD-1
when provided by AD-2. This could be done as part of AD-1/AD-2
contracts. Note that except for implied multicast-specific elements,
the options listed here are not unique or novel for IP multicast, but
they are more important for services novel to the operators than for
operationally well-established services (such as unicast). We
therefore detail them as follows:

o Usage information logs at an aggregate level.

o Usage failure instances at an aggregate level.

o Grouped or sequenced application access: performance, behavior,
   and failure at an aggregate level to support potential
   application-provider-driven strategies. Examples of aggregate
   levels include grouped video clips, web pages, and software-
   download sets.

o Security logs, aggregated or summarized according to agreement
   (with additional detail potentially provided during security
   events, by agreement).

o Access logs (EU), when needed for troubleshooting.

o Application logs ("What is the application doing?"), when needed
   for shared troubleshooting.

o Syslogs (network management), when needed for shared
   troubleshooting.

The two ADs may supply additional security logs to each other, as
agreed upon in contract(s). Examples include the following:

o Information related to general security-relevant activity, which
   may be of use from a protection or response perspective: types and
   counts of attacks detected, related source information, related
   target information, etc.

o Aggregated or summarized logs according to agreement (with
   additional detail potentially provided during security events, by
   agreement).

4.4. Operations - Service Performance and Monitoring Guidelines

   "Service performance" refers to monitoring metrics related to
   multicast delivery via probes. The focus is on the service provided
   by AD-2 to AD-1 on behalf of all multicast application sources
   (metrics may be specified for SLA use or otherwise). Associated
   guidelines are as follows:

   o Both ADs are expected to monitor, collect, and analyze service
      performance metrics for multicast applications. AD-2 provides
      relevant performance information to AD-1; this enables AD-1 to
      create an end-to-end performance view on behalf of the multicast
      application source.

   o Both ADs are expected to agree on the types of probes to be used
      to monitor multicast delivery performance. For example, AD-2 may
      permit AD-1's probes to be utilized in the AD-2 multicast service
      footprint. Alternately, AD-2 may deploy its own probes and relay
      performance information back to AD-1.

   "Service monitoring" generally refers to a service (as a whole)
   provided on behalf of a particular multicast application source
   provider. It thus involves complaints from EUs when service problems
   occur. EUs direct their complaints to the source provider; the
   source provider in turn submits these complaints to AD-1. The
   responsibility for service delivery lies with AD-1; as such, AD-1
   will need to determine where the service problem is occurring -- in
   its own network or in AD-2. It is expected that each AD will have
   tools to monitor multicast service status in its own network.

   o Both ADs will determine how best to deploy multicast service
      monitoring tools. Typically, each AD will deploy its own set of
      monitoring tools, in which case both ADs are expected to inform
      each other when multicast delivery problems are detected.

   o AD-2 may experience some problems in its network. For example,
      for the AMT use cases (Sections 3.3, 3.4, and 3.5), one or more
      AMT relays may be experiencing difficulties. AD-2 may be able to
      fix the problem by rerouting the multicast streams via alternate
      AMT relays. If the fix is not successful and multicast service
      delivery degrades, then AD-2 needs to report the issue to AD-1.

   o When a problem notification is received from a multicast
      application source, AD-1 determines whether the cause of the
      problem is within its own network or within AD-2. If the cause is
      within AD-2, then AD-1 supplies all necessary information to AD-2.
      Examples of supporting information include the following:

      * Kind(s) of problem(s).

      * Starting point and duration of problem(s).

      * Conditions in which one or more problems occur.

      * IP address blocks of affected users.

      * ISPs of affected users.

      * Type of access, e.g., mobile versus desktop.

      * Network locations of affected EUs.

   o Both ADs conduct some form of root-cause analysis for multicast
      service delivery problems. Examples of various factors for
      consideration include:

      * Verification that the service configuration matches the product
         features.

      * Correlation and consolidation of the various customer problems
         and resource troubles into a single root-service problem.

      * Prioritization of currently open service problems, giving
         consideration to problem impacts, SLAs, etc.

      * Conducting service tests, including tests performed once or a
         series of tests over a period of time.

      * Analysis of test results.

      * Analysis of relevant network fault or performance data.

      * Analysis of the problem information provided by the customer.

   o Once the cause of the problem has been determined and the problem
      has been fixed, both ADs need to work jointly to verify and
      validate the success of the fix.

4.5. Client Reliability Models / Service Assurance Guidelines

   There are multiple options for instituting reliability architectures.
   Most are at the application level. Both ADs should work these
   options out per their contract or agreement and also with the
   multicast application source providers.

   Network reliability can also be enhanced by the two ADs if they
   provision alternate delivery mechanisms via unicast means.

4.6. Application Accounting Guidelines

   Application-level accounting needs to be handled differently in the
   application than in IP unicast, because the source side does not
   directly deliver packets to individual receivers. Instead, this
   needs to be signaled back by the receiver to the source.

   For network transport diagnostics, AD-1 and AD-2 should have
   mechanisms in place to ensure proper accounting for the volume of
   bytes delivered through the peering point and, separately, the number
   of bytes delivered to EUs.

5. Troubleshooting and Diagnostics

   Any service provider supporting multicast delivery of content should
   be able to collect diagnostics as part of multicast troubleshooting
   practices and resolve network issues accordingly. Issues may become
   apparent or identifiable through either (1) network monitoring
   functions or (2) problems reported by customers, as described in
   Section 4.4.

   It is recommended that multicast diagnostics be performed, leveraging
   established operational practices such as those documented in
   [MDH-05]. However, given that inter-domain multicast creates a
   significant interdependence of proper networking functionality
   between providers, there exists a need for providers to be able to
   signal (or otherwise alert) each other if there are any issues noted
   by either one.

   For troubleshooting purposes, service providers may also wish to
   allow limited read-only administrative access to their routers to
   their AD peers. Access to active troubleshooting tools -- especially
   [Traceroute] and the tools discussed in [Mtrace-v2] -- is of specific
   interest.

Another option is to include this functionality in the IP multicast
receiver application on the EU device and allow these diagnostics to
be remotely used by support operations. Note, though, that AMT
does not allow the passing of traceroute or mtrace requests;
therefore, troubleshooting in the presence of AMT does not work as
well end to end as it can with native (or even GRE-encapsulated) IP
multicast, especially with regard to traceroute and mtrace. Instead,
troubleshooting directly on the actual network devices is then more
likely necessary.

The specifics of notifications and alerts are beyond the scope of
this document, but general guidelines are similar to those described
in Section 4.4. Some general communications issues are as follows.

o Appropriate communications channels will be established between
   the customer service and operations groups from both ADs to
   facilitate information-sharing related to diagnostic
   troubleshooting.

o A default resolution period may be considered to resolve open
   issues. Alternately, mutually acceptable resolution periods could
   be established, depending on the severity of the identified
   trouble.

6. Security Considerations

6.1. DoS Attacks (against State and Bandwidth)

Reliable IP multicast operations require some basic protection
against DoS (Denial of Service) attacks.

SSM IP multicast is self-protecting against attacks from illicit
sources; such traffic will not be forwarded beyond the first-hop
router, because that would require (S,G) membership reports from the
receiver. Only valid traffic from sources will be forwarded, because
RPF ("Reverse Path Forwarding") is part of the protocols. One can
say that protection against spoofed source traffic performed in the
style of [BCP38] is therefore built into PIM-SM / PIM-SSM.

Receivers can attack SSM IP multicast by originating such (S,G)
membership reports. This can result in a DoS attack against state
through the creation of a large number of (S,G) states that create
high control-plane load or even inhibit the later creation of a valid
(S,G). In conjunction with collaborating illicit sources, it can
also result in the forwarding of traffic from illicit sources.

Today, these types of attacks are usually mitigated by explicitly
defining the set of permissible (S,G) on, for example, the last-hop
routers in replicating IP multicast to EUs (e.g., via (S,G) access
control lists applied to IGMP/MLD membership state creation). Each
AD (say, "ADi") is expected to know what sources located in ADi are
permitted to send and what their valid (S,G)s are. ADi can therefore
also filter invalid (S,G)s for any "S" located inside ADi, but not
sources located in another AD.

In the peering case, without further information, AD-2 is not aware
of the set of valid (S,G) from AD-1, so this set needs to be
communicated via operational procedures from AD-1 to AD-2 to provide
protection against this type of DoS attack. Future work could signal
this information in an automated way: BGP extensions, DNS resource
records, or backend automation between AD-1 and AD-2. Backend
automation is, in the short term, the most viable solution: unlike
BGP extensions or DNS resource records, backend automation does not
require router software extensions. Observation of traffic flowing
via (S,G) state could also be used to automate the recognition of
invalid (S,G) state created by receivers in the absence of explicit
information from AD-1.

The second type of DoS attack through (S,G) membership reports exists
when the attacking receiver creates too much valid (S,G) state and
the traffic carried by these (S,G)s congests bandwidth on links
shared with other EUs. Consider the uplink to a last-hop router
connecting to 100 EUs. If one EU joins to more multicast content
than what fits into this link, then this would also impact the
quality of the same content for the other 99 EUs. If traffic is not
rate adaptive, the effects are even worse.

The mitigation technique is the same as what is often employed for
unicast: policing of the per-EU total amount of traffic. Unlike
unicast, though, this cannot be done anywhere along the path (e.g.,
on an arbitrary bottleneck link); it has to happen at the point of
last replication to the different EU. Simple solutions such as
limiting the maximum number of joined (S,G)s per EU are readily
available; solutions that take consumed bandwidth into account are
available as vendor-specific features in routers. Note that this is
primarily a non-peering issue in AD-2; it only becomes a peering
issue if the peering link itself is not big enough to carry all
possible content from AD-1 or, as in Use Case 3.4, when the AMT relay
in AD-1 is that last replication point.

Limiting the amount of (S,G) state per EU is also a good first
measure to prohibit too much undesired "empty" state from being built
(state not carrying traffic), but it would not suffice in the case of
DDoS attacks, e.g., viruses that impact a large number of EU devices.

6.2. Content Security

   Content confidentiality, DRM (Digital Rights Management),
   authentication, and authorization are optional, based on the content
   delivered. For content that is "FTA" (Free To Air), the following
   considerations can be ignored, and content can be sent unencrypted
   and without EU authentication and authorization. Note, though, that
   the mechanisms described here may also be desirable for the
   application source to better track users even if the content itself
   would not require it.

   For inter-domain content, there are at least two models for content
   confidentiality, including (1) DRM authentication and authorization
   and (2) EU authentication and authorization:

   o In the classical (IP)TV model, responsibility is per domain, and
      content is and can be passed on unencrypted. AD-1 delivers
      content to AD-2; AD-2 can further process the content, including
      features like ad insertion, and AD-2 is the sole point of contact
      regarding the contact for its EUs. In this document, we do not
      consider this case because it typically involves service aspects
      operated by AD-2 that are higher than the network layer; this
      document focuses on the network-layer AD-1/AD-2 peering case but
      not the application-layer peering case. Nevertheless, this model
      can be derived through additional work beyond what is described
      here.

   o The other model is the one in which content confidentiality, DRM,
      EU authentication, and EU authorization are end to end:
      responsibilities of the multicast application source provider and
      receiver application. This is the model assumed here. It is also
      the model used in Internet "Over the Top" (OTT) video delivery.
      Below, we discuss the threats incurred in this model due to the
      use of IP multicast in AD-1 or AD-2 and across the peering point.

   End-to-end encryption enables end-to-end EU authentication and
   authorization: the EU may be able to join (via IGMP/MLD) and receive
   the content, but it can only decrypt it when it receives the
   decryption key from the content source in AD-1. The key is the
   authorization. Keeping that key to itself and prohibiting playout of
   the decrypted content to non-copy-protected interfaces are typical
   DRM features in that receiver application or EU device operating
   system.

   End-to-end encryption is continuously attacked. Keys may be subject
   to brute-force attacks so that content can potentially be decrypted
   later, or keys are extracted from the EU application/device and
   shared with other unauthenticated receivers. One important class of

content is where the value is in live consumption, such as sports or
other event (e.g., concert) streaming. Extraction of keying material
from compromised authenticated EUs and sharing with unauthenticated
EUs are not sufficient. It is also necessary for those
unauthenticated EUs to get a streaming copy of the content itself.
In unicast streaming, they cannot get such a copy from the content
source (because they cannot authenticate), and, because of asymmetric
bandwidths, it is often impossible to get the content from
compromised EUs to a large number of unauthenticated EUs. EUs behind
classical "16 Mbps down, 1 Mbps up" ADSL links are the best example.
With increasing broadband access speeds, unicast peer-to-peer copying
of content becomes easier, but it likely will always be easily
detectable by the ADs because of its traffic patterns and volume.

When IP multicast is being used without additional security, AD-2 is
not aware of which EU is authenticated for which content. Any
unauthenticated EU in AD-2 could therefore get a copy of the
encrypted content without triggering suspicion on the part of AD-2 or
AD-1 and then either (1) live-decode it, in the presence of the
compromised authenticated EU and key-sharing or (2) decrypt it later,
in the presence of federated brute-force key-cracking.

To mitigate this issue, the last replication point that is creating
(S,G) copies to EUs would need to permit those copies only after
authentication of the EUs. This would establish the same
authenticated "EU only" copy that is used in unicast.

Schemes for per-EU IP multicast authentication/authorization (and, as
a result, non-delivery or copying of per-content IP multicast
traffic) have been built in the past and are deployed in service
providers for intra-domain IPTV services, but no standards exist for
this. For example, there is no standardized RADIUS attribute for
authenticating the IGMP/MLD filter set, but such implementations
exist. The authors of this document are specifically also not aware
of schemes where the same authentication credentials used to get the
encryption key from the content source could also be used to
authenticate and authorize the network-layer IP multicast replication
for the content. Such schemes are technically not difficult to build
and would avoid creating and maintaining a separate network
traffic-forwarding authentication/authorization scheme decoupled from
the end-to-end authentication/authorization system of the
application.

If delivery of such high-value content in conjunction with the
peering described here is desired, the short-term recommendations are
for sources to clearly isolate the source and group addresses used
for different content bundles, communicate those (S,G) patterns from
AD-1 to AD-2, and let AD-2 leverage existing per-EU authentication/
authorization mechanisms in network devices to establish filters for
(S,G) sets to each EU.

6.3. Peering Encryption

Encryption at peering points for multicast delivery may be used per
agreement between AD-1 and AD-2.

In the case of a private peering link, IP multicast does not have
attack vectors on a peering link different from those of IP unicast,
but the content owner may have defined strict constraints against
unauthenticated copying of even the end-to-end encrypted content; in
this case, AD-1 and AD-2 can agree on additional transport encryption
across that peering link. In the case of a broadcast peering
connection (e.g., IXP), transport encryption is again the easiest way
to prohibit unauthenticated copies by other ADs on the same peering
point.

If peering is across a tunnel that spans intermittent transit ADs
(not discussed in detail in this document), then encryption of that
tunnel traffic is recommended. It not only prohibits possible
"leakage" of content but also protects the information regarding what
content is being consumed in AD-2 (aggregated privacy protection).

See Section 6.4 for reasons why the peering point may also need to be
encrypted for operational reasons.

6.4. Operational Aspects

Section 4.3.3 discusses the exchange of log information, and
Section 7 discusses the exchange of program information. All these
operational pieces of data should by default be exchanged via
authenticated and encrypted peer-to-peer communication protocols
between AD-1 and AD-2 so that only the intended recipients in the
peers' AD have access to it. Even exposure of the least sensitive
information to third parties opens up attack vectors. Putting valid
(S,G) information, for example, into DNS (as opposed to passing it
via secured channels from AD-1 to AD-2) to allow easier filtering of
invalid (S,G) information would also allow attackers to more easily
identify valid (S,G) information and change their attack vector.

From the perspective of the ADs, security is most critical for log
information, as it provides operational insight into the originating
AD but also contains sensitive user data.

Sensitive user data exported from AD-2 to AD-1 as part of logs could
be as much as the equivalent of 5-tuple unicast traffic flow
accounting (but not more, e.g., no application-level information).
As mentioned in Section 7, in unicast, AD-1 could capture these
traffic statistics itself because this is all about traffic flows
(originated by AD-1) to EU receivers in AD-2, and operationally
passing it from AD-2 to AD-1 may be necessary when IP multicast is
used because of the replication taking place in AD-2.

Nevertheless, passing such traffic statistics inside AD-1 from a
capturing router to a backend system is likely less subject to
third-party attacks than passing it "inter-domain" from AD-2 to AD-1,
so more diligence needs to be applied to secure it.

If any protocols used for the operational exchange of information are
not easily secured at the transport layer or higher (because of the
use of legacy products or protocols in the network), then AD-1 and
AD-2 can also consider ensuring that all operational data exchanges
go across the same peering point as the traffic and use network-layer
encryption of the peering point (as discussed previously) to
protect it.

End-to-end authentication and authorization of EUs may involve some
kind of token authentication and are done at the application layer,
independently of the two ADs. If there are problems related to the
failure of token authentication when EUs are supported by AD-2, then
some means of validating proper operation of the token authentication
process (e.g., validating that backend servers querying the multicast
application source provider's token authentication server are
communicating properly) should be considered. Implementation details
are beyond the scope of this document.

In the event of a security breach, the two ADs are expected to have a
mitigation plan for shutting down the peering point and directing
multicast traffic over alternative peering points. It is also
expected that appropriate information will be shared for the purpose
of securing the identified breach.

7. Privacy Considerations

   The described flow of information about content and EUs as described
   in this document aims to maintain privacy:

   AD-1 is operating on behalf of (or owns) the content source and is
   therefore part of the content-consumption relationship with the EU.
   The privacy considerations between the EU and AD-1 are therefore
   generally the same (with one exception; see below) as they would be
   if no IP multicast was used, especially because end-to-end encryption
   can and should be used for any privacy-conscious content.

   Information related to inter-domain multicast transport service is
   provided to AD-1 by the AD-2 operators. AD-2 is not required to gain
   additional insight into the user's behavior through this process
   other than what it would already have without service collaboration
   with AD-1, unless AD-1 and AD-2 agree on it and get approval from
   the EU.

   For example, if it is deemed beneficial for the EU to get support
   directly from AD-2, then it would generally be necessary for AD-2 to
   be aware of the mapping between content and network (S,G) state so
   that AD-2 knows which (S,G) to troubleshoot when the EU complains
   about problems with specific content. The degree to which this
   dissemination is done by AD-1 explicitly to meet privacy expectations
   of EUs is typically easy to assess by AD-1. Two simple examples are
   as follows:

   o For a sports content bundle, every EU will happily click on the
      "I approve that the content program information is shared with
      your service provider" button, to ensure best service reliability,
      because service-conscious AD-2 would likely also try to ensure
      that high-value content, such as the (S,G) for the Super Bowl,
      would be the first to receive care in the case of network issues.

   o If the content in question was content for which the EU expected
      more privacy, the EU should prefer a content bundle that included
      this content in a large variety of other content, have all content
      end-to-end encrypted, and not share programming information with
      AD-2, to maximize privacy. Nevertheless, the privacy of the EU
      against AD-2 observing traffic would still be lower than in the
      equivalent setup using unicast, because in unicast, AD-2 could not
      correlate which EUs are watching the same content and use that to
      deduce the content. Note that even the setup in Section 3.4,
      where AD-2 is not involved in IP multicast at all, does not
      provide privacy against this level of analysis by AD-2, because
      there is no transport-layer encryption in AMT; therefore, AD-2 can
      correlate by on-path traffic analysis who is consuming the same

content from an AMT relay from both the (S,G) join messages in AMT
and the identical content segments (that were replicated at the
AMT relay).

In summary, because only content to be consumed by multiple EUs is
carried via IP multicast here and all of that content can be
end-to-end encrypted, the only privacy consideration specific to IP
multicast is for AD-2 to know or reconstruct what content an EU is
consuming. For content for which this is undesirable, some form of
protections as explained above are possible, but ideally, the model
described in Section 3.4 could be used in conjunction with future
work, e.g., adding Datagram Transport Layer Security (DTLS)
encryption [RFC6347] between the AMT relay and the EU.

Note that IP multicast by nature would permit the EU's privacy
against the content source operator because, unlike unicast, the
content source does not natively know which EU is consuming which
content: in all cases where AD-2 provides replication, only AD-2
knows this directly. This document does not attempt to describe a
model that maintains such a level of privacy against the content
source; rather, we describe a model that only protects against
exposure to intermediate parties -- in this case, AD-2.

8. IANA Considerations

   This document does not require any IANA actions.

9. References

9.1. Normative References

   [RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P.
              Traina, "Generic Routing Encapsulation (GRE)", RFC 2784,
              DOI 10.17487/RFC2784, March 2000,
              <https://www.rfc-editor.org/info/rfc2784>.

   [RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A.
              Thyagarajan, "Internet Group Management Protocol,
              Version 3", RFC 3376, DOI 10.17487/RFC3376, October 2002,
              <https://www.rfc-editor.org/info/rfc3376>.

   [RFC3810] Vida, R., Ed., and L. Costa, Ed., "Multicast Listener
              Discovery Version 2 (MLDv2) for IPv6", RFC 3810,
              DOI 10.17487/RFC3810, June 2004,
              <https://www.rfc-editor.org/info/rfc3810>.

   [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter,
              "Multiprotocol Extensions for BGP-4", RFC 4760,
              DOI 10.17487/RFC4760, January 2007,
              <https://www.rfc-editor.org/info/rfc4760>.

   [RFC4604] Holbrook, H., Cain, B., and B. Haberman, "Using Internet
              Group Management Protocol Version 3 (IGMPv3) and Multicast
              Listener Discovery Protocol Version 2 (MLDv2) for
              Source-Specific Multicast", RFC 4604,
              DOI 10.17487/RFC4604, August 2006,
              <https://www.rfc-editor.org/info/rfc4604>.

   [RFC4609] Savola, P., Lehtonen, R., and D. Meyer, "Protocol
              Independent Multicast - Sparse Mode (PIM-SM) Multicast
              Routing Security Issues and Enhancements", RFC 4609,
              DOI 10.17487/RFC4609, October 2006,
              <https://www.rfc-editor.org/info/rfc4609>.

   [RFC7450] Bumgardner, G., "Automatic Multicast Tunneling", RFC 7450,
              DOI 10.17487/RFC7450, February 2015,
              <https://www.rfc-editor.org/info/rfc7450>.

   [RFC7761] Fenner, B., Handley, M., Holbrook, H., Kouvelas, I.,
              Parekh, R., Zhang, Z., and L. Zheng, "Protocol Independent
              Multicast - Sparse Mode (PIM-SM): Protocol Specification
              (Revised)", STD 83, RFC 7761, DOI 10.17487/RFC7761,
              March 2016, <https://www.rfc-editor.org/info/rfc7761>.

   [BCP38]   Ferguson, P. and D. Senie, "Network Ingress Filtering:
              Defeating Denial of Service Attacks which employ IP Source
              Address Spoofing", BCP 38, RFC 2827, May 2000,
              <https://www.rfc-editor.org/info/rfc2827>.

   [BCP41]   Floyd, S., "Congestion Control Principles", BCP 41,
              RFC 2914, September 2000,
              <https://www.rfc-editor.org/info/rfc2914>.

   [BCP145]   Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage
              Guidelines", BCP 145, RFC 8085, March 2017,
              <https://www.rfc-editor.org/info/rfc8085>.

9.2. Informative References

   [RFC4786] Abley, J. and K. Lindqvist, "Operation of Anycast
              Services", BCP 126, RFC 4786, DOI 10.17487/RFC4786,
              December 2006, <https://www.rfc-editor.org/info/rfc4786>.

   [RFC6347] Rescorla, E. and N. Modadugu, "Datagram Transport Layer
              Security Version 1.2", RFC 6347, DOI 10.17487/RFC6347,
              January 2012, <https://www.rfc-editor.org/info/rfc6347>.

   [INF_ATIS_10]
              "CDN Interconnection Use Cases and Requirements in a
              Multi-Party Federation Environment", ATIS Standard
              A-0200010, December 2012.

   [MDH-05]   Thaler, D. and B. Aboba, "Multicast Debugging Handbook",
              Work in Progress, draft-ietf-mboned-mdh-05, November 2000.

   [Traceroute]
              "traceroute.org", <http://traceroute.org/#source%20code>.

   [Mtrace-v2]
              Asaeda, H., Meyer, K., and W. Lee, Ed., "Mtrace Version 2:
              Traceroute Facility for IP Multicast", Work in Progress,
              draft-ietf-mboned-mtrace-v2-22, December 2017.

Authors' Addresses

   Percy S. Tarapore (editor)
   AT&T

   Phone: 1-732-420-4172
   Email: tarapore@att.com

   Robert Sayko
   AT&T

   Phone: 1-732-420-3292
   Email: rs1983@att.com

   Greg Shepherd
   Cisco

   Email: shep@cisco.com

   Toerless Eckert (editor)
   Huawei USA - Futurewei Technologies Inc.

   Email: tte+ietf@cs.fau.de, toerless.eckert@huawei.com

   Ram Krishnan
   SupportVectors

   Email: ramkri123@gmail.com