ISSN:0975-9646

A Contemporary Study of Application Layer Multicast Protocols in aid of Effective Communication

M. Anitha[#1], P. Yogesh[*2]
[#] *Department of Computer Science and Engineering,* [*]*Department of Information Science and Technology*
Anna University College of Engineering, Chennai, India.

Abstract **– IP multicasting, on par with unicast, is a powerful solution to one-to-many and many-to-many communication. The challenges faced while implementing it includes slow deployment due to the infrastructure changes needed and the address resolution problem. Hence this does not suit the current growing applications over the internet such as peer-to-peer file sharing system, which requires multimedia data to be transmitted across peers. End System Multicast (ESM) or the Application Layer Multicast (ALM) becomes the alternative solution to overcome the challenge of overheads incurred in the traditional multicasting. Unlike IP multicast the end system multicast does not require any infrastructure changes. This easy deployment feature helped ALM gain popularity. Many recent research works have been carried out to improve the performance of multicasting multimedia in applications such as video conferencing, Video-on-demand, gaming etc. Many protocols for ALM have been proposed by researchers taking into account the current trends in internet usage. In this article we have described a set of ALM protocols and classified them based on some of its characteristics. We have also compared its performance based on some of its properties, with the common evaluation metrics being scalability, stress, stretch and latency.**

Keywords: Application Layer Multicast, IP Multicast, Wireless Mesh Network

I. INTRODUCTION

The Internet today has grown into a vast repository of knowledge. This low cost technology acts as a gateway to huge quantity of information which leads to a dramatic change in data communication. As the number of multimedia users keeps increasing, there is a huge traffic in the network because of the huge volume of data transmitted. The frequency with which a user needs to send the same packets is very high. The fact that there are more number of packets with same content transmitted in the network makes it more congested. Applications such as audio/video streaming, video conferencing, online gaming used by multimedia users involves multiple receivers and a single sender or multiple senders and multiple receivers. The way the Internet communicates can be broadly classified into four types, namely one-to-one (Unicast), one-to-many (Multicast), many-to-many (Broadcast) and one-to-one-of-many (Anycast).

Multicasting makes it possible for the sender to send packets over the network only once to multiple users. In this, if the paths from the source to the destinations are same, then only one packet is sent along the network. But if there are different paths for different destinations then the packets are duplicated onto the routers and forwarded to the next hop of the network. This is the fundamental concept of IP multicast. This was introduced by Steven Deering in 1988 [1]. This leads to an efficient utilization of resources in the network. Successful working of IP Multicast hinges on the ability of the network to intelligently route the packets across the network so that it reaches the desired destination. It is the responsibility of the routers to set up and also in tearing down the multicast sessions. The hosts with interests in a particular group should inform the routers to which they are attached to, to join their group. This process is accomplished by Internet Group Management Protocol (IGMP).

Even though IP multicast is more efficient, there are many disadvantages which include the slow deployment. This drawback makes the IP multicast not suitable for the increase in multi user applications like gaming, Video on Demand (VoD) and Video Conferencing etc. So, to handle such multi user applications Application Layer Multicasting (ALM) emerged as a solution [2],[3] [4]. In this technique, the end hosts are responsible for the forwarding and duplicates the packets when needed. The nodes participating in multicasting forms an overlay network. Application Layer Multicast does not require any change in the underlying network infrastructure [4].

This survey gives an overview of the ALM protocols and its working principles. The ALM protocol can be classified based on the arrangement of the end hosts in the overlay network as mesh first, tree first and hierarchical clustering approaches. Each classification has different protocols having different methodologies for the arrangement of the end hosts. They almost tend to optimize the performance metrics of ALM like, latency, bandwidth, Relative Delay Penalty (RDP) and the most important parameter stress and stretch. However variations exist among these protocols in terms of efficiency and overhead. In this paper, we are going to highlight them in terms of their relative advantages and disadvantages.

II. IP MULTICAST DEPLOYMENT CHALLENGES

The major problem of IP multicast is its deployment [4]. The challenges include:

Group Management:

A key decision on how to manage a group of nodes in a multicast session must be made b y a protocol designer after the decisions on application domain and deployment level are made. Basic group management deals with users identifying the multicast sessions, joining the session, leaving the session and about the contribution to the session.

Address allocation:

From a globally shared address space, assigning each application a unique address is referred to as multiple address allocation. There are 228 distinct addresses in the

current IP protocol version IPv4. Due to the lack of address allocation mechanism, ISPs face a threat as they have to deal with angry customers those who are forced to carry unwanted data.

Network Management:

In comparison to inter domain where Rendezvous Point (RP) and associated sources lie in different domains Intra domain multicast deployment is relatively easy. Source pruning for specific multicast groups as well as source specific joins are provided by the Source Specific Multicast (SSM) Model and IGMPv3. Group management, network management and address allocation problems are to some extent alleviated by SSM model.

Multicast Security:

Multiple entities participate in the multicast session but they do not have any trusted relationship with others which complicates the process of providing security. Authentication, authorization, encryption and data integrity are some mechanisms that are provided.

Lack of proper business model:

The deployment of IP multicast becomes slow because of the lack of proper business scenarios. Multicast motivates Internet Service Providers (ISP) as it leads to considerable saving of bandwidth which is more costly when compared to the deployment and management costs. Initial cost is not high in multicast. The large collection of servers and available network bandwidth attracts a very large audience. A new protocol was developed and named as IPv6, which will solve the problem of address allocation in IPv4. The 32 bit unique addresses will be completely exhausted in 2008 and 2018 as per report of the two leaders of IETF's working group [5]. The solution is to replace IPv4 routers with IPv6.

The IPv6 capable routers are compatible with IPv4, but vice versa are not possible. This has to be kept in mind as a practical implication in implementing the Ipv6 [6]. Replacing few hundreds of the many IPv4 systems takes much time by announcing a flag day when all the Internet systems will be shut down. But this is not possible for the current Internet systems, as there are millions of users and administrators. One approach is making the IPv6 system to follow dual stack approach where, the IPv6 systems will have a complete implementation of IPv4 referred as IPv6/IPv4 system [7]. Another approach is the tunneling concept [7]. Tunneling occurs when two IPv6 systems communicate with each other having intervening IPv4 systems, then a tunnel is created so that the intervening IPv4 systems need not worry about the payload. The entire IPv6 packet will be as a payload of IPv4 packet, which will be handled as a normal IPv4 packet by the IPv4 intervening system. Only the IPv6 router will come to know that the IPv4 packet contains a Ipv6 packet with the original payload. But the problem is how fast the deployment of IPv6 is going to happen or it may not happen at all [8]. Then ALM became the solution for these problems to multicast the data among a group. ALM follows a similar concept of tunneling but it happens between the end hosts who participate in the multicasting.

III. ADVANTAGES OF ALM DEPLOYMENT

Today's Internet is loaded with multimedia information, which cannot be handled by the current Internet infrastructure. ALM protocols provide a conducive way to transmitting data even though its efficiency is less when compares with the IP multicast. ALM protocols are used in many applications like video conferencing, Video on Demand, Video surveillance, gaming, etc. ALM does not require any support from the network infrastructure. Packets forwarding, multicast functionalities, multicast tree construction and group formation are moved to application layer [9].

ALM provides a new way of overcoming the deployment issue in IP multicasting over heterogeneous network. This technique used for multicasting addresses scalability issue and also improves the performance of the network [10], which is achieved through the overlay network constructed using the end systems. This is a logical network built on top of the end hosts. The ALM protocols are developed to handle the topology change, occurring due to the end hosts joining and leaving at their wish, which leads to a dynamic network. The applications in the recent trends require multi sender and multiple receivers. But, this has to be done without overloading the existing network infrastructure. ALM comes as a definite solution at a lesser cost and also speeds up deployment. These protocols work with the users than with routers. They are capable of handling data trans-coding, error recovery, and flow control based on the application involved [11]. End System Multicast are independent of routers, whose deployment is done by end systems using some application codes [12].

IV. CATEGORIES OF ALM PROTOCOLS

The deployment of ALM protocol [13] can be made both at the infrastructure level and at end system level. Specific and dedicated servers/proxies are required for infrastructure level, where they self organize into an overlay network. Only the unicast service from the infrastructure is expected by the ALM protocols and all the multicasting functionalities are handled by the end hosts. It is the business and marketing issues that drive the choice between infrastructure level and the end system level rather than purely technological ones. The existing Internet infrastructure available to them is used by the end systems which share the forwarding load of a multicast session. Unlike in the case of infrastructure level it does not expect to pay more in this case. The efficiency in proxy based infrastructure is increased by including a representative of an existing IP based islands to construct overlay network. The disadvantages in the deployment of proxies over the internetwork are additional costs, reducing adaptability and lesser optimization for specific applications [13]. End system ALM enjoys, more flexibility, adaptability to specific application domains and immediate deployment over the Internet. But it may not scale well. It requires end system to take the responsibility of forwarding. It should deal with low bandwidth of end systems. There are two basic approaches for transmitting the data in the path: mesh-first and tree-first. In mesh-first topology, mesh is explicitly created at the beginning and the tree is created later using source as the chosen root. The source specific trees are

constructed by using any of the IP multicast protocols eg: - Distance Vector Routing Multicast Protocol (DVMRP) [14]. The quality of the tree constructed depends on the quality of the mesh chosen. The tree topology can be determined by the proper selection of the mesh neighbours and the metrics. It is more suitable for multi-source applications with high control overhead.

In the tree-first approach, the tree is directly built where the members select their parents from the known members of the tree. The tree first approach gives a direct control over the tree unlike the mesh-first approach. It provides a control over the selection of the parent neighbours with enough resources [4]. It also provides independent actions that can be applied to each member. Moreover, it also lowers the communication overhead. Whenever a node changes its parent, its descendants are not aware of that and hence they are dragged towards the new node by their parent which makes the tree uneven and less efficient. Reorganization of tree is a challenging issue in tree first approach. Any node in the overlay may leave the tree either gracefully or abruptly. The children attached to that node are affected. The children should be attached to a new parent to further receive the data. This process takes some time. This delay incurred requires some buffering in each node. This buffering technique is more complex in transmitting multimedia data eg: - Video on Demand. Tree structure has fundamental limitations both for high bandwidth multicast and for high reliability. Bandwidth is guaranteed to be monotonically decreasing moving down the tree. Any loss high up the tree will reduce the bandwidth available to receivers lower down the tree [15].

Multicasting based on clustering is one more classification of ALM protocols which addresses the scalability issue of the ALM. These types of protocols are also classified as implicit protocols [10]. The data topology is embedded in the clustered hierarchy. The clustering based ALM protocols support more number of nodes. The protocol selects a cluster head from each cluster and that node takes the responsibility of its cluster members. This cluster head periodically checks for its cluster size and decides on the changes if required.

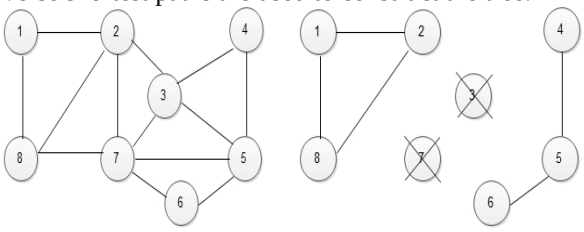
V. ALM PROTOCOLS

End System Multicast protocols or ALM protocols addresses many issues related to IP multicast. The data packets are sent as unicast packets through the tunnel created between the hosts participating in the multicast. This strategy accelerates the deployment of ALM [3]. All ALM protocols organize the members into data and control topologies. The control topology is responsible for the node's join/leave process. The data topology is a subset of control topology which provides the information related to the data path that is used for the transmission of the information among the group members.

V.1. Mesh First Approach

1) Narada

In this protocol Narada [4], [16] every node in the mesh maintains a list which contains all the states of the other members in the group. Whenever a new member wants to join the group, it obtains the current list of the members who belong to the group. After receiving the list from the rendezvous point using the bootstrap method the new member sends a join message to several randomly chosen members from the list. The new member joins the group as soon as it receives the acceptance message from any one member of the group. It starts updating the table and this information is forwarded to all the members of the group. Similarly if a member wants to leave the group it sends a leave message to all its neighbours and this information is also forwarded to all the members of the group. The protocol follows a source specific multicast tree to forward the data which is computed by the members of the group which run a variant of a Distance Vector Protocol (DVR). Reverse shortest paths are used to construct the tree.



(a) Sample Overlay (b) Mesh Partition

Fig.1 Narada Failure Mechanism

The main advantage of Narada is its robustness. The members failure or the act of abruptly leaving the group is handled as a fail-stop failure model [16]. As an example from Fig. 1b, if two nodes say 3 and 7 fails, leads to mesh partitioning. Members in the group maintains the list of nodes and periodically checks for the response from the failed nodes. This process is done for a particular amount of time, after which the nodes are declared as dead nodes. The nodes which detected the partition send probe messages to the failed nodes with certain probability value. The probability should be chosen carefully, so that the loss of probe messages or the replies is minimal.

Narada does a periodical refinement of the mesh because of the factors like; initial selection is random, partition repair, dynamic behaviour of the users and dynamic conditions of the underlying network which makes the mesh to be not an optimal one. Each member computes the cost and utility of the existing link as well as the link to a random node which is not its neighbours.

2) Scattercast

Scattercast proXies (SCXs) [4], [17] uses a protocol named Gossamer to self organize the mesh structure. A new mesh is generated for every multicast session. This protocol works for fixed number of SCXs. The architecture consists of SCXs which are placed strategically around the Internet. The overlay is constructed between the SCXs using the unicast links. A cluster is formed with three components namely, front end, network module and a collection of SCXs. Front is used to interact with the client and the network module helps in constructing efficient overlay. The new node which wants to join the group first contacts the nearest cluster and requests for an SCX. Then the cluster checks for the SCX and if does not find any, then it sets up a new SCX and sends the IP address of the new SCX to the new member. The SCXs runs a Distance Vector Routing

(DVR) protocol on top of the mesh to construct the multicast delivery tree which is a source specific tree.

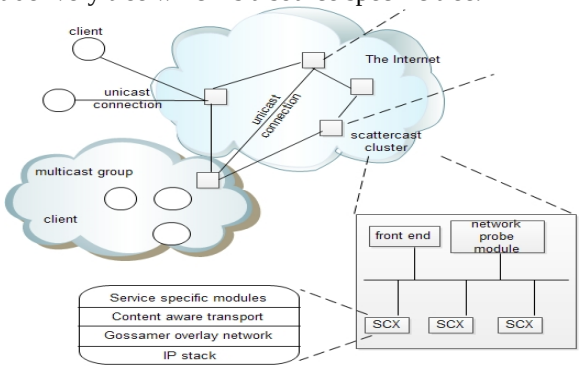


Fig. 2 Scattercast architecture

The fig.2 shows that the architecture consists of three main components. The constraints in constructing the tree are the degree of the SCX and the delay between the source and destination. To optimize the mesh the SCX probes periodically and compares the potential links and the existing links. The SCXs maintains a routing table which consists of costs associated with the paths. The cost of a path is the sum of the costs of all links along the path. If the new cost computed on optimizing the mesh is lesser than the existing link the new cost is substituted for the old one. To avoid loops in the route instead of just maintaining the routing cost it also maintains complete information of all the paths like in Border Gateway Protocol (BGP) [18]. If mesh partitions are detected then RP is chosen randomly. RP will send periodically refresh messages, if any of the SCX does not receive the message it contacts that particular RP and reconnects the mesh. Since the ScatterCast network is composed of SCXs located at IP end-points rather than within IP routers, the path taken by a ScatterCast data stream will inevitably incur additional latency [17].

3) RMX

Reliable Multicast proXies RMX [19] is a hybrid approach for reliable multicast communication. The heterogeneous receivers are split into small number of homogeneous receivers. A divide and conquer approach is used. The RMX allows for the notion of semantic reliability as opposed to data reliability, that is, reliability of information rather than that of the representation of the information. The RMX architecture builds on the ScatterCast [17] model by integrating application-specific intelligence and semantics into the forwarding service. A node that needs any data first sends the request to the local data group. If the data is not available then the request is sent to the next higher level in the hierarchy.

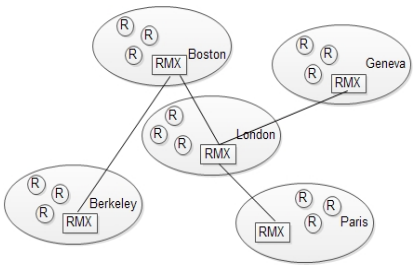
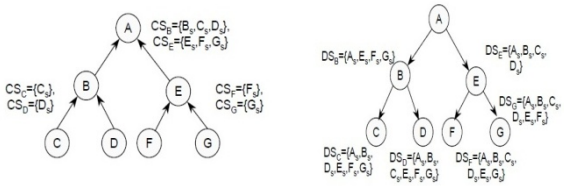


Fig. 3 RMX architecture

Fig. 3 shows the connectivity among the RMX's. The data forwarding is done b y the Scalable Reliable Multicast (SRM) protocol [20]. Buffer management is done at each RMXs' so that they do not loose data. Only static placement of RMX is considered. Network failure recovery is not handled.

3) Bullet

Bullet, a scalable and distributed algorithm that enables nodes spread across the Internet to self-organize into a high bandwidth overlay mesh [15]. The sender who wants to send the data splits the data into individual blocks. These blocks are further divided into objects. These objects are sent to different points in the network. The nodes receive some data from their parents. It is node's responsibility to obtain the missing data which is done by using a distributed algorithm. This uniformly spreads the data across the overlay. Bullet nodes self organizes into overlay tree. Starting point is the root of the node. The data transmitted to the children is a disjoint set. RanSub [21] is the approach used to locate the missing data. RanSub distributes random subsets of participating nodes throughout the tree using collect and distribute messages. Collect messages traverses from the leaves towards the root and the distribute messages just in the opposite manner.



(a) Collection phase (b) Distribute phase

Fig. 4 Ransub Phases

There are two phases of this process. They are the distribute phase and the collect phase. In the former phase a random set of participants are distributed to the nodes, the latter sends the collected data up in the tree. As an example the fig. 3a, shows the collect phase and the fig. 3b shows the distribution phase. Bullet uses a TCP Friendly Rate Control (TFRC) protocol [22] as the transport layer protocol. Bullet is capable of functioning on top of essentially any overlay tree. A Bullet receiver views data as a matrix of sequenced packets with rows equal to the number of peer senders it currently has. Apart from eliminating the overhead required in traditional distributed tree construction it also achieves throughput twice as that of the traditional bandwidth tree.

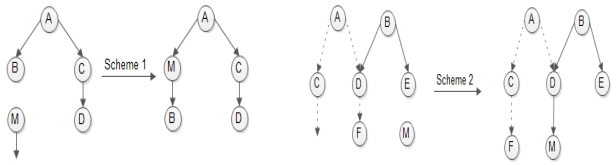
4) HOMA

The protocol Heuristic Overlay Multicast Algorithm (HOMA) [23] is an application layer multicast protocol. It uses a heuristic routing to construct efficient multicast trees on the application layer. This protocol supports small scale multi party video conferencing applications. As it is for small group of participants scalability is not an issue. Hence a centralized approach is followed for group management by the Rendezvous point (RP) which itself is a

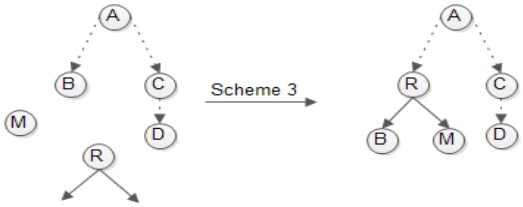
member in the conference. Back up RP is used to avoid single point failure. Cost is not a major problem for a small group video conference. The desire is to achieve lower end-to-end delay. Due to this reason source specific trees are constructed to deal with multiple data sources.

Conferencing applications require low latencies and high data rate between end hosts. So, multicast trees should be constructed considering both delay and bandwidth requirements. In order to ensure, the effective utilization of the available bandwidth of the conference members efficiently, the participants are not allowed to watch the video of all other conference members simultaneously [24]. Hence there is no need for transmitting videos. An efficient overlay multicast routing algorithm is implemented in HOMA to construct multicast trees. At the starting stage of the conference session all the members are connected as full mesh and whenever the participants are interested in viewing the video they send a request to RP. Then RP will try to attach the node with the multicast tree.

Fig. 5a shows that the new node is attached to the existing multicast tree as it has available out-degree. If more than one source has the capability of accepting a child, then these sources are compared and the one with the best utility is chosen. If requesting node with the available out degree chooses a parent to join the tree and if it checks that the parent is not having available out-degree, then it swaps itself with the chosen parent as node M does. Fig. 5b shows that if another tree has an out-degree space it is shifted to that tree if it cannot join in the chosen tree B. The third scheme 5c uses a reflector to accommodate the attachment of the new member which assists in forwarding its flows, even if it is not part of the receiver set of a source [25]. Leave requests are also handled when a participant wants to discontinue watching a video or change the video source. If the leaving node is a leaf or has one child, it is simply removed from the tree. Otherwise, if it has two or more children, the leaving node will stay in the multicast tree as a reflector. The performance metrics used are rejection rate and Relative Delay Penalty (RDP).



(a) Scheme 1 (b) Scheme 2



(c) Scheme 3

Fig. 5 Topological changes in HOMA

Table I shows the comparison of various mesh-first protocols discussed in this section. NARADA and HOMA are ALM protocols which support small group size and hence scalability is not an issue. This leads to less control overhead. This can be managed by the protocol. But the major challenge of NARADA is the dissemination of changes in the mesh as the information should reach all the members of the group. NARADA's failure recovery mechanism is less efficient when compared to the other protocols as it uses the group members for the failure recovery. All other mesh first based protocols use multiple RPs. The information about the changes in the mesh happens to be the responsibility of RP which relieves the group members from the burden. The RP's may be a group member or any central control. To avoid single point failure the two protocols use multiple points as a backup RP. The ScatterCast and RMX are hybrid approaches where the protocols maintain the properties of IP multicast even though they are application layer multicast protocols.

V.2. Tree First Approach

1) YTMP

Yoid [27], [28] is a suite of protocols called as Yoid Topology Management Protocol (YMTP). It allows the end hosts to replicate and forward the data required for distribution for a given application. It directly creates the data delivery tree. It has a direct control over various aspects of the tree structure. Yoid generates two topologies. One is the shared tree and the other is mesh topology. The two topologies are created and optimized for different purposes. The tree is optimized for efficiency and the mesh is for robustness. Yoid is stack of protocols consisting of identification, transport and distribution protocols. The tree-first approach protocol creates a shared tree and the members interested should find their parents b y themselves. The degree bound of each node limits the number of children to that node.

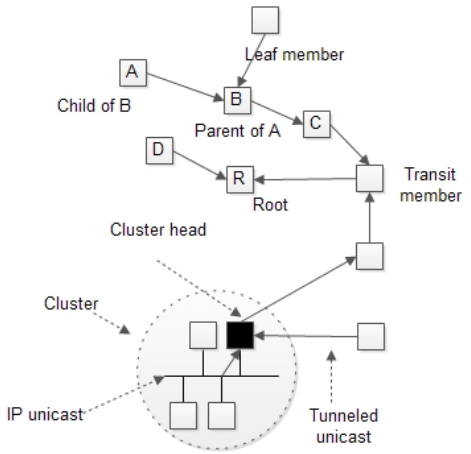


Fig. 6 Components of Yoid tree

Fig. 6 shows the components involved in a Yoid tree in which the boxes represent a member. The lines represent the links in the tree. The arrows represent the relationship between the members in the tree. It has one or more rendezvous host. If a new member wants to join it has to contact the Rendezvous host which is not in the group. It will respond with a list of members that are already part of the multicast group in the tree. Then the member queries the

list of members to find its parent. The conditions for joining the particular parent are that the newly joined node should not form a loop in the tree and also the newly joined node should have the degree for attaching the children to it. The Yoid chooses the best potential parent from the available potential parents. If the new member cannot find any potential parent then it announces itself as the root of the shared tree. During the changes, the tree may get partitioned. In such situations, one member in each tree partition will declare itself to be the root. In that case, the RP arbitrates in merging the different tree fragments. Periodically each member seeks other potential parents for better points of attachment in the shared tree. Yoid incorporates loop detection and avoidance mechanisms when members change parents in the tree.

2) HMTP

Host Multicast Tree Protocol (HMTP) [27],[12] is a hybrid application layer multicast protocol. It uses the tree-first approach and has some similarities with the Yoid protocol. Yoid protocol creates the mesh explicitly. But this protocol does not do that but instead each member in the tree maintains list of some members who are in the group. It periodically updates this list. The protocol automates the interconnection of IP multicast enabled islands and also provides multicast delivery to end hosts where IP multicast is not available. The member (router) in the IP multicast acts as a Designated Member (DM) for an island. These DMs are responsible for tree construction and tree maintenance.

In tree construction, members in HMTP are responsible for finding parents on the shared tree. Two members are said to be neighbours if they are connected by a tunnel. The path from the leaf to the root is called as root path. Each multicast group members requires a Host Multicast Rendezvous Point (HMRP). The security and group policies are implemented in the HMRP. This RP always knows the root of the tree. A node who wants to join the group has to send a request to the HMRP which will provide the information of the root of the tree. Starting from the root, at each level of the tree it tries to find a member, close to itself. If the number of children of a member which is close is less than its degree bound, then the new member joins as a child to its identified parent. If it is not able to identify the potential parent it proceeds to the next level and tries to find a potential parent among the children of the closest node.

When a member wants to leave the group it has to inform its parent and its children. The parent node deletes the information of the node from its list. The children do not have capability of finding a new parent as they do not have all the information of the nodes in the tree. So, it is sole responsibility of the leaving node to find the new parent for its children. Fig. 7 below shows the architecture of the HMTP.

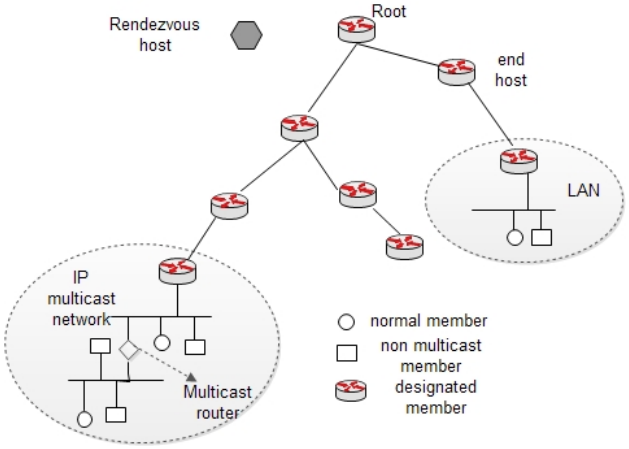


Fig. 7 Host Multicast tree protocol

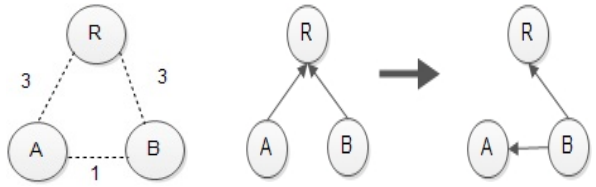
Members on its path to the root. Periodically, each member tries to find a better (i.e. closer) parent on the tree, by re-initiating the join process from some random member on its root path. Knowing the entire root path allows members to detect loops. HMTP employs a loop detection and resolution mechanism, instead of loop avoidance. Unlike Yoid, HMTP does not explicitly create a mesh. However, each member periodically discovers and caches information about a few other members that are part of the tree. In the specific case when the RP is unavailable, the knowledge of such members is used to recover the tree from partitions.

3) BTP

ALM Protocol	Group/Tree Management	Tree building algorithm	Applications	Failure recovery mechanism	Evaluation metrics
Narada	Group members	DVR	Video Conferencing	Group members	Latency, Bandwidth, Stress
ScatterCast	Multiple SCXs	DVR	Internet MP3 radio and Electronic white board for online presentations	Multiple SCXs	Average latency, Cost ratio variation
Bullet	Underlying tree overlay	Gossip algorithm	Large file transfer and Multimedia streaming	Underlying tree overlay	Bandwidth, Scalability
RMX	Local data group	SRM	Shared electronic white board, Infocast [26]	-	Data loss
Homa	RP	Greedy algorithm	Multiparty video conferencing	Backup RPs	Rejection rate, RDP

TABLE I Mesh First Approach Protocol Comparison

Banana Tree Protocol (BTP) [4], [10], [29] is a simple protocol. This protocol was designed for the file sharing program, Jungle Monkey (JM) [30]. The implementation of this protocol requires two more supporting protocols: Banana Tree Simple Multicast Protocol (BTSMP) and Banana Tree File Transfer Protocol (BTFTP). The former is used to advertise the files that can be downloaded and the latter is used for one-to-many distribution. It is tree based topology where the root is the source that created the tree. It uses a bootstrap mechanism for the join process. The node who wants to join the group needs to first contact any of the node already existing in the group. That node becomes the parent of the new node. If there are no nodes existing in the group then the new node becomes the root node. Fig. 8a shows a simple topology with three nodes. The nodes change their parent in order to optimize the tree. This is known as parent switching. This reduces the tree cost and latency. The switching takes place only between the siblings and the grandparent. This helps in avoiding loops. Fig. 8 illustrates the sibling's switch (eg:- A switches to B). If a sibling is not found then the switching takes place with the root R. Each node who needs to change their parent receive



(a) Simple Overlay (b) Switching

Fig. 8 BTP optimization Technique

the list of siblings and grandparents and then it checks for their closeness. The sibling list can be obtained by IDMaps [31]. If the chosen sibling or the grandparent is closer than that of its parent then the node will send a request message to its newly chosen node. When a node is in its parent switching stage it does not allow any other nodes to choose it as a new parent.

4) ALMI

Application Level Multicast Infrastructure (ALMI) [4], [11] is a tree based centralized protocol. The approach in this protocol simplifies the routing process. It is suitable for only small size group applications. ALMI consists of a session controller who computes Minimum Spanning Tree (MST) and session members. These two components of the protocol communicate between users using the control plane. The session controller may be a session member or a special purpose server. The tree is the data delivery path. Each member in the group reacts to the latency to a set of members because latency parameter is optimized in this protocol. The controller reconstructs the tree based on the latency updates received from the members over a time period. It also updates the information of the reconstructed tree to all the members in the group.

Any new member who wants to join the group has to first approach the controller. The controller creates a node ID for the new member and sends this information along with the ID of the node which will act as a parent of the new node. The new node then sends a GRAFT message to the parent and starts receiving the data. Sometimes loops are caused due to loss of update information of the tree and also members may have different versions of MST. So the control data is not consistent in the group. To avoid the confusion over different versions of MST, a number is assigned to each version which is assigned to the tree incarnation field of the packet structure. The members send the packet with tree version number. The number is maintained in a cache so that a node can verify the MST number through which it receives the data. If the number is not the same as that of the number in the cache then the node approaches the controller for an updation of the new MST. Additionally ALMI has an error recovery mechanism which is supported by data naming interface.

5) Overcast

Overcast [4], [32] is a single source multicast protocol used to build an efficient multicast tree. The bandwidth is the optimizing parameter. Latency parameter is not considered because Overcast is not intended for interactive applications. The new member is placed farther from the root without affecting the bandwidth. A periodical evaluation is done by each node to ensure its position by measuring the bandwidth to its siblings, parent and grandparent. Due to this, Overcast is tolerant towards the root node failure circumstance. If root node failure becomes an issue then backup parents or backup tree may get implemented. The new node contacts the root and takes it as the current parent and computes the direct bandwidth to it. If the bandwidth of the children is more when compared to its parent then the child node becomes the potential parent and the iteration starts again. When more than one child is eligible to change as a potential parent then the proximity measure is taken so that the closest node will become the new potential parent. Each member periodically sends refresh messages to its parents. If the parent does not receive any message from the child for a particular time period then the child and its descendants are considered as dead.

Up/Down protocol is used to know the status of the node in the tree which requires some statistical information. Each node in the tree maintains a table consisting of the details of the nodes lower in hierarchy. This leads to the situation where the root node maintains all the information related to all the nodes in the tree.

6) TBCP

Tree Building Control Protocol (TBCP) [4], [33] is a protocol used to build overlay spanning trees among the multicast group members. It is a degree constrained protocol. The strategy of the protocol is the placement of the node. Node's out-degree is checked and the number of children is decided. The root is the main sender of the data to the group. When a new node needs to join the group it sends a request message to the root node. The root node responds to

the request message with the list of children. Then the new node computes the distances between itself and the root and also the distances between itself and all the children nodes specified by the root. This information is sent to the root where it selects an optimal configuration for the new node after evaluating all the possible configurations. The node then joins the group if everything fits in well. Otherwise the new node has to restart the joining process from the node from where it was redirected.

The existing nodes perform a join procedure in order to improve the performance of the tree thereby optimizing the tree. In order to increase the efficiency of the tree, receivers of the same domain are grouped under a single sub-tree. A score function is defined as, the maximum value relating the parent, the new node and the children listed by the parent to the new node based on the distances.

7) Delaunay Triangulation Protocol

The Delaunay Triangulation protocol constructs overlay topology using the logical addresses drawn from the coordinate space [34]. Routing protocols are not necessary for the construction of the multicast tree. Scalability is achieved through the distributed implementation so that no entity needs to maintain the information of the entire group. A set of vertices form a triangulation graph, in which, if a circle is drawn it encloses three vertices of a graph. No node will be found inside the circle. Delaunay Triangulation property is used for the overlay construction. This property is explained in [35]. The geographical locations are assigned to each node which is used to connect the nodes as a topology. The connected topology should satisfy the Delaunay property.

A new node (N) who wants to join the group sends a request with its coordinate space to DT server. The server replies with a message containing any node (X) already in the group. The message will contain the logical and physical addresses of the node. The server also ensures that the coordinate space of the new node (N) is less than that of the chosen node (X). Then X, using the DT property goes for neighborhood test for the new node. If N fails the test then X forwards the request to another node (Y) in the group. If N passes the neighborhood test then Y is closer to N than X, else Y will again forward it to some other node (D) which will be closer to N than Y. If N passes the test then N becomes the candidate neighbours of D. The new node sends heartbeat messages to all other nodes so that it gets promoted to neighbours from being candidate neighbours. When a node (N) in the group wants to leave the group, then sends goodbye messages to all its neighbours and the DT server. The server and the neighbours removes N from their list. Node leaving the overlay network disconnects the nodes i.e., some nodes fails in the neighborhood test. This is rectified by sending Hello Neighbours messages to each other.

8) Bayeux

Bayeux [4], [36] is an efficient source specific ALM protocol. Like Tapestry [37] this protocol also uses a prefix based routing system. The methodologies for wide area location and the routing architecture are taken from Oceanstore [38]. Bayeux multicast session is identified by a session name and a unique ID for each instance of the session. This information together forms a tuple to identify a multicast session. Each session is secured by some security algorithms like SHA-1 [39]. A new node who wants to join the session should get the tuple and request to the root for the further process. Fig. 9 shows the joining procedure and tree maintenance.

The protocol uses Tapestry location mechanism to avoid single point failure as the root is responsible for the join/leave procedure. Packet duplication is also avoided by creating a cluster for the receiver's ID. The idea behind this is to get a longest suffix ID which is shared by different nodes. This allows the protocol to send only one packet. First Reachable Link Selection (FRLS) is the protocol used by Bayeux for packet delivery. This choice reduces duplication of packets.

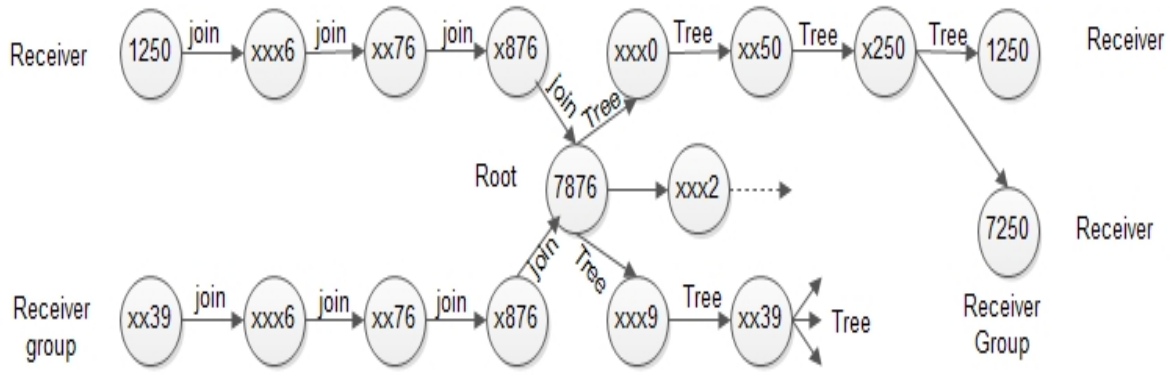


Fig. 9 Joining procedure and tree maintenance

9) Hypercast

Hypercast [4], [40] is a protocol which organizes the multicast end user in an n- dimensional logical cube known as Hypercube. This cube is not used for data transmission. It is used to transfer the control information. This results in avoidance of ACK implosion problem [41]. The n-dimensional cube contains 2^n nodes. Each node is designated with a binary string 0, 1. Based on this value the nodes obtain their position in the hypercube. The binary strings are formed based on the Gray code [42] so that the codes vary by one bit value.

Two nodes are said to be neighbours only if their bit strings differ only in one bit. Each node maintains the address of its few neighbours. The multicasting tree can be built even with a cube that has less than 2n nodes. The algorithm is very simple. The node computes its parent node by just taking the two strings assigned to the node as it is varied by only one bit. Just inverting one bit is sufficient to identify the parent. The ACK implosion problem is eliminated by moving all control messages towards the parent node. . The number of group members need not always be even, hence there are incomplete hypercubes. This situation is handled by using the property of compactness.

The size of the hypercube is restricted to $\log_2 N$ using this property. A hypercube is said to be stable if all the nodes have ancestor except one. The one with the ancestor is called the Hypercube Root (HRoot). The hypercube is unstable as nodes join and leave at any time. So, in order to bring back the hypercube to its stable state this protocol uses the Duplicate Elimination (DUEL) and Address Minimization (Admin) mechanisms. Fig. 10 shows the member arrangement in cube and the tree.

Table II shows the comparison of various tree first approach protocols discussed in this section. Tree based approaches support large group of members. Each member holds information of only a small number of group members. Separate algorithms are required for loop detection and avoidance as the members joining the group chooses their parent on their own. Choice of selecting the best parent is available for the member who wants to join the group. Since the group members organize themselves as tree, the control over the data topology tree is also well defined. In tree based approaches the fan-out is controlled well. The control overhead is reduced as there is no need for communicating the control information to the entire multicast group. Despite having more advantages, tree based approaches lags in terms of balancing the tree i.e., if a node changes its parent it takes its descendants also with it. Now the descendants will have a different ancestor without their knowledge.

(a) Members in Hypercube (b) Members in tree

Fig. 10 Hypercast Protocol

TABLE II Tree First Approach Protocol Comparison

ALM Protocol	Group/Tree Management	Loop avoidance	Applications	Failure recovery	Evaluation metrics
YOID	Rendezvous Host	Coordinated loop avoidance and emergency loop avoidance algorithm	Video conferencing	-	-
HMTP	HMRP	Loop detection and resolution	-	Surviving group members	Tree cost, Link load, Delay ratio
BTP	Group members	Parent switching	File sharing program	Group members	Latency, Cost, Degree closeness
ALMI	Session controller	Bi-directional parent-child relation	Multisender multicast	Backup session controllers	Delay
Overcast	Root	-	High quality video and live streams	Hierarchical information at nodes	Bandwidth
TBCP	Root	-	-	-	Mean and maximum delay, link stress
Bayeux	Root	-	Multimedia streaming	Tapestry location mechanism	Relative Delay Penlty, Physical link stress
DT Protocol	-	-	-	-	-
Hypercast	Group members	-	-	Beacon messages and address minimization algorithm	Number of packets/bytes transmitted, Time taken for stability

V.3. Hierarchical Cluster based Multicasting

1) Kudos

In this protocol, members are arranged in a hierarchical structure. Kudos [4], [43] organizes the members in a two layer hierarchy. This approach of organizing the members hierarchically increases the scalability of the network. Clustering and mesh management are two issues that should be handled in this hierarchical arrangement of the nodes. In each layer a mesh is constructed among the members of the same level. If there are N nodes, this protocol arranges the members as N/2 in each cluster. In this protocol, members are arranged in a hierarchical structure. The solid circles are the cluster heads and all other circles are called as members of the clusters.

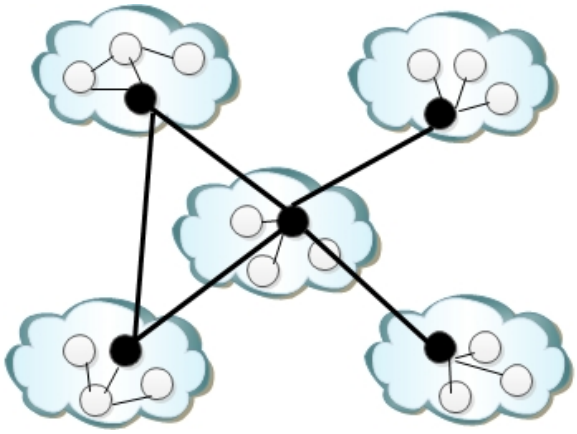


Fig. 11 Two level Hierarchy

Fig. 11 shows the two level hierarchy of the protocol where the solid circles are the cluster head of that group. Cluster head is identified which is the closest node to all other nodes in the cluster. A mesh based ALM protocol NARADA [3] is extended to form a mesh overlay. This is done at the lower layer. All the cluster heads from the lower layers are moved to the upper layer which again connects themselves as a mesh. The operations done at this level are to, add, delete, swap and partition detection and repair of the tunnels created during the overlay construction. The next process is clustering which does its work in three phases: migration, splitting and diffusion. The node who wants to join the group enters into the migration phase. The new node randomly chooses the cluster using boot strapping mechanism [44].

The head of the cluster replies with a list of all other cluster heads. The new node selects from them and computes the latency. It chooses a cluster which is having less latency and migrates from the current head to the newly found head. This is done only if the latency between the node and the new head is less than twice as that of the latency between the node and the current head. This limits the unnecessary migration. The cluster goes to the splitting phase, whenever the cluster size grows beyond N/2 where the cluster is split into two. After the split, the head of the old cluster remains the same but the head for the new cluster is chosen based on the latency information provided by the old cluster head. The cluster enters into the diffusion stage, whenever the number of nodes in a cluster goes below N/2, due to some failure or as a result of a node leaving the cluster. The nodes in such clusters move to the neighboursing cluster. The children nodes from different clusters cannot form an overlay among them. This reduces the efficiency even though it achieves scalability due to the hierarchical arrangement.

2) NICE

The NICE [4], [45] protocol uses hierarchical clustering concept. The members at the lowest level are clustered. Each cluster is of size between k and 3k− 1, where k is a constant. This constant k determines the set of members that are close to each other. From those clusters one cluster head is chosen based on the center of the cluster and taken to the next level. Like this all the clusters in the lowest level will send their cluster heads to the next level. The choice of the cluster guarantees that a new member is able to quickly find its appropriate position. The members at the bottom of the hierarchy maintain (soft) state about a constant number of other members, while the members at the top maintain such state for about O(log N) nodes.

Fig. 12 shows the hierarchical arrangement of the members. The member hierarchy is used to define both the control and data overlay topologies. In the control topology, all members of each cluster peer with each other and exchange periodic refreshes between them. The source member sends a packet to all its peers on the control topology which is used to determine the data topology. The new member joins the lowest layer cluster that is closest to itself with respect to the distance metric but this is done by a series of refinements from the topmost layer till it reaches the bottom layer.

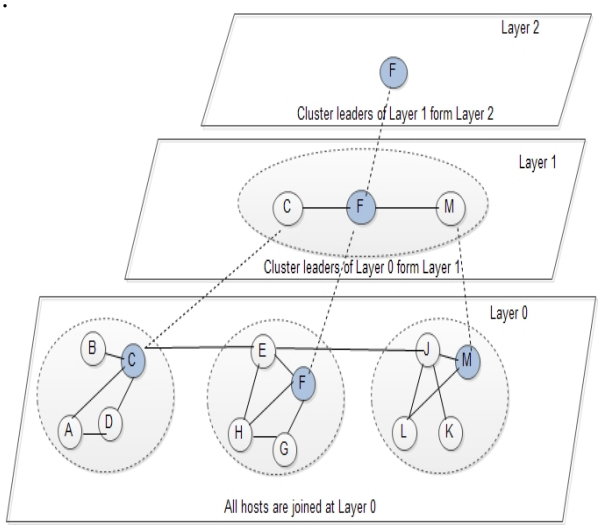


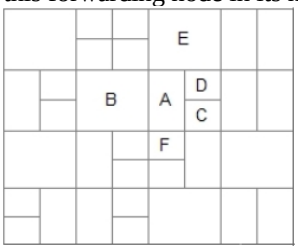
Fig. 12 NICE protocol

The new node first contacts the RP, the RP will then send the information of the cluster heads of the immediate lower layer. Then the node contacts those clusters and so on until it is mapped to a lower layer L_0. The delay incurred in the joining process is compensated by providing the new node

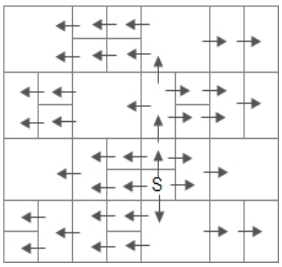
with the data by the cluster of the current layer. This results a situation where the node need not wait without data until it joins its correct cluster. Cluster management is done by split and merge process in order to avoid any violation in the size of the cluster. This protocol supports low bandwidth data stream applications.

3) CAN

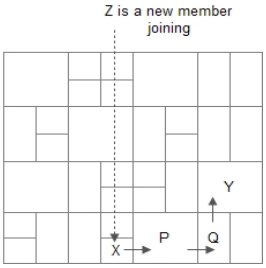
Content Addressable Network (CAN) [27], [46] is an application level infrastructure where a set of end hosts implement a distributed hash table on an Internet wide scale. The members of the CAN form a virtual d dimensional Cartesian coordinate space. Each member obtains a particular portion from this coordinate space. In the control topology, two members are peer with each other only if their corresponding regions in the dimensional space lie adjacent to each other. The data topology is implicitly defined by performing directed flooding on the control topology. The node who wants to send the data is termed as source and it forwards the data to all its neighboursing nodes. The node receiving the packet will forward the packet only to the nodes which will have this forwarding node in its neighbours list.



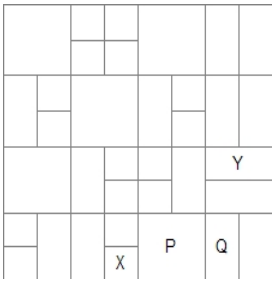
Panel 0



Panel 1



Panel 2



Panel 3

Fig. 13 CAN

Panel 0 in Fig. 13 is a two dimensional coordinate space partitioned into 34 zones by 34 members participating in the multicast group. A-F are the members marked with their zones in the coordinate space. This shows the control topology in which member A has 5 neighbours. Panel 1 in Fig. 13 shows the flooding happening from the source to all members in the control topology. The nodes in the control topology forward the message. The forwarding will continue only if the packet has not travelled half of the coordinate space. This ensures that the packet does not get into the loop. Each member has a cache to identify whether the packet is a duplicate or not. Panel 3 in Fig. 13 shows a new member, Z wants to join the CAN. It queries the RP to find at least one existing member, X , that has already joined CAN. Z picks a random point in the coordinate space. The goal of the joining member is to find the member Y which owns this randomly chosen point is done. This is done by routing through the CAN. The protocol then splits the zone owned by Y into two, and the ownership of one of the halves is transferred to Z as shown in panel 3 of Fig. 13.

The assignment procedure of zones of the coordinate space to members of the CAN ignores the relative distances between the members in constructing the overlay. As a consequence, neighbours on the CAN may be far apart and thus, the multicast overlay paths can have high stretch. To remedy this situation, "distributed binning "scheme, where members that are close to each other are assigned nearby zones in the coordinate space, is used.

4) Scribe

Scribe [27], [47] is a large scale event notification system that uses application layer multicast to disseminate data on topic- based publish subscribe groups. Scribe is built on top of Pastry which is a peer-to-peer object location and routing substrate overlaid on the Internet. The control topology is similar to that of the Pastry's control topology. Each member in Pastry is assigned a random node identifier, which may be generated by computing cryptographic hash like SHA-1[39] of the member's public key. Pastry organizes the members into an overlay in which messages can be routed from a member to any other member by knowing the node identifier of the latter. The members are represented by rectangular boxes. The corresponding node identifiers are marked inside the box. The node identifiers are represented as a sequence of digits.

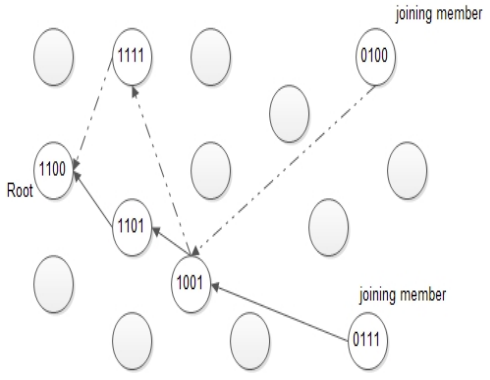


Fig. 14 Scribe

Fig. 14 shows the joining mechanism in Scribe. A routing table is maintained by all the members. The routing table provides information of a set of members with common prefixes in the overlay. As in the Fig. b=1 i.e., the prefix match is done one bit at a time. Assume the group identifier is 1100. If a new member with node identifier as 0111 wants to join the group, Pastry routes the join message to 1001 and then to 1101 until it reaches the root 1100. If a node 0100 wants to join as shown in dot-dash line pattern, 1001 adds the new node to its children list.

The routing is performed by each member forwarding the message to a member available in the routing table which will have the closest prefix. When no such member with the closest prefix is available, the message is forwarded to a member in the leaf set which is closer to the destination identifier than its own identifier. Each multicast group has its own identifier. The member whose node identifier is numerically closest to the multicast group identifier becomes the RP for that group.

The data topology for a multicast group in Scribe is the union of the Pastry unicast paths from the different group members to the RP. A member can join the multicast group if it has proper credentials which are provided to a node by the Pastry, if it has not violated any of the security issues. A member joining the multicast group sends a join message using the multicast group identifier as the destination identifier which is routed by the Pastry substrate to the RP. The member to join the Scribe multicast group should be joined in the Pastry group.

If the node leaves the group and if it does not have children, then it sends a leave message to its parent and hence the message travels upwards till it reaches root. Scribe handles the failure of RP (root) as it is replicated in k closest nodes to the RP (root). If there is a node failure in the multicast tree then, Scribe delivers the packets out of order to its members. This is handled by a simple mechanism of invoking some handlers which will take care of the forwarding of the message, the sequence number of the messages and the details of the joins.

5) SHMHD

An ALM protocol called Scalable Hierarchical Multicast for High Definition streaming media (SHMHD) [48] is proposed by YouWei Zhang et al. This protocol efficiently utilizes the end to end bandwidth to achieve HD media transmission. Based on the available bandwidth the hosts transmit a basic unit of data along the data delivery path. This protocol works in many-to- many transmission styles. Multiple parents collaborate for a data and send to the requested end host. RP maintains the information of membership. The data piece is split into smaller blocks. A recipient can request particular data blocks from several parents simultaneously in line with the available bandwidth between parents and itself. The hosts are arranged in a logical hierarchical layer. Fig. 15a shows the hierarchical arrangement of hosts. Host A and B receives information from S, C selects A as its parent. Information for C is received from host A. But, in another case as shown in fig. 15b C selects A and B as its parents and the information is received from both the parents. HD streaming demand can be either met by A or B, but it is also possible to make both the hosts jointly achieve the demand. In order to do it jointly the hosts sends consecutive blocks separately. There is also another chance for host C to select its parents. It can also choose S as its parent as shown in Fig. 15c.

HD streaming media based on its demand cannot be achieved in layer1, as C receives information from S. Once the hosts find their position in the structure they cannot move to any other layer. The hosts can receive and send data for upto k parents and children respectively. When a node wants to join the group contacts the RP provides information of some peers at the lower layers. Then the host measures the approximate bandwidth which is mapped with the number of blocks as shown in equation 1.

$$NUM\_BK = Trunc\left(\frac{Avail\_BW}{Block\_size}\right) \quad (1)$$

Avail_BW is the available bandwidth and Block_size is calculated as in equation 2.

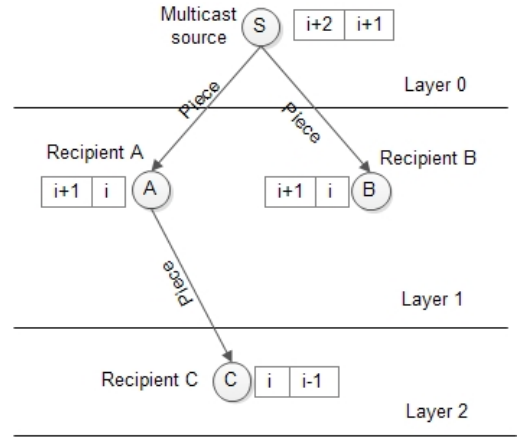
$$Block\_size = 2^{(trunc(\log_2 piece-length/k)+1)} \quad (2)$$

The procedure is repeated until the new host finds the proper peer is located. The new member then joins the tree and starts requesting for the data blocks. The value obtained in equation 1 is sorted in descending order by each host. The Num_BKs are checked with the k value, if it exceeds then computes upto k– 1 Num_BKs. From this the minimum set qualified parents are obtained and then node who intended to join the group attaches itself with its parent. After joining the tree the hosts' requests for the data by specifying the block number, start block and end block of it. The blocks transmitted by the parents are in proportion to the available bandwidth. This is to ensure that the parent is load balanced. When a node leaves the group gracefully in order to avoid the performance degradation, back up parents are employed. Keep Alive messages are used to detect the disgraceful departure of the nodes. The performance is compared with the Coopnet protocol [49] based on RDP, link stress and the Quality of Service (QoS).

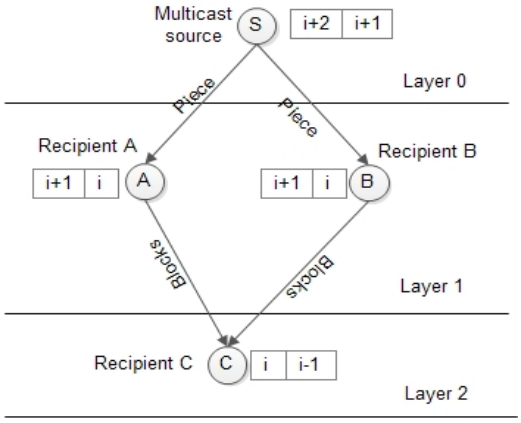
6) SDDM

Jing Li et al proposes a Short Delay Degree-constrained hierarchical Multicast protocol SDDM [50] for application layer multicast. The tree construction is based on the Fibonacci series. Bandwidth capacity heterogeneity is considered for the construction of the hierarchical structure.

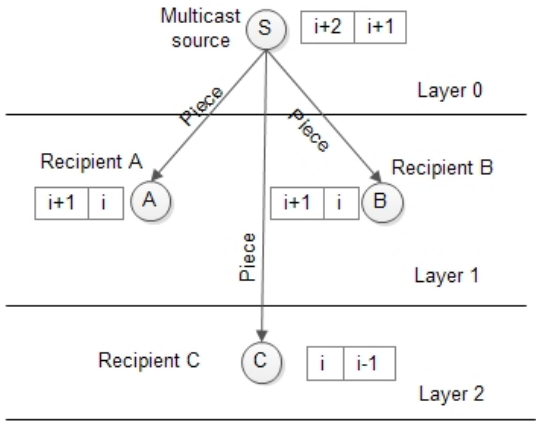
SDDM uses a concept of local areas which means all end hosts connect to a same router directly or through some network components. Each local area consists of k to 3k– 1 end hosts similar to NICE [45]. Members in each local area are divided into layers, the leaders of the clusters are moved to next higher cluster until a single node layer is formed which is the root of the hierarchical structure.



(a)



(b)



(c)

Fig. 15 hierarchy division and data delivery in SHMHD

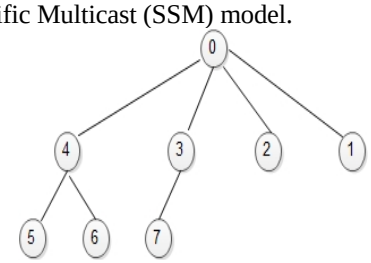
Members in each cluster are connected by a shared tree for data delivery constructed by the Fibonacci series. Each member m_i has to maintain d_i, the delay of the link from itself to cluster head, deg_i, degree of the member and li, the packet processing delay of the member. Based on these parameters, weight of a member is computed as shown in equation 3. α and β are balance factors. The parameters n_{max}, deg_{max} and l_{max} are the maximum value that the parameters n_i, deg_i and l_i takes respectively. The equation 3 calculates weight and is considered for members of the cluster being in the same local area.

Equation 4 calculates if the members of the cluster are not from the same local area. Based on these weights the members are sorted in the ascending order and sequenced.

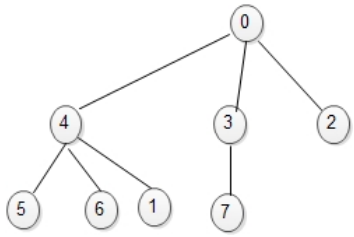
After the sequencing of members, tree construction is done using the Fibonacci series. Fig. 16a shows the tree constructed using Fibonacci series. The tree constructed using just based on the Fibonacci series does not have the degree constrained property.

If the node exceeds the out-degree, it leads to under utilization of some of the links of that particular node. So another algorithm which builds a multicast tree based on the Fibonacci series with the degree constrained property as shown in Fig. 16b. Here the out-degree is fixed as 3. Each member sends a refresh message to maintain the cluster. A member v who wants to join the group contacts a RP which is close to it. The RP then checks its local core list and selects a local core which will be close to v and informs about the new node. Then it is the cluster head's responsibility to find a suitable position for the member in the tree. When a member wants to leave the group sends a remove message to the leader. Accordingly some modifications are done to the parent and children of the leaving node if there exist. SDDM is compared with the NICE [45] and OMNI [51] based on metrics like delay and overhead. It out performs both the protocols.

Fig. 17 shows the join process in this Bincast protocol. Each node has its score computed using equation 5. As an example, if a node interested in joining the group has to compute its score (eg: - 3100). The node then sends a join message towards the source. The source sends a build message through a midway header. This header directs the build message to a bin whose score is close to the new node's score (eg:-3200). Similarly the build message is propagated downwards until the new node gets its appropriate parent. This process creates a k-ary tree where k is the number of bins in each level. A secondary k-ary tree is constructed among the source, local headers and midway headers using Source Specific Multicast (SSM) model.



(a) Multicast tree construction using Fibonacci series



(b) Modified multicast tree

Fig. 16 SDDM multicast tree construction methods

$$w_i = \alpha \frac{n_i}{n_{\max}} + \beta\left(1 - \frac{\deg_i}{\deg_{\max}}\right) + (1-\alpha-\beta)\frac{l_i}{l_{\max}} \quad (3)$$

$$w_i = \alpha\left(1 - \frac{n_i}{n_{\max}}\right) + \beta\left(\frac{\deg_i}{\deg_{\max}}\right) + \gamma\left(1 - \frac{\deg_i}{\deg_{\max}}\right) + (1-\alpha-\beta-\gamma)\frac{l_i}{l_{\max}} \quad (4)$$

7) Bincast

Reza Besharati et al proposes a novel, stable and low overhead ALM approach using binning technique. Clustering is done by grouping nearby receivers [52]. Bincast uses a constant number of landmarks, which are special nodes like DNS servers across the Internet so that the user can find its cluster. It then constructs a k-ary tree between cluster members. The most stable node based on the fan-out and age [53] is selected as the local header in each bin. It monitors cluster membership events and decides to split or merge the cluster if necessary. The local header is the root of the k-ary tree.

This kind of two tier construction improves the stability criteria. It is worth noting that the described scoring function is flexible and interchangeable.

When a new member m_{new} is interested in joining the multicast session, pings predefined number of landmarks and computes score S_m score is computed as shown in eqn. 5.

$$S_m = \sum_i D_{li} \times 3^{i-1} \quad (5)$$

where, D_{li} is the RTT to the landmark i. It first finds a bin that its range contains S_m. Then it sends the score by means of a join message toward the source for which the source responds with the parent message accepting it as a direct child. If the source is not able to accept then it sends a build message which is empty path list, to the next header in the bin. The message is filled by IP addresses of the nodes whoever receives it. Once if the new node can be occupied with the score S_m mentioned by it, the new node receives the build message with the updated path list. Hence, the node joins the group. If the bin is empty, m_{new} is the first member of the bin and is selected as its local header. If the number of landmarks is l, the order of join process with N being number of end hosts and the k is the number of bins, then the order of join process is given as, $l + \log_k N$. The internal nodes also maintain a timer in order to keep track of the time to make the new node as a stable node. The resulted tree is a k-ary tree. Each local header must keep a list of bin members with their scores and arrival times. Since the bin size is limited by a threshold, this list is also small. Each non-header member must maintain a list of its children as well as a list of overlay nodes on the path from the source to itself. If the bin size exceeds the threshold, then splitting is done.

Bin is split into k new bins, where, each bin consists of 1/k nodes. It is the responsibility of the header to inform about the split to all the members, So that the nodes can join their bin using join message with the scores. The internal nodes must periodically update its children with a Hello message. Once the failure is detected based on the join message sent by the detector the header chooses a stable node to percolate upward in the tree. If the node leaves gracefully, then it sends a remove message to all its children who will then contact the closest header. The performance metrics used for the evaluation are, total data delivery of all receivers, total stress of all links and control overhead. Bincast is compared with the NICE [45] based on these evaluation parameters and the result shows that its performance is better than NICE.

Table III shows the comparison done among the hierarchical application layer multicast protocols. The performance of hierarchical multicasting is well appreciated on basis of scalability issue. These types of ALM protocols support few thousands of end users. As the end users participate in multicasting the network becomes dynamic. This dynamics are well handled by the hierarchical overlays when compared with the tree and mesh overlays. Heterogeneous networks are supported by this kind of architecture.

The clustering helps in reducing routing of the information In the multicast tree as the entire group is split into clusters of small number of nodes, the information maintained in the node is minimal. The overlay converges very fast.

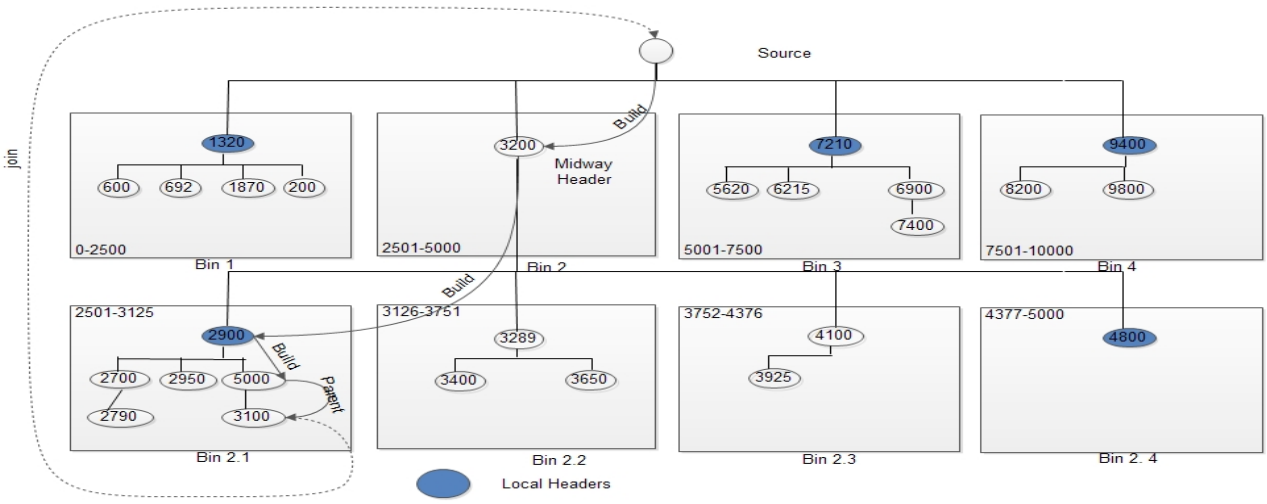


Fig 17. Bincast

TABLE III Hierarchical Alm Protocol Comparison

ALM Protocol	Group/Tree Management	Applications	Failure recovery mechanism	Evaluation metrics
Kudos	Randomly chosen member	Broadcasting	Refresh messages	RDP, QDP
NICE	RP	Low bandwidth data stream applications	-	Stress, Stretch, Latency
CAN	RP	-	-	-
Scribe	Pastry substrate [54]	Instant messaging	Paxos [55]	Delay, Stress, Stretch
SHMHD	RP	-	Backup parents	RDP, Link stress, QoS
SDDM	Set of RP	-	-	Delay, Overhead
Bincast	Landmarks	-	Hello messages and Bin leaders	Total data delivery for all receivers, Total stress, Control overhead

VI. CONCLUSION

Application layer multicast, with the advantage of easy deployment capability, is a new approach to provide multicast services to group applications. In this end-system architecture, end-hosts organize themselves into overlay network which take care of multicast functionalities. This overlay network has the capability to handle the group dynamics and address the scalability issues. In this article, various ALM protocols are described and analyzed based on their characteristics. Depending upon the application a specific protocol can be chosen. Though ALM has less performance efficiency when compared with IP multicasting, newer ALM approaches are being developed to facilitate emerging trends in Internet usage.

The Application Layer Multicast provides a practical solution for multimedia communication among a group of members. Qualities of Service parameters are also considered while designing ALM protocols. The tree and mesh topology follow a flat architecture rather than following a hierarchical architecture. The topology is formed among the group members in the same logical level. But whereas in hierarchical design, the group members form a topology, in which the members participating the multicast functionalities are at different logical level. The hierarchical scheme provides a good scalability. The overhead is also low as the group members' information is localized, so that a node needs to maintain only small amount of information. But additional overhead is incurred in maintaining the clusters. Mesh structures can be used when the group size is small because the control overhead exchanged between the nodes is limited which will not add much traffic to the network.

The Internet trend sees more mobile nodes into the multicast sessions. The mobility of the nodes cannot be restricted leading to a problem of Line of Sight (LoS) which causes interruption amidst of sending and receiving the video streams. This interruption is not appreciable in multimedia streaming applications. Another issue to be considered is to provide good quality video stream over rural areas or under privileged areas. As the video streaming requires some infrastructure for a quality video to be transmitted, building an infrastructure is not possible in these areas as it incurs longer time and also more cost. Wireless Mesh Networks (WMN) comes as a practical solution whose deployment is very easy, fast and the cost is also very less. WMNs are self organized and self healing networks [56]. WMNs consist of mesh routers and mesh clients. Multicasting video streams with QoS guarantee is an open issue to be solved. A cross layer based ALM protocol is to be designed, as the mobile nodes have limited battery power. This makes ALM protocols to be collaborated with the physical network which is the mesh routers. A cross layering approach may enhance the process of multicasting multimedia data over WMNs. Developing an efficient ALM protocol in general and developing such a protocol for wireless networks in specific still remains as a open and challenging research problem.

REFERENCES

[1] S. Deering. Host extensions for ip multicasting. RFC 1112, 1989.

[2] Suman Banerjee, Bobby Bhattacharjee, and Christopher Kommareddy. *Scalable application layer multicast*, Proceedings of SIGCOMM 2002, pp. 205-217.

[3] Yang hua Chu, Sanjay G. Rao, Srinivasan Seshan, and Hui Zhang. A case for end system multicast. Proceedings of the 2000 ACM SIGMETRICS international conference on Measurement and modeling of computer systems, pp. 1-12, volume 28, 2000.

[4] Milena JANIC. *Multicast in Network and Application Layer*. PhD thesis, October 2005.

[5] F. Solensky, *IPv4 Address Lifetime Expectations* Adddison Wesley, 1996.

[6] James F. Kurose and Keith W. Ross. *Computer Networking: A Top-Down Approach Featuring the Internet* Adddison Wesley, 1999.

[7] R. Gilligan and E. Nordmark. Transition mechanisms for ipv6 hosts and routers. RFC 1993, 1996.

[8] L. Garber. Steve deering on ip next generation. *IEEE Computer*,vol. 32, n. 4, pp. 11-13, 1999.

[9] Kostas Katrinis and Martin May. *P2P Systems and Applications*, vol. 3485. Springer- Verlag Berlin Heidelberg-LNCS, pp. 157-170, 2005.

[10] C.K. Yeoa, B.S. Leea, and M.H. Erb. A survey of application level multicast techniques. *Elsevier Computer Communications*, vol. 27, n . 15, pp. 1547-1568, 2004.

[11] D. Pendarakis, S. Shi, D. Verma, and M. Waldvogel. *Almi: An application level multicast infrastructure, In* Proc. of 3rd Usenix Symposiumon Internet Technologies and Systems (USITS), pp. 153-162, 2001.

[12] Beichuan Zhang, S. Jamin, and L. Zhang. Host multicast: A framework for delivering multicast to end users. Computer Networks: The *International Journal of Computer and Telecommunications Networking - Overlay distribution structures and their applications*, vol. 50, n. 6, pp. 781-806, 2006.

[13] Mojtaba Hosseini, Dewan Tanvir Ahmed, Shervin Shirmohammadi, and Nicolas D. Georganas. A survey of application-layer multicast protocols. *IEEE communications Surveys*, vol. 9, n. 3, pp. 58-74, 2007.

[14] D.Waitzman, C. Partridge, and S. Deering. Distance Vector Multicast Routing Protocol. RFC 1075, 1998.

[15] D. Kostic, A. Rodriguez, J. Albrecht, and A. Vahdat. *Bullet: High bandwidth data dissemination using an overlay mesh*, Proceedings

of 20th ACM Symp. on Operating Sys. Principles, pp. 282-297, 2003.
[16] F.B. Schneider. Byzantine generals in action: Implementing fail-stop processors. *ACM transactions on Computer Systems*, vol. 2, n. 2, pp. 145-154, 1984.
[17] Yatin Chawathe. Scattercast: an adaptable broadcast distribution framework. *Multimedia Systems*, vol. 9, n. 3, pp. 104-118, 2003.
[18] Lougheed .K And Rekhter .Y. A border gateway protocol (bgp). RFC 1247, 1989.
[19] Yatin Chawathe, Steven McCanne, and Eric A. Brewer. *Rmx: Reliable multicast for heterogeneous networks*. Proceedings of the 19th Joint Conf. of the IEEE Computer and Communica- tions Societies, 2000.
[20] Sally Floyd, Van Jacobson, C. Liu, Steven McCanne, and L. Zhang. *A reliable multicast framework for light-weight sessions and application level framing, Proceedings of* ACM SIGCOMM 95, pp. 342-356, 1995.
[21] Dejan Kostic, Adolfo Rodriguez, Jeannie Albrecht, Abhijeet Bhirud, and Amin Vahdat. *Using random subsets to build scalable network services*. In Proceedings of USENIX Symposium on Internet Technologies and Systems, 2003.
[22] Sally Floyd, Mark Handley, Jitendra Padhye, and Jorg Widmer. *Equation-based congestion control for unicast applications*. Proceedings of SIGCOMM 2000, pp. 43-56, 2000.
[23] S. Nari, H. R. Rabiee, A. Abedi, and M. Ghanbari. *An efficient algorithm for overlay multicast routing in videoconferencing applications.* In Proceedings of 18th International Conference on Computer Communications and Networks, pp. 1-6, 2009.
[24] Mojtaba-Hosseini. *Design of a multi-sender 3d videoconferencing application over an end system multicast protocol*. pp. 480-489, 2003.
[25] Mojtaba Hosseini. End system multicast routing for multi-party videoconferencing applications. *Computer Communications*, vol. 29, n. 11, pp. 2046-2065, 2006.
[26] Tina Wong, Thomas Henderson, Suchitra Raman, Adam Costello, and Randy Katz. *Policy-based tunable reliable multicast for periodic information dissemination*, In Proceedings of WOSBIS,, 1998.
[27] Suman Banerjee and Bobby Bhattacharjee. A comparative study of application layer multicast protocols, 2001.
[28] P. Francis. Yoid: Extending the multicast internet architecture. White Paper, 1999.
[29] David A. Helder, Sugih Jamin, Banana Tree Protocol, an End-host Multicast Protocol, 2000.
[30] Jungle Monkey homepage. http://www.junglemonkey.net, 2000.
[31] Sugih Jamin, Cheng Jin, Yixin Jin, Dan Raz, Yuval Shavitt, and Lixia Zhang. *on the placement of internet instrumentation*. In Proc. of IEEE INFOCOM, 2000.
[32] J. Jannotti, D. Giford, K. Johnson, M. Kaashoek, and J. OToole. *Overcast:reliable multicasting with an overlay network*, In Proceeding of OSDI'00 the 4th conference on Symposium on Operating System Design & Implementation, Vol. 4, pp. 14 – 14, 2000.
[33] L. Mathy, R. Canonico, and D. Hutchinson. *An overlay tree building control protocol*. In Proc. of 3.rd International COST264 Workshop on Networked Group Communication (NGC01), pp. 78-87, 2001.
[34] Jorg Liebeherr, Michael Nahas, and Weisheng Si. Application-layer multicasting with delaunay triangulation overlays. *IEEE JSAC*, vol. 20, n. 8, pp. 1472-1488, 2002.
[35] R. Sibson. Locally equiangular triangulations. *Comput. J.*, vol. 21 n. 3, pp. 243-245, 1977.
[36] S. Q. Zhuang, B. Y. Zha, and A. D. Joseph. *Bayeux: An architecture for scalable and fault-tolerant wide-area data dissemination*. In Proceedings of 11th ACM/IEEE NOSSDAV01, pp. 11-20, 2001.
[37] Zhao B. Y., Kubiatowicz J. D., And Joseph A. D. Tapestry: An infrastructure for fault-tolerant wide-area location and routing. Technical report, Tech. Rep. UCB/CSD-01-1141, University of California at Berkeley, Computer Science Division, 2001.
[38] Kubiatowicz J. *Oceanstore: An architecture for global-scale persistent storage*. In Proceedings of ASPLOS, 2000.
[39] Robshaw M. J. B. Md2, md4, md5, sha and other hash functions. Tech. Rep. TR-101 4, RSA Labs, 1995.
[40] Jrg Liebeherr and Tyler K. Beam. *Hypercast: A protocol for maintaining multicast group members in a logical hypercube topology*. In Networked Group Communication, pp. 72-89, 1999.
[41] J. Crowcroft and K. Paliwoda. *A multicast transport protocol*. In Proc. ACM Sigcomm '88, pp. 247-256, 1988.
[42] M.J. Quinn. Parallel Computing: Theory and Practice. (McGraw-Hill, 1994).
[43] Sushant Jain, Ratul Mahajan, David Wetherall, and Gaetano Borriello. Scalable self organizing overlays. Technical report, 2002.
[44] C. Cramer, K. Kutzner, and T. Fuhrmann. Bootstrapping locality-aware p2p networks, In proceedings of the IEEE International Conference on Networks (ICON), pp. 357-361, 2004.
[45] Suman Banerjee, Bobby Bhattacharjee, and Christopher Kommareddy. *Scalable application layer multicast*, In proceedings of conference on Applications, technologies, architectures, and protocols for computer communications 2002.
[46] Sylvia Ratnasamy, Mark Handley, Richard Karp , and Scott Shenker, *Application-Level Multicast using Content-Addressable Networks*, In Proceeding of NGC '01 the Third International COST264 Workshop on Networked Group Communication, pp. 14-29, 2001.
[47] Miguel Castro, Peter Druschel, Anne-Marie Kermarrec, and Antony Rowstron. Scribe: A large-scale and decentralized application-level multicast infrastructure. *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 20,n. 8, 2002.
[48] You-Wei Zhang, Chan-Le Wu, and You-Wei Zhang. *A scalable multicast scheme for high definition streaming media*. In proceedings of 2009, International Conference on Multimedia Information Networking and Security, pp. 291 - 295, 2009.
[49] Venkata N. Padmanabhan, Helen J. Wang, and Philip A. Chou. Distributing streaming media content using cooperative networking. Technical Report MSR-TR-2002-37, Carnegie Mellon University, 2002.
[50] Jing Li, Yong Wang, Mei Xue, and Zhong Tang. *A short delay degree-constrained hierarchical application layer multicast protocol*, In proceedings of 2009 International Conference on Web Information Systems and Mining, pp. 674 - 681, 2009.
[51] Suman Banerjee, Christopher Kommareddy, Koushik Kar, Bobby Bhattacharjee, Samir Khuller. OMNI: An Efficient Overlay Multicast Infrastructure for Real-time Applications,2005.
[52] Reza Besharati, Mozafar Bag-Mohammadi, and Mashallah Abbassi Dezfouli. *A topology-aware application layer multicast protocol*. In 7th IEEE Consumer Communications and Networking Conference (CCNC), pp. 1 - 5, 2010.
[53] D. Stutzbach and R. Rejaie. *Understanding churn in peer-to-peer networks*. In Proc of ACM IMC, pp. 189-202, 2006.
[54] Antony Rowstron and Peter Druschel. "*pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems*". In Proc. IFIP/ACM Middleware 2001, 2001.
[55] L. Lamport. The part-time parliament. Technical report, Digital Equipment Corporation Systems Research Center and Palo Alto and CA,, 1989.
[56] Ian F. Akyildiz, Xudong Wang, and Weilin Wang, "Wireless mesh networks: a survey". *Computer Networks and ISDN Systems* ,vol. 47, pp. 445-487, 2005.

AUTHORS' INFORMATION

M. Anitha received B.E in electronics and communication engineering from Madurai Kamaraj University, Madurai, Tamilnadu, India in 1998 and ME in multimedia technology from Anna University, Chennai, Tamilnadu, India. She is currently working as a visiting faculty in computer science and engineering, College of Engineering, Anna University, Chennai, India. Her research interests include wireless mesh networks, multimedia networks particle swarm optimization.

P. Yogesh received his B.E and M.E in computer science and engineering from Madurai Kamaraj University, Madurai, Tamilnadu, India in 1988 and 1996 respectively. He received his PhD degree in the faculty of Information and Communication Engineering in 2007 from Anna University, Chennai, Tamilnadu, India. He has published around 20 articles in various international journals. He has also published around 20 articles in various international conferences. His research areas include infrastructureless wireless networks, multimedia communication and network security.Dr. Yogesh, Associate Professor, Palanichamy. He is a life member of Indian Society of Technical Education and life member of Computer Society of India.