Application Layer Multicast Extensions to RELOAD

Mario Kolberg
University of Stirling
Stirling, United Kingdom
mko@cs.stir.ac.uk

John Buford
Avaya Labs Research
Basking Ridge, NJ, USA
buford@samrg.org

Abstract

Native multicast deployment is relatively slow and linked with a number of issues. However, there are a number of native multicast regions. Application Layer Multicast (ALM) can be used in areas of the network where there is no native multicast available. The SAM (Scalable Adaptive Multicast) Research group within the IRTF is investigating hybrid approaches to multicast, involving native deployments were available and ALM in other regions. SAM is using a P2P overlay to connect the nodes. Here we describe a protocol and API extensions to RELOAD for constructing Scalable Adaptive Multicast (SAM) sessions using hybrid combinations of ALM, native multicast, and multicast tunnels. The Automatic Multicast Tunneling (AMT) relay and gateway elements are employed for interoperation between native regions and ALM regions.

1. Introduction

Multicast is important for communications applications such as small group video conferencing and IPTV[1]. In hybrid multicast schemes, native multicast islands are interconnected using tunnels [2] or overlays with Application Layer Multicast (ALM) [3][4][5][6][7]. A hybrid protocol has been defined [8] which is based on integrating AMT (Automatic Multicast Tunnel) tunneling mechanism for native multicast (NM) [9] with application layer multicast protocol mechanisms.

The approach presented here centers around peers are connected to a DHT based P2P overlay. Tree forming and management operations as well as multicast data are routed over the overlay. Clearly, an overlay may support multiple concurrent multicast trees. Peers may be distributed throughout the network, that is in network regions which support native multicast (NM) as well as regions where it is not available. Consequently, some trees will only contain peers that use ALM, that is peers that do not have NM connectivity. Some other multicast trees will contain peers with a combination of ALM and NM. Although the overlay could also be used to form trees which consists of only native multicast peers, pure native multicast approaches are properly better suited if such peers are all in the same network region. AMT tunneling is expected to be used for such a tree if the peers are in different native multicast reagions. For the work in this paper we assume peers are able to determine, through configuration or discovery whether:

- they can connect to a NM router
- there is an AMT gateway accessible
- the peer supports AMT-GW functionality locally
- if MDR is supported in the region

The approach presented here uses RELOAD [10] as the distributed hash table (DHT) based overlay. RELOAD is a generic P2P overlay, and application support is defined by profiles called Usages. We define an Application Layer Multicast (ALM) Usage and a Hybrid ALM Usage.

We use AMT [9] to connect nodes in different native multicast regions. AMT permits AMT-R and AMT-GW functionality to be embedded in hosts or specially configured routers. We assume AMT-R and AMT-GW can be implemented in nodes.

This paper presents the following contributions: (1) extensions to the baseline RELOAD specification for ALM and hybrid ALM, and (2) a new P2P SIP uage for ALM and hybrid ALM.

2. Hybrid ALM Tree operations

Peers use the overlay to support multicast operations such as tree creation, joining, leaving, and re-forming a tree. Many algorithms for tree creation, management and usage, often with particular strengths for specific application areas, have been defined and our approach permits multiple such algorithms to be supported in the overlay. Therefore, multicast operation messaging must carry algorithm identification information.

For example, for small groups, the join point might be directly assigned by the rendezvous point, while for large trees the join request might be propagated down the tree with candidate parents forwarding their position directly to the new node.

In addition to these overlay level tree operations, some peers may implement additional operations to map tree operations to native multicast and/or AMT connections.

In the response to join requests, native multicast peers indicate details, such as their AMT-Relay or Gateway, their public IP address if they are in a private network. This information allows joining peers to establish features of existing tree peers. Native multicast peers use this information to establish if other native multicast peers

belong to the same or a different multicast region in the network, and ways to connect to it.

The joining peer can then decide to join the tree via ALM or AMT. Further, joining peers select the most suitable parent peer based on criteria such as lowest latency, throughput or highest capacity. The protocol implements a three way approach for this purpose: a peer which wishes to join a tree will send a join request. Multiple peers on the tree may offer to be a parent peer by responding with join accept messages. The joining peer will the select the most suitable parent peer and sends a join confirm message.
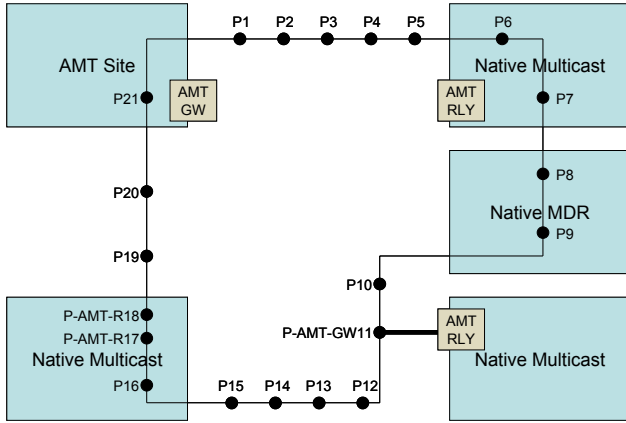

Figure 1: A sample SAM architecture.

In the above figure we show the hybrid architecture in six regions of the network. These are four native multicast regions (P6, P7) with one being based in multidestination routing (P8, P9), one AMT site which is linked to a native region via an AMT-GW (P21), and the rest of the network which uses ALM (e.g. P19, P20). All peers are connected in the overlay. AMT gateways (AMT-GW) do not have native multicast connectivity to the multicast backbone. AMT-Relays (AMT-R) do have native multicast connectivity to the multicast backbone. AMT Gateways provide a connection to the native multicast backbone via AMT Relays. Some native multicast peers themselves emulate the AMT relay behavior (P-AMT-R17, P-AMT-R18). There may also be peers which themselves behave as an AMT-GW (P-AMT-GW11) and which connect to multicast sources through an AMT-R. The following sections outline the algorithm for the particular scenarios.

2.1 ALM-Only Tree Algorithm

We have defined a simple base algorithm for forming a multicast tree in the overlay. Its main advantage is the use of the overlay routing mechanism for both control and data messages. The group creator does not have to be the root of the tree or even in the tree.

1. `groupID=create()` is used to create a new multicast group. This allocates a unique groupId to the tree. The root is the nearest peer in the overlay. The groupID will then be advertised out of band, most

likely by publishing it in the DHT where it can be discovered by other peers interested in joining the tree.

2. `joinTree(groupID)` is used to join a tree by sending this message to the root of the tree. Peers on the path to the root join the tree as forwarding points.

3. `leaveTree(groupID)` is used by peers who wish to leave the tree. The leave message is sent to each child node and to the parent node. If the parent node is a forwarding node and this is its only child, then it itself sends a leave message to its parent. A child node receiving a leave message from a parent sends a new join message to the groupID.

4. `multicastMsg(groupID, msg)` is used to forward a multicast message using one of the following two approaches: Firstly, in a Source Specific (SSM) tree, only the creator of the tree can be the source of a multicast message. It will send data messages to the tree root which then forwards them down the tree. Secondly, in an Any Source (ASM) tree, a node sending a data message sends the message to its parent and its children. Each node receiving a data message from one edge forwards it to its remaining tree edges.

2.2 ALM tree with joining peer at AMT site (AMT-GW)

In this case, the joining p eer is within an AMT region (such as P21) and attempts to connect to the tree using the ALM protocol. A number of peers which are already member of the tree may respond to this request. Based on these responses, there are the following cases. Please note the ALM approach can always be used as an alternative

1. Connecting to a native multicast peer in the same multicast region as the joining peer: The joining peer connects directly to the multicast peer.

2. Connecting to a native multicast peer in a different multicast region (which supports an AMT-R) as the joining peer (e.g. P6) or connecting to a peer which can function as a P-AMT-R (P-AMT-R17): the joining peer uses its AMT-GW to connect.

3. No native multicast peer responds: The joining native multicast peer uses ALM routing.

If the peer is not a joining peer and is on the overlay path of a join request and if its next hop is a peer in an NM region with AMT-R, it can select either overlay routed multicast messages or AMT delivered multicast messages. On the other hand, if its next hop is a peer outside of an NM region, then it can use ALM only or use AMT delivery as an alternate path.

2.3 ALM tree with NM joining peer using AMT-R

In this case, the joining peer is within a NM region (P6) and initially attempts to connect to the tree using the ALM protocol. Again, a number of peers which are members of the tree respond to the request. Based on these responses,

the following cases exist. As before the basic ALM approach is always possible as an alternative.

1. Connect to NM peer in the same region (P7). The joining peer may connect directly to the NM peer.
2. Connect to NM peer in a different region (P16). The joining peer may connect to the NM peer using AMT.
3. Connect to a peer which can function as P-AMT-GW (P-AMT-GW11): The joining peer may connect to the P-AMT-GW via its local AMT-R.
4. Connect to a peer which is in an AMT-GW region (P21): The joining peer may connect via the AMT-GW and its local AMT-R.
5. There is no NM peer in the tree. The NM peer uses basic ALM routing.

2.4 ALM tree with joining peer with P-AMT-R

Either the joining peer supports P-AMT-R (P-AMT-R18) or another peer in the multicast tree in the same region is P-AMT-R (P16 via P-AMT-R17) capable. The last three cases above apply here, replacing AMT-R with P-AMT-R.

2.5 ALM tree with joining peer with P-AMT-GW

If the tree includes a peer which is in a NM region which supports an AMT-R (P6, P7), or a peer featuring P-AMT-R (P-AMT-R17) the joining peer can connect to these using its AMT-GW functionality.

3. Protocol

The algorithms presented in the previous section are implemented as RELOAD [10] messages. RELOAD provides the mechanism to support a number of overlay topologies. This means that the hybrid overlay multicast framework can be used with P2P-SIP, but also, and perhaps more importantly, that the SAM framework is overlay agnostic.

As discussed earlier, there are a variety of ALM tree formation and maintenance algorithms. Our approach is meant to be algorithm agnostic, the same as RELOAD is overlay algorithm agnostic. We assume that all control messages are propagated using overlay routed messages. The message types needed for ALM behavior can be divided into the following three categories:

- Tree life-cycle (create, join, leave, re-form, heartbeat)
- AMT gateway advertisement and discovery
- Peer region and multicast properties

These are discussed in more detail in the following sections. In general, the parameter PeerId refers to the joining peer. ParentPeerId is the parent or candidate parent (during join) of a peer. GroupId indicates the id of the tree and Options contains particular options for a particular message.

3.1 Tree Lifecylce Messages

3.1.1 Create Tree

This message is used to create a new ALM tree in the overlay. The root node of the tree (according to common algorithms) is the peer with the id closest to and less than the `GroupId`. Just after creation, the tree has no children.

```
struct {
  NodeID PeerId;
  opaque SessionKey<0..2^32-1>;
  NodeID GroupId;
  Dictionary Options;
} CreateALMTree;
```

3.1.2 Join

This is the initial message a peer wishing to join a multicast tree (`PeerId`) sends. This causes the algorithm for peer join of a specific ALM group to be invoked. If successful, the PeerId is notified of one or more candidate parent peers with JoinAccept messages. These then get confirmed with JoinConfirm messages.

```
struct {
  NodeID PeerId;
  NodeID GroupId;
  Dictionary Options;
} Join;
```

3.1.3 Join Accept

This message is sent back to the joining peer indicating that the peer with id `ParentPeerId` is happy to act as its parent. A peer may receive more than one JoinAccept from different candidate parent peers. The joining peer accepts a peer as parent in the next step using the JoinConfirm message. A JoinAccept message which does not get confirmed or declined will expire.

```
struct {
  NodeID ParentPeerId;
  NodeID PeerId;
  NodeID GroupId;
  Dictionary Options;
} JoinAccept;
```

3.1.4 Join Confirm

A peer receiving a JoinAccept message from a peer with id `ParentPeerId` which it wishes to accept sends this message as confirmation. `Options` specified in this message are the options from the JoinAccept message the peer wishes to accept. This is the final step in this negotiation process.

```
struct {
  NodeID PeerId;
  NodeID ParentPeerId;
  NodeID GroupId;
  Dictionary Options;
} JoinConfirm;
```

3.1.5 Join Decline

A peer may receive multiple JoinAccept messages. It can decline any of these using this message. ParentPeerId identifies the peer which sent the JoinAccept message.

```
struct {
  NodeID PeerId;
  NodeID ParentPeerId;
  NodeID GroupId;
} JoinDecline;
```

3.1.6 Join via AMT Gateway

This message is a special form of the Join message indicating that the joining peer wishes to use AMT. If the parent peer is a P-AMT-R, then after the JoinAccept and JoinConfirm steps, an AMT tunnel is formed between the P-AMT-R to the `AMT-GW` which the Peer is associated with.

If the parent peer is a peer in a native multicast region, then after the JoinAccept and JoinConfirm steps, a tunnel is created between the native multicast peer's AMT-R and the specified AMT-GW. If the parent peer is an ALM peer, the request is propagated to other peers in the tree.

```
struct {
  NodeID PeerId;
  IpAddressPort AMT-GW;
  NodeID GroupId;
  Dictionary Options;
} JoinViaAMTGateway;
```

3.1.7 Join via Native Link

This allows a child to select a native multicast parent peer, overriding the selection based on the basic join method. The typical use is for a peer in a native multicast region to join the multicast group using the local native multicast path for the tree. This requires at least one other peer in the same multicast region to be a member of the tree already, or a peer in another multicast region being a member, and the two regions being linked via AMT. If neither is the case, the peer cannot join the tree using this method. The full message is shown below:

```
struct {
  NodeID ChildPeerId;
  NodeID ParentPeerId;
  NodeID GroupId;
  Dictionary Options;
} JoinWithNativeLink;
```

3.1.8 Leave

A peer which is part of an ALM tree identified by `GroupId` which intends to detach from either a child or parent peer sends a Leave message to the peer it wishes to detach from.

```
struct {
  NodeID PeerId;
  NodeID GroupId;
  Dictionary Options;
} Leave;
```

3.1.9 Leave via AMT Gateway

A peer which intends to detach from either a child or parent peer and which uses an AMT tunnel to connect to the peer sends a LeaveViaAMTGateway message to the peer it wishes to detach from. The request is transmitted to the AdjacentPeerId. This peer removes the specified PeerId from its children or parent lists.

If AdjacentPeerId is a P-AMT-GW, then it tears down the AMT tunnel from P-AMT-GW to the specified AMT-GW if no other children are using it. If AdjacentPeerId is a peer in a native multicast region, then the tunnel between its AMT-GW and the specified AMT-R is removed.

```
struct {
  NodeID PeerId;
  NodeID AdjacentPeerId;
  IpAddressPort AMT-GW;
  NodeID GroupId;
  Dictionary Options;
} LeaveViaAMTGateway;
```

3.1.10 Re-Form or Optimize Tree

This message triggers a reorganization of either the entire tree or a sub-tree. It may include information of specific peers of recommended parent or child peers to reconnect to. A peer receiving this message invokes local algorithms for selecting preferred parent and/or child peers.

```
struct {
  NodeID GroupId;
  NodeID PeerId;
  Dictionary Options;
} Reform;
```

If `PeerId` is specified then the tree is reorganized only at the sub-tree identified by PeerId. If this parameter is omitted the tree is reorganized starting from the root.

3.1.11 Heartbeat

A node signals to its adjacent nodes in the tree that it is alive. If a peer does not receive a Heartbeat message within N heartbeat time intervals, it treats this as an explicit Leave message from the unresponsive peer. N is configurable. In the message shown below `PeerId1` is the source of the hearbeat and `PeerId2` is its destination.

```
struct {
  NodeID PeerId1;
  NodeID PeerId2;
  NodeID GroupId;
} Heartbeat;
```

3.2 AMT gateway advertisement and discovery

Allows peer to disclose to other peers in the overlay their ability to act as a native-multicast gateway (as in AMT) for peers in a given region. We expect to use the P2P Publish and Lookup messages for this purpose. But to avoid collision with the semantics of those operations, we temporarily define shadow versions within the SAM extension. Publish stores an advertisement object for a peer with is an AMT gateway in the DHT for the overlay, under a given key.

```
struct {
  NodeID PeerId;
  ResourceID Key;
  opaque Region<0..2^32-1>;
  Dictionary Options;
} PublishAMTGateway;
```

```
struct {
  NodeID PeerId;
  ResourceID Key;
} LookupAMTGateway;
```

The parameters here are `PeerId`, which is the AMT gateway. `Key` is the key by which other peers can lookup the advertisement. There might be more than one key. `Region` is an id for a region, using a region identification scheme, such as the Autonomous System Number (ASN) [RFC1930].

3.3 Peer region and multicast properties

Peers can advertise the network region that they belong to and any of its native multicast properties. Similar to AMT Gateway advertisement and discovery, this uses the DHT for lookup and publishing.

```
struct {
  NodeID PeerId;
  ResourceID Key;
  opaque Region<0..2^32-1>;
  Dictionary Options;
} PublishPeerNMInfo;

struct {
  NodeID PeerId;
  ResourceID Key;
} LookupPeerNMInfo;
```

The parameters in these two operations are `PeerId`, which is the peer which is the AMT gateway. `Key` is the key by which other peers can lookup the advertisement. There might be more than one key.

4. RELOAD Usages

Applications of RELOAD are restricted in the data types that be can stored in the DHT. The profile of accepted data types for an application is referred to as a Usage. RELOAD is designed so that new applications can easily define new Usages. New RELOAD Usages are needed for hybrid multicast applications since the data types in base RELOAD and existing usages are not sufficient.

We define an ALM Usage and a Hybrid ALM Usage in RELOAD. The ALM Usage is sufficient for applications which only require ALM functionality in the overlay. The Hybrid ALM (HALM) Usage extends the ALM Usage so that hybrid native multicast and ALM trees can be used by applications.

The ALM Usage involves the following functions:

- ALM applications use the RELOAD data storage functionality to store a groupID when a new ALM tree is created in the overlay, and to retrieve groupIDs for existing ALM trees.
- ALM applications use the RELOAD data storage functionality to store a set of attributes for an ALM

tree, such as owner, tree size, tree height, tree formation algorithm, and join criteria.
- ALM applications and management tools use the RELOAD data storage functionality to store diagnostic information about the operation of tree, including average number of tree, delay from source to leaf nodes, bandwidth use, lost packet rate. In addition, diagnostic information may include statistics specific to the tree root, or to any node in the tree.

The Hybrid ALM Usage involves the following additional functions:

- HALM applications use the RELOAD data storage functionality to store a set of attributes for an AMT Gateway that can connect to at least one node in the overlay.
- HALM applications use the RELOAD data storage functionality to store a set of attributes about a native multicast region associated with an AMT Gateway.
- HALM applications and management tools use the RELOAD data storage functionality to store diagnostic information about the operation of AMT and ALM interconnections.

5. Conclusions and Further Work

This paper presents a hybrid application layer multicast approach, which uses native multicast where available. AMT tunneling is used between multicast islands. Nodes are connected using a DHT based overlay. Specifically, this paper presents algorithms for the different scenarios and corresponding messages for the RELOAD overlay. We are currently implementing a testbed for experimentation.

6. References

[1] J. Buford, M. Kolberg. Hybrid Overlay Multicast Simulation and Evaluation. IEEE CCNC 2009. Jan. 2009.

[2] B. Zhang, S. Jamin, and L. Zhang. Universal IP multicast delivery. Int'l Workshop on Networked Group Communication, 2002

[3] X. Jin, H.-S. Tang, S.-H. Chan and K.-L. Cheng, Deployment Issues in Scalable Island Multicast for Peer-to-Peer Streaming," IEEE Multimedia Magazine, vol. 16, issue 1, pp. 72-80, Jan.-Mar. 2009.

[4] X. Jin, K.-L. Cheng, and S.-H. Chan, Scalable Island Multicast for Peer-to-Peer Streaming, Hindawi Journal of Advances in Multimedia special issue on Multimedia Networking, vol. 2007, Article: 78913.

[5] X. Jin, K.-L. Cheng, and S.-H. Chan, SIM: Scalable Island Multicast for Peer-to-Peer Media Streaming, in Proceedings IEEE International Conference on Multimedia Expo (ICME), pp. 913-916, 2006.

[6] M. Waehlisch, T. Schmidt. Multicast routing in structured overlays and hybrid network. in: Handbook of Peer-to-Peer Networking. (eds. S. Shen, H. Yu, J. Buford, M. Akon). Springer-Verlag. Forthcoming.

[7] M. Wählisch, T. C. Schmidt, G. Wittenburg. A Generalized Group Communication Network Stack and its Application to Hybrid Multicast, 28th IEEE INFOCOM. Student Workshop, 2009.

[8] J. Buford, M. Kolberg, T. Schmidt, M. Waehlisch, Application Layer Multicast Extensions to RELOAD, Internet Draft draft-kolberg-sam-baseline-protocol-01, Work in progress, July 2010.

[9] D. Thaler, M. Talwar, A. Aggarwal, L. Vicisano, T. Pusateri. Automatic IP Multicast Without Explicit Tunnels (AMT). IETF draft-ietf-mboned-auto-multicast-10, Work in progress. March 2010.

[10] Jennings, C., Lowekamp, B., Rescorla, E., Baset, S., and H. Schulzrinne, "REsource LOcation And Discovery (RELOAD) Base Protocol", draft-ietf-p2psip-base-10 (work in progress), August 2010.