22nd AIAA International Communications Satellite Systems Conference & Exhibit 2004
9 - 12 May 2004, Monterey, California

AIAA 2004-3176

Framework for IP Multicast in Satellite ATM Networks

Ayan Roy-Chowdhury* and John S. Baras†
Department of Electrical and Computer Engineering and Center for Satellite and Hybrid Communication Networks, Institute for Systems Research, University of Maryland College Park, MD 20742

This paper proposes a design for IP multicast routing in hybrid satellite networks. The emergence of IP multicast for Internet group communication has placed focus on communication satellites as an efficient way to extend the multicast services for groups with distributed membership in wide-area networks. This poses interesting challenges for routing. Hybrid satellite networks can have both wired and wireless links and also combine different link-layer technologies like Ethernet and ATM. No proposed IP multicast routing protocol for wired networks offers an integrated solution for such networks. This paper attempts to provide a solution by proposing a design for IP multicast routing in wide-area networks that have terrestrial Ethernet LANs interconnected by ATM-based satellite channels. The paper reviews the multicast services offered by IP and ATM, and proposes a multicast routing framework that combines PIM-SM protocol for terrestrial multicasting with the ATM MARS and VC mesh architecture for multicast routing over the satellite links. Modifications are made to the standard protocols to suit the unique needs of the network being considered. The feasibility of the proposed design is tested by performing simulations. The proposed framework is presented in detail, along with analysis and simulation results.

Nomenclature

IP	=	Internet Protocol
ATM	=	Asynchronous Transfer Mode
PIM-SM	=	Protocol Independent Multicast – Sparse Mode
MARS	=	Multicast Address Resolution Server
LAN	=	Local Area Network
VC	=	Virtual Channel

I. Introduction

IP multicast routing is a network layer mechanism that provides resource-efficient communication services for applications that send the same data to multiple recipients simultaneously. The source transmits a single copy of the data; an intermediate router makes a copy of each incoming multicast packet to retransmit on each outgoing link towards the destinations reachable from it. This makes efficient use of network bandwidth compared to sending multiple unicasts, where the source sends a copy of the packet separately to each receiver. Applications that can benefit from use of multicast include webcasts, shared workspace, video- and voice-conferencing, and online gaming.

Satellite networks offer a natural method to extend the multicast services in wide-area networks where the sources and recipients are widely separated from one another. Satellites offer high bandwidth for broadband services, as many multicast applications are. Their broadcast nature allows the sources to reach multiple recipients simultaneously. For *geostationary* orbit satellites, the transmission from the source to recipients can be accomplished in a single hop. Satellite networks are self- contained and require less infrastructure compared to terrestrial fiber-based networks, and hence can be set up rapidly.

There is, however, little support today for IP multicast services over satellites. Most IP multicast routing protocols have been proposed for networks that are either fully wired or wireless; they set up either a multicast tree

* Ph.D. student, Electrical and Computer Engineering, 1103 AV Williams College Park MD 20742, ayan@umd.edu.
† Professor, Electrical and Computer Engineering and Institute for Systems Research, 2147 AV Williams College Park MD 20742, baras@isr.umd.edu.

or a mesh through point-to-point communications between routers that are in proximity to one another. The protocols do not consider hybrid broadcast networks, such as those involving satellites, that can have both wired and wireless links and that can connect distant nodes in a single hop, or reach multiple nodes simultaneously through broadcast transmission. The IP multicast protocols also assume that Ethernet is used as the underlying access layer. Since Ethernet has native support for multicasting, mapping IP multicast to Ethernet multicast is relatively simple. But the multicast mapping becomes complicated when other link layer technologies are considered. For example, ATM has no native support for multicast and requires fairly complex mechanisms to support IP multicast over ATM links. Although there are several solutions for supporting IP multicast over different access layers, very little has been done in designing support for IP multicast in networks that combine multiple link layer technologies. There is an important need to address these issues in satellite networks, since there have been proposals for geostationary satellite networks that would interconnect geographically distributed high-speed terrestrial networks via ATM-based satellite links[15]. The satellite will have multiple spot beams for selective broadcast to different locations, and incorporate an ATM stack on-board for fast switching. Such a network will be very attractive for broadband multicast routing, but that will require finding efficient solutions to the above problems.

This paper addresses the above problems in a satellite network that has Ethernet-based terrestrial LANs of varying capacities inter-connected via ATM-based geostationary satellite links. The network can support multicast groups with dynamic membership and varying in size from several hundred to several million, with the sources and recipients widely distributed. The paper proposes a design for routing that integrates traditional IP multicast in the Ethernet LANs, with ATM support for IP multicast over the satellite links, for end-to-end multicast routing in the satellite network. The proposed design makes use of well-known multicast protocols with modifications to suit the unique needs of the network. The primary concern in the design is to optimize the flow of multicast control and data traffic over the satellite links, avoid redundant re-transmissions and support heterogeneous link characteristics in a single multicast group. To demonstrate the feasibility of the routing framework, design simulations are performed for different scenarios, which show that the proposed architecture has comparably good performance characteristics, and has low control overhead.

The rest of the paper is organized as follows. Section II covers the fundamental concepts of IP multicast and reviews some popular IP multicast protocols. Review of ATM multicasting is in section III. Section IV describes the network architecture and details the design of the proposed multicast routing framework. Simulation of the routing framework and the results of the simulation are given in section V. We present our conclusions in section VI.

II. IP Multicast Concepts and Routing Protocols

A. IP Multicast Fundamentals

The original IP multicast model, proposed in Ref. 12, is based on the notion of a *group*, identified by a unique *address*, and composed of a certain number of participants (senders and receivers). Here we review the basic concepts in IP multicast, based on the treatment in Ref. 13.

1) IP Address Space: The IP address associated with a multicast group is assigned from the class D address space. Some of these addresses are pre-assigned, while the others can be dynamically allocated at the time of group formation.
2) Member Registration: The IP multicast protocols make use of the Internet Group Management Protocol[14] (IGMP) to find out about the participants in a group. All receivers in a multicast group are required to explicitly register the multicast address for which they wish to receive data, by sending join requests to their local IGMP-enabled multicast routers. When a receiver wants to leave a group, it sends an explicit leave request. The receivers can join and leave at any time during a multicast session. IP multicast hence "maps" a multicast address to a set of receivers. Registration is required only for receivers, but not for the senders to a group. The recipients can be anonymous; the sources need not know who the receivers are, also the receivers do not know each other.
3) Multicast Tree: The multicast routers and the receivers together form the *multicast delivery tree*. The receivers are always at the leaves of the tree. The tree might have one or more root(s) or core(s), depending on the routing algorithm. The core(s), if present, is a (are) multicast router(s). The multicast tree can be either a *shared tree*, i.e., a single common tree for a multicast group; or, *source-specific shortest path trees*, where every source for a multicast group has its own individual tree rooted at the source.
4) Unidirectional or Bidirectional Forwarding: The multicast traffic in a group can be *unidirectional* or *bidirectional*. In unidirectional forwarding, the source(s) send the data packets to the core node; the data is then forwarded along the shared multicast tree to reach the set of receivers. In bidirectional forwarding, the

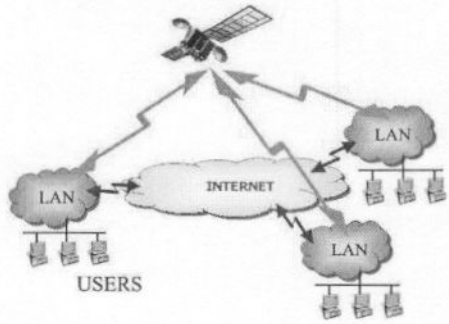
multicast traffic from the source does not necessarily have to go through the core router(s) to reach the recipients in the tree.

In summary, support for IP multicast in wired networks requires the following mechanisms:

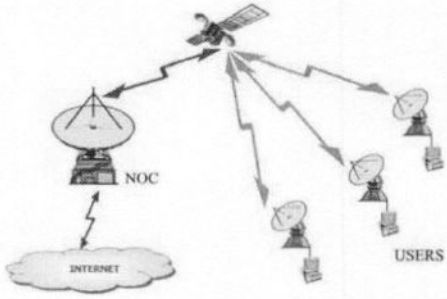
- Allocation of a class D address.
- Registration of the set of receivers.
- Setting up the multicast tree and dynamic membership management.
- Routing of traffic from the sources to the receivers along the multicast tree.

B. Wide-Area Multicast Routing via Satellites

Satellite networks have some inherent advantages in providing multicast service, as mentioned in section I.



a) Satellite Backbone Deployment



b) Satellite Direct-to-Home Deployment

Figure 1: Satellite Network Topologies[10]

There are two common topologies for support of multicast service in a satellite network[10]. A satellite can be deployed as a *backbone* for connecting LANs that are widely separated from one another. Each LAN has multiple terrestrial nodes and one or more satellite gateways that can uplink to and downlink from the satellite (Fig. 1(a)). The nodes in the LAN receive transmission from, and send to, the satellite via the gateway nodes. This topology is thus hierarchical in structure.

The other topology is the *direct-to-home* (DTH), in which there are multiple independent terrestrial nodes, each with its own connectivity to the satellite. The connections can be unidirectional or bidirectional. The network has a star topology and user terminals have no access to other networks. The ground terminals access the terrestrial core network through a gateway node located at the Network Operations Center (NOC) (Fig. 1(b)).

Most deployed satellites do not perform on-board switching or processing; instead, they broadcast the data packets on all outgoing links. Future satellites are planned to be more sophisticated, supporting multiple spot-beams covering different geographical regions over a large area. These satellites will be able to perform on-board switching and processing, and transmit the data packets only on the outgoing links, based on the spot-beams, that are necessary[1].

A geostationary satellite can connect large, widely-separated, terrestrial networks. The satellite will thus be a part of the multicast tree. If the networks in a multicast group are in different spot-beams, then the satellite will have to perform on-board switching for the multicast traffic. The challenge therefore is to design efficient routing protocols that would allow the satellite to do "selective" broadcast and send out the traffic only on the links that have receivers downstream. The relative simplicity of the satellite network can offer a simpler design for end-to-end multicast.

Most deployed satellites use their own link layer protocols. The amount of processing at the satellite is minimal. Since it is difficult to have a generic design based on proprietary protocols, one can look for standards that are closely matching. ATM is attractive since it supports very fast switching. There have been proposals for satellites with ATM switching support. It is a challenging task to design a multicast routing framework that integrates terrestrial Ethernet networks with ATM satellite channels, as discussed in section I.

D. Multicast Routing Protocols

There have been several proposals for multicast routing protocols in the literature. The various protocols can be classified as either intra-domain, i.e., managing the multicast tree within a domain, or inter-domain, i.e., for building the multicast tree across domains. We briefly outline some of the most popular ones, based on the treatment in Refs. 5, 13.

1. Intra-domain Multicast Routing Protocols

One of the earliest proposals for intra-domain multicast routing is the Multicast Open Shortest Path First (MOSPF) protocol[16]. MOSPF is the multicast extension to Open Shortest Path First (OSPF) unicast routing protocol. OSPF is extended to support multicast by the addition of group membership *link state advertisements* (LSAs).

MOSPF requires heavy computation at each on-tree router for computing the shortest path tree (SPT) per source. MOSPF is slow to react to frequent membership changes, and incurs a heavy control message overhead. Also, MOSPF needs to maintain routing state entry for every {source, multicast group}, even if the source transmits infrequently. The protocol hence scales poorly to large groups.

Distance Vector Multicast Routing Protocol[17] (DVMRP) computes the multicast routing paths based on the unicast routing tables constructed by the unicast Routing Information Protocol[18] (RIP). DVMRP uses "flood and prune" or Reverse Path Forwarding[2] (RPF) algorithm to construct the multicast tree. The flooding mechanism can incur a heavy overhead in large networks with many sources. Also, DVMRP is a *soft-state* protocol requiring periodic refresh of the multicast prune state in each router, therefore the multicast packets need to be flooded periodically. DVMRP can also have heavy overhead in terms of storage, since each on-tree router needs to maintain state for every source per group.

Core-Based Tree[8] (CBT) multicast routing protocol uses a shared bidirectional tree for a group, rooted at a core router. The single shared tree requires less state information to be maintained at each multicast router per group. However, using a single shared tree leads to "traffic concentration" on a few links that are part of the shared tree. This can be avoided if source-based trees are used. Also, the sender and the receivers are not necessarily connected by the shortest path when using the shared tree. Therefore the delivery delay can be higher compared to using source-based shortest path trees.

Protocol Independent Multicast[3] (PIM) has been proposed for multicast routing with the flexibility to support both source-based shortest path trees and core-based shared trees, while attempting to minimize the overheads associated with either approach. PIM comes in two flavors - PIM Dense Mode (PIM-DM) and PIM Sparse Mode (PIM-SM). PIM-DM has been designed for networks that are densely populated with members of a multicast group. PIM-DM builds the multicast tree using "flood-and-prune" RPF, as in DVMRP. The primary difference between DVMRP and PIM-DM is that PIM-DM is independent of the unicast routing protocol; it only requires that a unicast routing protocol exists to construct the unicast routing tables; PIM-DM uses the unicast routing tables to build the multicast tree.

PIM-SM has been designed as a multicast routing protocol for a sparsely populated network. We describe PIM-SM in some detail here, since an understanding of the protocol is necessary for the multicast framework that we propose.

The definition of a region as *sparse* requires any of the following conditions to be true[5]:

- The number of networks/domains with members is smaller than the total number of networks/domains.
- Group members are widely distributed.
- The overhead of flooding all the networks with data followed by pruning networks with no members in them is significantly high. In addition, the groups are not necessarily small and hence dynamic alteration of the groups with a large number of members must be supported.

PIM-SM supports both shared tree and shortest path trees. PIM-SM uses the concept of a central node for a multicast group, similar to CBT. The central node in PIM-SM is called the *Rendezvous Point* (RP). A unique RP for each group is determined based on the multicast group address, by a special router called the *Bootstrap Router* (BSR). In PIM-SM, the routers responsible for managing group membership in the leaf subnets are called the *Designated Routers* (DRs). When any receiver wants to join the multicast group, its DR sends an explicit "join" request to the RP. The join message is processed by all the routers between the receiver and the RP; the routers save the state information for the group. When a sender wants to multicast to a group, its DR initially encapsulates the data packets and unicasts them to the RP, which then forwards the de-capsulated data packets to the receivers along the RP-rooted shared multicast tree (RPT). A shortest path tree rooted at the sender is created if the sender's traffic increases beyond a pre-determined threshold. All the routers on the shared tree between the RP and the receivers send a "join" message towards the source and a "prune" message towards the RP, thereby creating the source-rooted SPT. The RP itself joins the SPT. Once the source-rooted tree is created, the source forwards the data packets along the SPT, and not the RPT. The RP continues to receive a copy of the multicast data packet (in native format), and forwards the packet along the shared RP tree. This is done because there might still be receivers who are receiving from the shared tree. It also ensures that new receivers who join the group are able to receive data packets for the group till the time they switch to the SPT. The unicast routing information is derived from the unicast routing tables, independently of the unicast routing protocol that constructed them. PIM-SM uses "semi-soft" states - the state information in each on-tree router has to be periodically refreshed (by sending join/prune message for each active entry in the PIM routing table). The periodic messages can reflect changes in topology, state or membership information. If the periodic update message is not received from a downstream router within the pre-set timeout period, the state entry is deleted from the upstream router's local memory. Since the state information is periodically refreshed, PIM-SM does not need an explicit *tear down* mechanism to remove state when a group ceases to exist.

PIM-SM creates large routing tables and requires significant memory at the routers to store the multicast state. The complexity of processing at the routers is also high. However, the protocol has many attractive features such as fast join to the multicast tree, low latency for high data rate sources, robustness to loops and node failures, independence of the unicast protocol, scalability, and inter-operability with other multicast protocols, which have led to its wide acceptance.

2. Inter-domain Multicast Routing Protocols

Several IP-level protocols have been proposed for managing a multicast group across different domains. Hierarchical DVMRP[9] (HDVMRP) organizes a network into a two-level hierarchy – the top-level consisting of non-overlapping regions and the lower level consisting of subnets within regions. DVMRP is proposed as the inter-region multicast protocol. Any multicast protocol can be used for multicast within a region. The regions are interconnected through border routers that exchange information about the regions in the top-level only. HDVMRP floods data packets to the border routers of all regions, and border routers that are not part of the group send prunes toward the source network to stop receiving packets. This implies a large overhead and maintenance of state per source, even when there is no interest for the group. HDVMRP also requires encapsulating the data packets for transit between the regions, which adds additional overhead.

Hierarchical PIM[20] (HPIM) was designed to overcome the drawback in PIM that the placement of the RP can be sub-optimal for a sparsely distributed group in a large network. HPIM uses a hierarchy of RPs for a group. Each candidate RP belongs to a certain level. An RP at a higher level has a wider coverage area. A receiver would send join messages to the lowest level RP (which is its local DR), which in turn would join an RP at the next higher level and so on, till the top-level RP is reached. The hierarchy of RPs helps in detecting loops and in decoupling control flow from the data flow. However, it is difficult to come up with a hierarchical placement of RPs without extensive knowledge of the network topology and the receiver set. Also, the tree in HPIM does not perform well in terms of delays from the source to receivers, especially in the case of local groups.

The combination of PIM-DM and PIM-SM was an early proposal for inter-domain multicast routing - PIM-DM to be used for intra-domain routing, while PIM-SM will connect the domains. Thus, PIM-DM will maintain source-rooted trees at every domain, that will be connected by a shared tree (and source-rooted trees) constructed by PIM-SM. The approach cannot be applied to a large heterogeneous network since the mechanism to advertise RPs and the maintenance of soft state entries in PIM-SM will have heavy control overhead. The amount of state entries required to be maintained is also not feasible for an inter-domain protocol (one state entry for the shared tree, and then as many as the number of source-specific trees available).

Border Gateway Multicast Protocol[21] (BGMP) is designed to inter-operate with any multicast routing protocol employed intra-domain, e.g., PIM-SM, CBT, DVMRP, etc. BGMP associates each multicast group with a root or core and constructs a shared tree of domains. The root is an entire domain in BGMP, and not a single router. Specific ranges of the class D address space are associated with various domains. Each of these domains is selected as the shared tree root for all groups whose addresses are in its range. The architecture of BGMP consists of domains or autonomous systems, and border routers with two components: (1) BGMP component and (2) Multicast Interior Gateway Protocol (M-IGP) component. The M-IGP component can be any intra-domain multicast routing protocol.

BGMP runs on the border routers and constructs a bi-directional shared tree that connects individual multicast trees built in a domain. In order to ensure reliable control message transfer, BGMP runs over TCP. As stated in Ref. 13, due to bidirectional forwarding, BGMP is not adequate for asymmetrical routing environments. Moreover, BGMP can only support source-specific delivery criteria in limited cases, for keeping the protocol simple. To obtain a globally available multicast routing solution, the use of BGMP necessitates that inter-operability problems, specific to the M-IGP being used, be solved.

E. ATM Support for Multicast

ATM networks based on UNI 3.0/3.1[6,7] do not provide the native multicast support expected by IP; ATM specifications do not have the concept of abstract group address for multicasting. Therefore if a sender wants to multicast data to a group of recipients, it has to know apriori the individual ATM addresses of the set of recipients, and it needs to set up multicast connections rooted at itself, to the set of receivers before it can send the data packets. This is in contrast to IP, where the multicast model is receiver-initiated.

1. ATM Point-to-Multipoint VC

One-to-many traffic flow in ATM is done using a unidirectional *point-to-multipoint virtual connection* (p2mpVC) (Fig. 2), which is specified in UNI 3.0/3.1. The point-to-multipoint VC is initiated from the sender ATM endpoint by opening a point-to-point virtual connection to the first receiver ATM endpoint by explicit ATM signaling mechanism. The sender subsequently adds "branches" to the point-to-point VC, specifying the other receiver ATM addresses; the signaling ensures that branches are created in the intermediate ATM switches on the path from the

sender to the set of receivers as appropriate. The sender is also responsible for connection tear down when it ceases data transmission. The source transmits a single copy of each cell; cell replication happens at the ATM switches where branching occurs.

In UNI 3.0/3.1, an ATM node who wants to receive cannot add itself to the point-to-multipoint VC. If the set of recipients changes during the lifetime of the connection, the source must explicitly add or remove any new or old recipients, by specifying the leaf node's actual unicast ATM address.

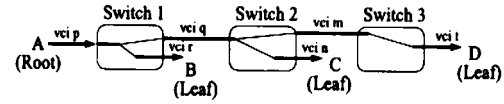


Figure 2: Point-to-multipoint VC

2. ATM Multipoint-to-Multipoint Communication Model

Emulating multipoint-to-multipoint service in ATM networks based on UNI 3.0/3.1 can be done using either a *VC mesh*, or, a *multicast server* (MCS). The VC mesh is the simpler approach: each ATM sender creates its own unidirectional point-to-multipoint VC to the set of receivers as the leaf endpoints. Nodes that are both sources and receivers for a group will originate a single point-to-multipoint VC (as a sender) and then terminate a branch of one other VC for every other sender of the group. This results in a crisscrossing of VCs across the ATM network, hence the term *multicast mesh* or *VC mesh*. Fig. 3 shows a VC mesh with four ATM nodes, each of which acts both as source and receiver.

The primary advantages of the VC mesh approach are optimal data path performance, low latency, and differential service. The major disadvantages are high usage of resources and heavy signaling load.

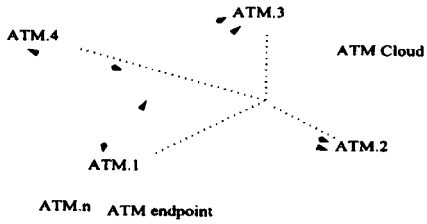


Figure 3: VC Mesh Architecture

The multicast server (MCS) architecture attempts to overcome the drawbacks of the VC mesh approach by using servers to forward multipoint-to-multipoint traffic. The MCS attaches to the ATM network and acts as a proxy group member. It terminates point-to-point VCs from all the endpoints, either sources or receivers, and originates one point-to-multipoint VC which is sent out to the set of all group members. The basic function of the MCS is to reassemble ATM Adaptation Layer Service Data Units (AAL SDUs) from all the sources and retransmit them as an interleaved stream of AAL SDUs out to the recipients.

The main advantages of the MCS architecture are low resource consumption and low signaling overhead. The main drawbacks include traffic concentration on the links leading to the MCS, high latency in data delivery, single point of failure and *reflected packets* to the source.

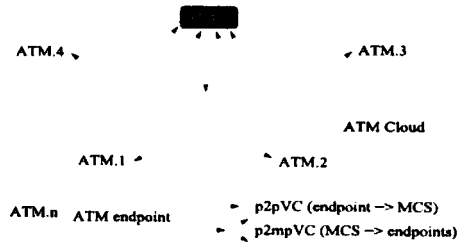


Figure 4: MCS Architecture

3. IP Multicast Support in ATM: MARS Architecture

In order to make IP multicast work over ATM, the use of *Multicast Address Resolution Server* (MARS) has been proposed in Ref. 4. MARS (Fig. 5) is used to map IP multicast addresses to the ATM addresses of the endpoints belonging to the IP multicast group. The MARS keeps a table of {Class D address, ATM address$_1$, ATM address$_2$,..., ATM address$_n$} mappings for every layer 3 multicast group that has one or more members.

The set of IP/ATM endpoints managed by a single MARS is known as a *cluster*. In the traditional model, the IP hosts are grouped into clusters and each such cluster has a MARS. The clusters are interconnected using IP multicast routers. Thus *inter-subnet* multicasting is still done using IP multicast routing protocols, while the *intra-subnet* multicasting is done using ATM with the help provided by MARS[5].

Two types of VCs are used to carry control messages between a MARS and its MARS clients:

1) A transient point-to-point VC to the MARS carries query/response activity initiated by the MARS client. There is one such VC for every MARS client connected to the MARS.

2) For control messages propagated by the MARS, the MARS uses a semi-permanent point-to-multipoint VC that has all its MARS clients as leaf nodes. This VC is known as the *ClusterControlVC (CCVC)*. Before a MARS client may use a given

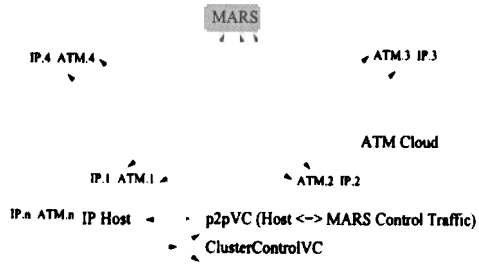


Figure 5: MARS Architecture

MARS, it must register with the MARS, allowing the MARS to add it as a new leaf of the CCVC.

An ATM endpoint who wants to send to an IP multicast group, queries the MARS for the list of ATM addresses of the multicast group members. On receiving the list from the MARS in a reply message, the endpoint proceeds to send the multicast traffic to the endpoints. The actual transfer of the multicast traffic can be done using either the VC mesh or the MCS architecture

III. Framework for IP Multicast Routing in Satellite ATM Networks

A. Satellite Network Architecture

The network architecture we consider is shown in Fig. 7. The architecture has a group of networks geographically separated and spread over a wide area. They constitute the "subnetworks" in the overall network. The subnetworks are connected to each other by satellite links using a geostationary satellite. The subnetworks are Ethernet-based, while the satellite links are ATM-based. Each subnetwork connects to the satellite using one or more satellite gateways or satellite terminals.

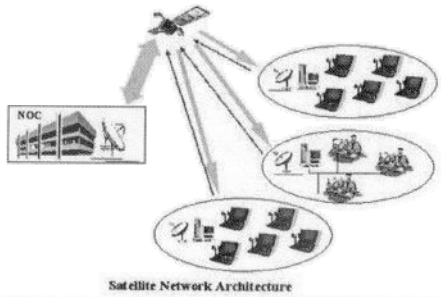


Figure 7: Satellite Network Architecture

The satellite is an ATM switch with spot-beam technology and the ability to do switching between beams, but has no support for IP. There is a network operations center (NOC) from which the operation of the satellite is controlled, through a dedicated connection. The geostationary satellite links involve high delay, of the order of 250ms in a single-hop (for example, Spaceway[1]). The uplink bandwidth is also constrained to approximately 1.54 Mbps. These are important considerations when we design the multicast routing framework.

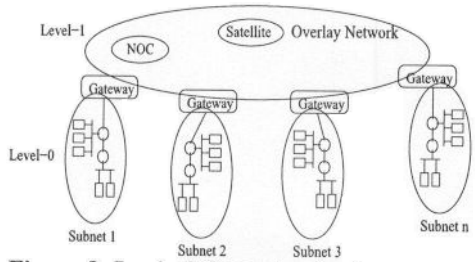


Figure 8: Logical Grouping in the Satellite Network Architecture

B. IP/ATM Multicast Routing Framework

The network architecture described in section III.A forms a natural hierarchy, as shown in Fig. 8. The network can be considered to be composed of terrestrial domains or subnetworks at the lower level. The satellite gateways connected by the satellite links form an overlay that interconnects the terrestrial subnetworks. Transmission between the subnetworks is done through the satellite by switching between the spot-beams to connect the source and destination domains. Therefore, the design of a framework for IP multicasting routing for this network involves two components:

- "Traditional" IP multicast routing in each Ethernet-based subnetwork. This is similar to the intra-domain IP multicast routing. Therefore it involves the selection of a suitable IP multicast routing protocol.
- IP multicast over ATM for inter-domain multicast routing. This requires the design of a suitable mechanism to multicast IP over the ATM-based satellite links.

1. Selection of Intra-domain Multicast Routing Protocol

The selection of a suitable IP multicast protocol for efficient and scalable intra-domain multicast routing within each subnetwork depends on the multicast group size and the dynamics of member joins and leaves. The terrestrial networks that we consider can be large with the multicast group members widely dispersed in each subnetwork. At the same time, the total number of group members in each subnetwork can be high, though a fraction of the total hosts in the subnet. We can therefore term the group as "sparse". PIM-SM has been proposed as a candidate protocol for multicast routing in sparse networks. Although PIM-SM is a complex multicast routing protocol, it has several features that make it attractive:

- It can efficiently manage a multicast group with low control message overhead.
- It allows fast receiver joins to a multicast group due to the presence of the shared tree.
- Initial source transmission is rapid and has low overhead due to the register mechanism.
- PIM-SM ensures low end-to-end latency for sources that require it by using source-specific trees.
- It can scale well if the number of group members increases.
- Use of centralized RP in PIM-SM facilitates the design of security framework for data confidentiality[11]

We therefore select PIM-SM as the protocol for intra-domain multicast routing.

2. *Selection of Inter-domain Multicast Routing Protocol*

The inter-domain multicast in our network architecture involves sending IP packets over ATM connections. Our inter-domain architecture is a "one-hop" ATM network, with one switch (the satellite) that can reach all the nodes (the satellite gateways) simultaneously.

None of the inter-domain protocols discussed in section II takes into consideration the unique characteristics of the satellite medium. We wish to minimize the amount of control and data traffic that flow over the satellite links, due to their high latency and constrained uplink bandwidth. BGMP, which is a popular inter-domain protocol, would create point-to-point TCP connections between the satellite gateways (BGMP peers). The root domain for every class D group will need to be one of the subnetworks; this therefore will mean unnecessary retransmissions - one to the root domain, and then from the root domain to all other domains, via the same overlay network. Also, since there will be point-to-point TCP connections between BGMP peers, the traffic will need to be replicated multiple times from the source border router to the receivers, which is a wasteful use of the satellite broadcast capabilities. The other inter-domain protocols also suffer from similar drawbacks when applied *as is* to our overlay network.

However, the VC mesh and MCS architectures can be well applied to the overlay network. The MCS architecture is ideally suited - the satellite can be the MCS, with each source sending only one copy of each cell on the uplink, which the satellite replicates and broadcasts using a point-to-multipoint VC to the receivers. But the MCS architecture suffers from several drawbacks when applied to the network:

a) The network has only one physical node (the satellite) that can act as the MCS. A single MCS can serve only one IP multicast group at a time, as it has no way to differentiate between traffic destined for different groups, since when IP multicast packets are fragmented into cells, the group information is lost till the cells are reassembled at the receivers. The single MCS can be extended to serve multiple groups by creating multiple logical instances of the MCS, each with different ATM addresses (e.g. a different SEL field value in the node's Network Service Access Point Address (NSAPA)[22]). But the SEL field is only 8 bits; therefore there can be at most 256 groups. This is a limitation for scalability that should be avoided.

b) To support even one group that can have multiple sources, the MCS needs to be able to do segmentation and re-assembly for every cell it receives, since AAL5 does not support cell level multiplexing of different AAL SDUs on a single outgoing VC. This involves higher latency. Also, we assume that the satellite has very limited switching functionality, and cannot do any extended processing.

c) A slightly more complex approach to support multiple groups using a single MCS would be to add minimal network layer processing into the MCS. This would require that every cell is re-assembled into the original IP multicast packet, the MCS checks the group address in each packet, and then the packet is again segmented into cells and sent out on the appropriate point-to-multipoint VC for the group. This will result in significantly higher latency due to the processing required, and necessitate sizeable buffers at the satellite, especially when the sources have high data rate. Also, the processing at the MCS will be complex and will require it to support an IP stack. No satellite to date has support for IP processing in it, and we make no assumption to that effect.

Based on the above reasons, we do not design our framework using the MCS architecture for routing in the overlay. Instead, we select the VC mesh architecture. Although the VC mesh has higher resource consumption in comparison to the MCS, the expected throughput is higher and end-to-end latency is lower (since the mesh does not need the intermediate AAL SDU reassembly that must occur in MCS), and makes no additional demand on the capabilities of the satellite, except that it be an ATM switch that supports UNI 3.0/3.1 signaling.

We describe in detail our framework in the next section. The framework is based on the technical description of PIM-SM and its message formats provided in Ref. 19, and on the description of ATM support for IP multicast and the signaling mechanism and message formats that are detailed in Ref. 22.

3. *Description of the Multicast Routing Framework*

IP Multicast Framework in each Subnet:

- Each subnetwork is a PIM-SM domain and runs standard PIM-SM multicast protocol in the routers.
- Routers directly connected to the end hosts also run standard IGMP.
- One or more satellite terminals in a subnetwork are configured to act as Rendezvous Points (RPs) for all the multicast groups in the subnetwork. We term the subnet RPs the "local" RPs. The local RPs create the shared multicast tree for the multicast groups in their subnet.
- A router in each subnetwork is configured to act as the *bootstrap router* (BSR) for the subnetwork, for selecting the active local RP, from amongst the list of candidate RPs in the subnetwork, in situations where a selection is needed (when there are multiple gateways in a subnetwork configured to act as RPs for PIM-SM multicast groups). Every subnetwork has its own BSR.

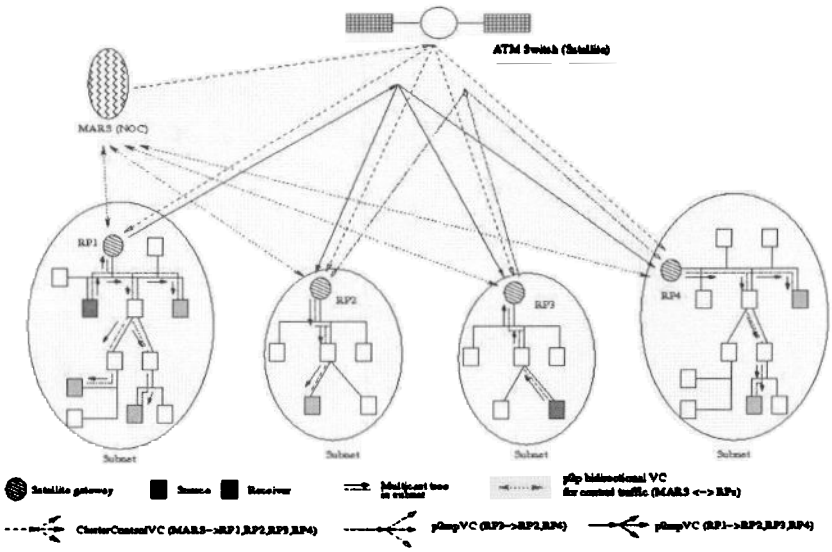


Figure 9: The IP/ATM multicast framework

4. ATM Multicast Framework over the Satellite Links

To facilitate the exchange of IP multicast data between subnetworks, we make use of the MARS with VC mesh architecture. The IP packets are carried as ATM cells over the point-to-multipoint virtual connections between the senders' RPs and receivers' RPs (the RP of a subnetwork that has a source is termed "sender RP" or "source RP", whereas the RP of the subnetworks that have the receivers are termed "receiver RPs". An RP might be both a source RP and a sender RP, and there can be multiple in each category for the same group).

The framework is detailed below.

- A Multicast Address Resolution Server (MARS) is used to maintain a mapping of IP multicast addresses to ATM addresses. We define the MARS in our architecture to be located at the NOC.
- The satellite terminals have ATM interfaces with unique ATM addresses. These terminals are the ATM endpoints at the ATM level in the overlay network. The ATM interfaces of the satellite terminals together form an ATM cluster that is managed by the MARS. The ATM address of the MARS is known to all the ATM endpoints in the ATM cluster.
- All ATM connections go over the ATM switch located at the satellite.
- Many-to-many multicast is done over the ATM "cloud" using point-to-multipoint VCs from each source RP to the set of receiver RPs per multicast group. This therefore implements the VC mesh architecture. Multiple senders to the same multicast group, located in the same subnet, will share one point-to-multipoint VC to reach receivers in other subnets. Senders for different groups in the same subnet will use different point-to-multipoint VCs.
- Each receiver RP will terminate one branch of a point-to-multipoint VC for every external source RP to the group. If there are receivers for multiple groups in the subnetwork, the receiver RP will terminate branches of separate point-to-multipoint VCs per group and per external source RP.
- All satellite terminals that are configured to act as RPs, register their ATM addresses with the MARS on startup, following the procedure defined in Ref. 22.
- A point-to-multipoint VC exists from the MARS to all the registered ATM endpoints in the subnets - this is the ClusterControlVC (CCVC) which is used by the MARS to advertise changes to group membership for all groups.

The multicast framework is given in Fig. 9. With the above framework, the operation of a multicast group when a source becomes active is detailed in the following section. An extended description of the operation of the multicast architecture can be found in Ref. 11.

5. Creation of a Multicast Group When a Source Becomes Active

When a host in a subnetwork wants to send data to a multicast group that previously did not exist, the chain of events is as follows (refer to Fig. 10).

1) The source (host A) in subnet 1 sends the data to be multicast to its designated router (DR) for forwarding to the multicast group *G*.
2) The DR computes the (local) RP in subnet 1 for the multicast group *G* and unicasts a REGISTER message (encapsulated data packet) to the RP.
3) The RP decapsulates the data packet and creates (*, *G*) entry for group *G* in its multicast routing table.
4) The REGISTER message for the new group triggers the IP module at the RP to send a request to its ATM module to query the list of receivers for the group in other subnets.

5) The ATM module at the source RP sends a MARS REQUEST message to the MARS.
6) The MARS, on receiving the request from its MARS client, searches the local database for the mapping {IP multicast group, list of ATM endpoint addresses}. Since the group is new, no prior mapping exists in the MARS database. MARS therefore creates an entry for the multicast group in its address mapping table (and adds the ATM address of the source RP to the table entry for the group). MARS then sends a MARS NAK message to the source RP (or a MARS MULTI message with the requesting ATM endpoint address as the only member address).
7) On receiving the MARS NAK, the source ATM module waits a pre-determined delay period before sending a new MARS REQUEST to the MARS.
8) When a host B in subnet 2 wants to receive data from group G, its DR sends a PIM JOIN (*, G) message to the local RP for group G.
9) RP in subnet 2 checks that it is not part of the multicast tree for group G. It therefore creates (*, G) state for group G. It also triggers the IP module at the RP to send a request to its ATM module to register with the MARS for receiving external traffic for group G.
10) The ATM module, on receiving the request from the IP module, sends a MARS JOIN message to the MARS for group G.
11) The MARS adds the ATM address of subnet 2 RP to the list of endpoints for group G.
12) The MARS JOIN message is propagated by the MARS over the CCVC to all registered ATM endpoints. Thus the RP in subnet 1 is updated about the change in the group membership. This leads to some inefficiency since all endpoints will get the membership update information, but the information is useful only to the source RPs. We therefore propose that the MARS maintain a separate point-to-multipoint VC to only the source RPs, and inform them of changes to the group membership using MARS MULTI message format. This would require additional database storage at the MARS to differentiate between the source RPs and the receiver RPs.

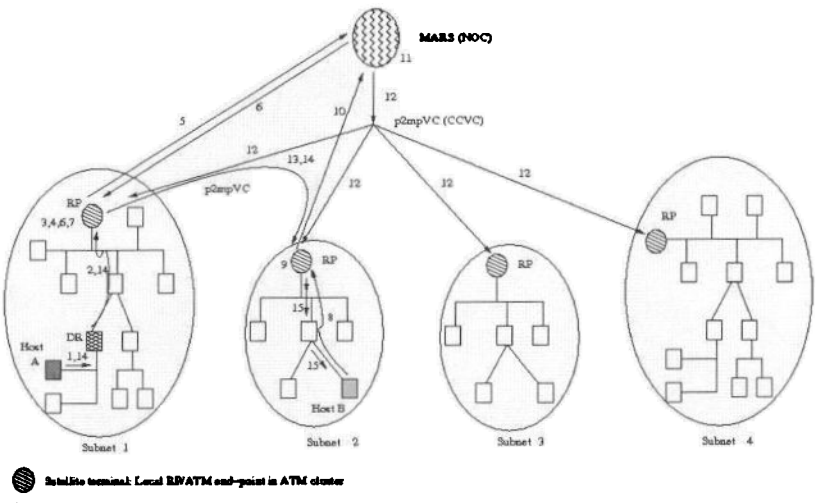


Figure 10: Creation of one multicast group across subnets

13) The ATM interface of the RP in subnet 1 gets the addresses of the receiver ATM endpoints from the MARS JOIN message. It then creates a point-to-multipoint VC over the satellite ATM switch to the set of ATM endpoints following standard procedures as given in Ref. 22. The ATM module at the source RP also sends a message to its IP module to inform the RP of the presence of receivers outside the subnet. The IP-ATM interface is therefore added to the outgoing interface (*oif*) list for the multicast group G in the local IP multicast tables.
14) Data flows in native IP format along the shared RP tree in subnet 1, to all local receivers. The packets are received by the IP-ATM interface at the source RP, where they are segmented into ATM cells and multicast to the receiver RPs over the satellite point-to-multipoint VC.
15) The ATM cells are received by the IP-ATM interface of the RP in subnet 2, where they are reassembled into the corresponding IP packet and forwarded to the IP module. The IP module forwards the packet to the PIM-SM module based on the multicast destination address. PIM-SM adds the IP-ATM interface to the incoming interface list (*iif* list) for the multicast group, and forwards the packet on the outgoing interfaces (based on the *oif* list) to the receivers along the shared tree rooted at the RP in subnet 2. The IP multicast tree is thus set up spanning multiple subnets.

IV. Routing Framework Simulation and Results

We have verified the validity and feasibility of our framework through simulations using Opnet Modeler, version 9.0[23]. The Opnet Modeler version 9.0 has support for PIM-SM, but it does not support ATM multicast. There is also no support for ATM point-to-multipoint connection. We implemented the basic MARS architecture with VC mesh in the Modeler software, and made modifications to PIM-SM and the MARS/VC Mesh specifications for our design. The implementation issues in our framework design are discussed in Ref. 11.

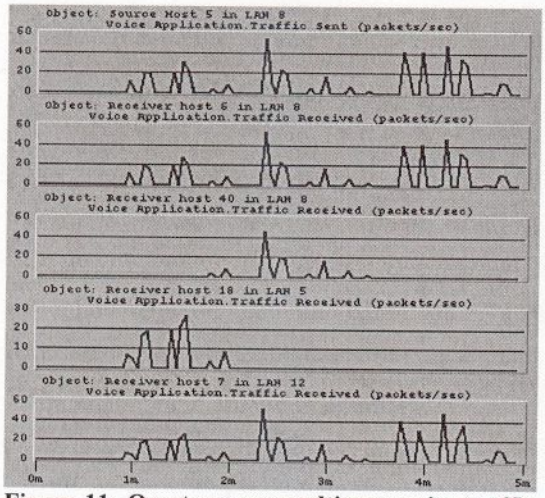


Figure 11: One-to-many multicast: voice traffic profile. *X-axis is the simulation time in minute; Y-axis is traffic sent or received in packets/sec*

For the simulation, the network configuration has 15 network domains spread over a region the size of the continental US; the domains are connected by a geostationary satellite. The MARS is located at the NOC. There are 50 nodes in each subnetwork. Each domain has one satellite gateway that acts as the PIM-SM RP for the multicast groups in its domain.

Simulations have been run for both many-to-many multicast and one-to-many multicast, with each simulation being run for 300 seconds. Simulations were done separately for voice and video traffic. For many-to-many multicast, there are 3 sources in each domain, for one group.

To compare the performance of the multicast framework, we performed simulations using the above scenario for two more cases:

1) Default PIM-SM, with a single RP for a multicast group across all domains; the RP is located in one of the terrestrial subnetworks.
2) Default PIM-SM, with a single RP for a multicast group across all domains; the RP is located at the NOC.

In both cases, there is one multicast tree constructed by PIM-SM spanning the entire network, including the satellite links. There is no ATM MARS/VC mesh architecture.

The above scenarios are selected since the end-to-end multicast tree that we attempt to build in our framework can be done using default PIM-SM; the major issue then is the placement of the single RP, which is sub-optimal in both the above cases for our large network.

The results are given in figures 11 to 14. Figure 11 gives the profile of the voice traffic sent by the source in one-to-many multicast, and the traffic received at selected group members, both in the subnet local to the source, and in remote subnets. The amount of traffic received by a host depends on the duration it is a member of the multicast group, hence some receivers get less than others. Figure 11 validates our design and shows that the framework is correct. All receivers, both local and remote with respect to the source, receive the multicast group traffic correctly, for the time duration that they are members of the group.

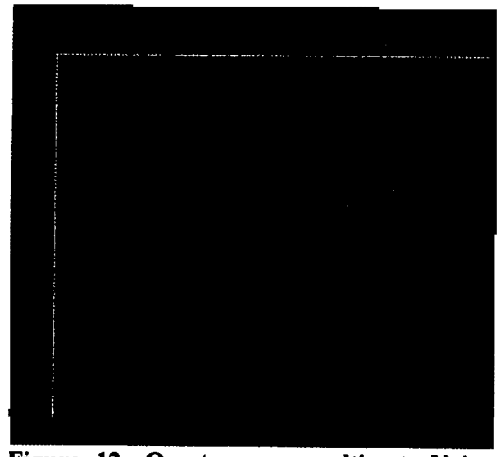


Figure 12: One-to-many multicast: Voice UBR cell loss ratio. *X-axis is the simulation time in minutes; Y-axis is the UBR CLR.*

IP multicast packets are assigned Unspecified Bit Rate (UBR) service category when they are segmented into ATM cells. Figure 12 shows the UBR cell loss ratio (CLR) in the satellite links for the three scenarios, for voice traffic in the one-to-many case. Our framework minimizes the transmission over the satellite links in comparison to the other two, and hence the UBR CLR is the least.

The end-to-end delay for video and voice applications is shown in figures 13 and 14 respectively. The perceived delay at the application is a very important criterion; our framework has less delay compared to the others, as the graphs show. This is due to minimizing the control traffic over the satellite links, and also avoiding redundant

transmissions of the data traffic. This contributes to lower overhead in creating and maintaining the multicast tree, thereby keeping the end-to-end latency lower than other cases.

V. Conclusions

In this work we have proposed a framework for IP multicast routing in a wide-area satellite network that has terrestrial Ethernet-based networks connected via ATM-based satellite links. We selected PIM-SM for the intra-domain multicast routing in the terrestrial networks; and IP-over-ATM multicast using MARS and VC mesh for inter-domain multicast routing over the satellite channels. We have proposed modifications to the protocols to adapt them to our network. Specifically, we have introduced the concept of active peer RPs for the same PIM-SM multicast group, one RP per subnetwork. We have also made additions to the RP functionality to allow seamless end-to-end multicast in a group spread across different areas. Our additions are lightweight, and do not involve any major change to existing RP functions. We have also used the MARS with VC mesh concept to do inter-domain multicasting, which differs from the "traditional" use of MARS for intra-domain multicasting. We have performed simulations of our framework, and have shown that it performs well, and compares favorably to other models.

The routing framework proposed here avoids the problem of sub-optimal placement of RPs which would happen in such a large network if standard PIM-SM is used. This has the advantage that the amount of multicast control traffic over the satellite channels is reduced significantly. If standard PIM-SM is used, with the RP for a multicast group located in a remote subnetwork or the NOC, then every control message to the RP would have to go over the satellite channels, even if the receivers are located in the local domain only. This would be wasteful use of the satellite bandwidth, and also introduce additional delay. Also, the data traffic would have to flow to the RP since the shared RP tree would remain active always. This would happen even if there are no receivers in any remote location. Our framework solves this problem very effectively by localizing the PIM-SM control messages and data traffic to the subnetworks. The amount of MARS control traffic sent over the satellite links is much less, and done once when the group is set up or torn down, instead of for every source. Also, the data traffic is sent over the links if and only if there are receivers in other locations. The design makes minimal assumptions on the satellite capabilities, and allows

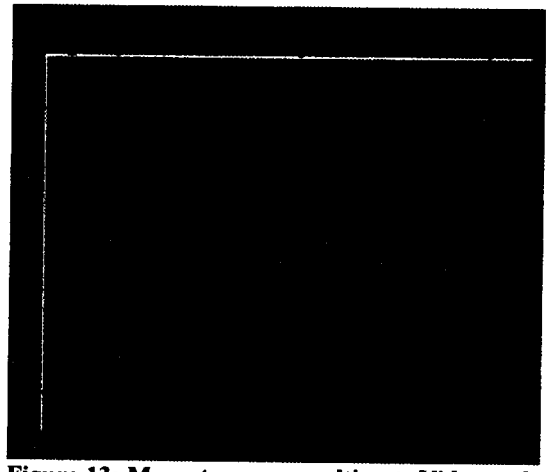


Figure 13: Many-to-many multicast: Video end-to-end delay. *X-axis is the simulation time in minutes; Y-axis is the delay in seconds.*

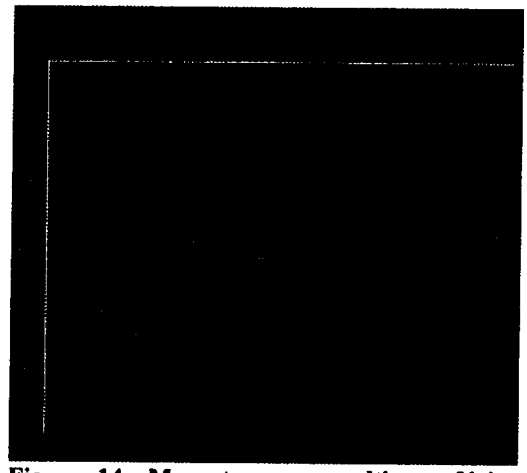


Figure 14: Many-to-many multicast: Voice end-to-end delay. *X-axis is the simulation time in minutes; Y-axis is the delay in seconds.*

the network to do IP-level multicast without requiring the satellite to support any IP stack. The framework can be adapted to networks that employ different "ATM-like" technology in the satellite links.

The work presented here does not give a quantitative analysis of the savings in control overhead in the proposed framework. We intend to do this as part of further work. Also, since we have considered IP-layer multicast routing, which is best effort, we have not taken into consideration losses due to channel error conditions, etc. However, that is an important area of research; in the future we plan to look at mechanisms for reliable transport of multicast traffic in the hybrid network, the reliable transport protocols being built upon the multicast routing framework proposed in this paper.

Acknowledgments

The first author thanks Dr. Majid Raissi-Dehkordi, Gun Akkor and Karthikeyan Chandrasekhar for helpful discussions on multicasting, and Nalini Bharatula for help with the simulation experiments.

The research work reported here was supported by a grant from Lockheed Martin Global Telecommunications, through Maryland Industrial Partnerships under contract number 251715, and by NASA under award number NCC 8235.

References

[1]Fitzpatrick, E.J., "Spaceway System Summary," Space Communications, Vol. 13, 1995, pp. 7-23.

[2]Dalal, Y. K., Metcalfe, R. M., "Reverse Path Forwarding of Broadcast Packets," *Communications of the ACM*, Vol. 21, No. 12, December 1978, pp. 1040-1048.

[3]Deering, S. E., Estrin, D., Farinacci, D., Jacobson, V., Liu, C-G., Wei, L., "The PIM Architecture for Wide-Area Multicast Routing," *IEEE/ACM Transactions on Networking*, Vol. 4, No. 2, April 1996, pp. 153-162.

[4]Armitage, G., "IP Multicasting over ATM Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 15, No. 3, 1997, pp. 445-457.

[5]Paul, S., "Multicasting on the Internet and Its Applications," Kluwer Academic Publishers, Boston, USA, 1998, pp. 9-127, 359-365.

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

[7]ATM Forum, "ATM User-Network Interface Specification Version 3.1," Prentice Hall, Englewood Cliffs, NJ, June 1995.

[8]Ballardie, A., Francis, P., Crowcroft, J., "Core Based Trees (CBT): An Architecture for Scalable Inter-Domain Multicast Routing," *Proceedings of ACM SIGCOMM '93*, September 1993.

[9]Thyagarajan, A. S., Deering, S. E., "Hierarchical Distance-Vector Routing for the MBone," *Proceedings of ACM SIGCOMM '95*, October 1995.

[10]Akkor, G., Hadjitheodosiou, M., Baras, J.S., "IP Multicast via Satellite: A Survey," Center for Satellite and Hybrid Communication Networks, University of Maryland College Park, Technical Report CSHCN TR 2003-1, 2003

[11]Roy-Chowdhury, A., "IP Routing and Key Management for Secure Multicast in Satellite ATM Networks," Master's Thesis, Electrical and Computer Engineering Dept., University of Maryland College Park, MD, 2003.

[12]Deering, S. E., "Host Extensions for IP Multicasting," Internet RFC 1112, URL: http://www.ietf.org/rfc/rfc1112.txt [cited November 2003].

[13]Ramalho, M., "Intra- and Inter-domain Multicast Routing Protocols: A Survey and Taxonomy," *IEEE Communications Surveys and Tutorials* [online journal], 3(1), URL: http://www.comsoc.org/livepubs/surveys/public/1q00issue/ramalho.html [cited November 2003].

[14]Cain, B., Deering, S., Kouvelas, I., Fenner, B., Thyagarajan, A., "Internet Group Management Protocol, Version 3". Internet RFC 3376, URL: http://www.ietf.org/rfc/rfc3376.txt [cited November 2003].

[15]Elizondo, E., Gobbi, R., Modelfino, A., Gargione, F., "Evolution of the Astrolink System," In *Proceedings 3rd Ka Band Utilization Conference*, 1997, pp. 3-7.

[16]Moy, J., "Multicast Extensions to OSPF (MOSPF)," Internet RFC 1584, URL: http://www.ietf.org/rfc/rfc1584.txt [cited November 2003].

[17]Waitzman, D., Partridge, C., Deering, S., "Distance Vector Multicast Routing Protocol," Internet RFC 1075, URL: http://www.faqs.org/rfcs/rfc1075.html [cited November 2003].

[18]Hedrick, C., "Routing Information Protocol," Internet RFC 1058, URL: http://www.ietf.org/rfc/rfc1058.txt [cited November 2003].

[19]Fenner, B., Handley, M., Holbrook, H., Kouvelas, I., "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)," Internet Draft, URL: http://www.ietf.org/internet-drafts/draft-ietf-pim-sm-v2-new-09.txt [cited March 2004].

[20]Handley, M., Crowcroft, J., Wakeman, I., "Hierarchical Protocol Independent Multicast," University College London, November 1995. Work in progress.

[21]Thaler, D., "Border Gateway Multicast Protocol (BGMP): Protocol Specification," Internet Draft, URL: http://www.ietf.org/internet-drafts/draft-ietf-bgmp-spec-06.txt [cited March 2004].

[22]Armitage, G., "Support for Multicast over UNI 3.0/3.1 based ATM Networks," Internet RFC 2022, URL: http://www.faqs.org /rfcs/rfc2022.html [cited November 2003].

[23]Opnet Inc., Opnet Modeler Software, Ver. 9.0.A., Bethesda, MD, 2002.