Multicast routing

Draft

Lucia Tudose
Nokia Research Center
E-mail: tudose@research.nokia.com

Abstract

Multicast routing is establishing a tree which is routed from the source node and contains all the multicast destinations.

Based on the way the state information is maintained, the existing multicast routing algorithms can be divided in three classes: source routing, distributed routing and hierarchical routing algorithms.

A distributed multicast routing protocol for real-time multicast applications is most thoroughly investigated. A delay bounded routing tree is a tree in which the accumulated delay from the source node to any destination along the tree does not exceed a pre-specified bound.

1. Introduction

Multicast is a kind of group of communications, which requires simultaneous transmission of messages from a source to a group of destinations. Real–time multicast refers to a multicast in which messages should be received by all destinations within a specified time delay. There are many applications relying on real-time multicast services. For example, in a video conferincing system, each participant interacts with the others by sending and receiving messages (which can be texts, pictures, audio sequences, or video frames). Each message, originating from a participant, must be delivered to all the others in a real-time manner.

In real-time communications, a connection (logical connection) from the source to destination(s) has to be established before any data transmission occurs. During the connection set-up, sufficient network resources (i.e. network bandwidth, buffer, etc.) are reserved at each network node on the connection , so that user required QOS (Quality of Services) can be guaranteed at run time.

Routing is an important step of connection set-up. Multicast routing is to find a routing tree and contains all destinations. One principal goal of multicast routing is to minimize the network cost. The network cost of a routing tree is defined as the total distance of all links in the tree, which is the total distance a multicast message should travel to reach all destinations. A tree with minimal overall cost is called a Steiner tree. Finding a such a tree in a network is a NP-complete problem. Many real-time applications also have a time constraint which requires the communication to be done within pre-specified delay bound. Finding such a tree which have a minimal network cost under a delay bound is called the delay bounded Steiner tree problem [2].

2. Routing Strategies

There are three routing strategies: source routing, distributed routing and hierarchical routing. They are classified according to how the state information is maintained and how the search of feasible path is carried out.

In source routing , each node maintains the complete global state, including the network topology and state information of every link. Based on the global state, a feasible path is locally computed at the source node. A control message is then sent out along the selected path to inform the intermediate nodes of their precedence and succesive nodes. A link-state protocol is used to update the global state at every node.

Drawbacks:

-high communication overhead for large-scale networks. The global state maintained at every node has to be update frequently enough to cope with the dynamics of network parameters, such as bandwidth and delay.

-approximate global state due to the overhead concern and nonnegligible propagation delay of state messages.

-high computation overhead at source especially in case of multicast when multiple constraints are involved.

In distributed routing the path is computed by distributed computation. Control messages are exchanged among the nodes, and the state information kept at each node is collectively used for the path search. Most distributed routing algorithms need a distance vector protocol or a link-state protocol to maintain a global state in a form of distance vectors at every node. The routing is done on a hop-by-hop basis.

Drawbacks: heavy communication cost, connection set-up time, and poor quality of the produced routing trees.

In hierarchical routing, nodes are clustered into groups, which are further clustered into higher-level groups recursively, creating a multilevel hierarchy. Each physical node maintains an aggregated global state. This state contains detailed state information about the nodes in the same group and aggregate state information about other groups. Source routing is used to find a feasible path on which some nodes are logical nodes representing groups. A control message is then sent along this path to establish the connection. When the border node of a group represented by a logical node receives the message, it uses the source routing to expend the path through the group.

3. **A distributed multicast routing protocol for real-time multicast applications**.

The method generates routing tree sequentially, that is, it extends the routing tree to one destination after another.

The protocol has the following advantages:

- Fully distributed. Each node operates based on its local routing information and the co-ordination with other nodes is done via network message passing.
- Near optimal trees under the delay bound. The network cost of the tree is close to optimal under the condition that the delay from a source to any destination along the routing does not exceed a pre-specified bound.
- Low communication cost and flexible in dynamic membership changes. It takes a small number of network messages for routing and the multicast membership can be dynamically changed without affecting the existing traffic on the connection.

The network is modelled by a connected graph G(V,E), where the nodes in the graph represent communication end-points and the edges represent links. For each link $e \in E$, d(e) is the distance of the link. It can be numbers of hops in wide area networks. It is assumed that the network delay is proportional to the distance to the link.

The network cost of a routing tree T is defined as:

$$\text{NetworkCost}(\text{T}) = \sum_{e \in T} d(e)$$

One objective of multicast routing is to optimize this network cost, so that the bandwidth consumption for each multicast will be minimal. In addition many interactive multimedia applications have a strigent delay constraint.

If D is a set of destination, $s \in V$ is the source node, and P(s,u) is the path from s to u along the routing tree T then the Bounded Delay requirement for each $u \in D$ is :

$$\sum_{e \in P(s,u)} d(e) \leq \Delta$$

The idea of the algorithm in constructing routing trees mimics Prim's MST [5] algorithm and is in combination with a distributed shortest-path algorithm. It is assumed that each node has information about the shortest path to every other node in the graph. This can be achieved by running the Bellman-Ford's distance-vector. Also, it is assumed there are no long live loops in shortest paths.

The construction of a routing tree starts with a tree containing only the source s. A destination in D is selected, which is the closest to the tree and for which the end-to-end delay from s to it along the tree is under the bound. The shortest path from the tree to this selected destination is added to the tree. By adding a path to the tree, all nodes on the path are included to the tree. At each step, an unused destination, which is the closest to the tree under the delay condition, is added to the tree. This operation repeats, until all destinations are in the tree.

Next follows an example to show how the protocol works and a delay-bounded tree is constructed in a distributed manner.

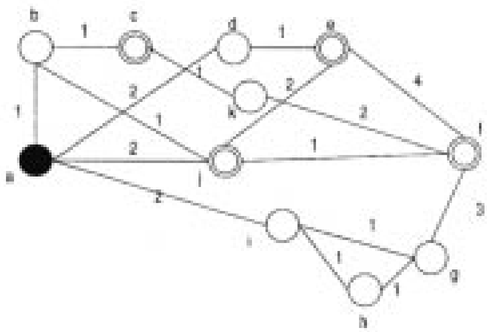


Fig. 2. A simple network graph.

Fig. 2 is a network graph, *a* node is the multicast source and destinations D={c, e, f, j}, Δ=3. The construction of the routing starts from the source *a*. Destination *c* is chosen to be first linked into the tree. After *c* is included in the tree, the tree contains path <abc>. Node *j* is included into the tree by path <bj>followed by the destination *f* in the same way. The routing constructed so far is shown in fig. 3.

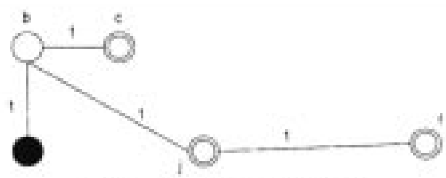


Fig. 3. A routing tree constructed so far.

When considering destination *e*, although is close to the tree node j, the end-to-end delay from *a* to it via *j* is 4>Δ. The shortest path from the tree to *e* under the delay bound is the shortest path from source a to it. Thus destination *e* is linked to the tree by path <ade>.

Fig. 4 is the final routing tree, which contains all destinations.

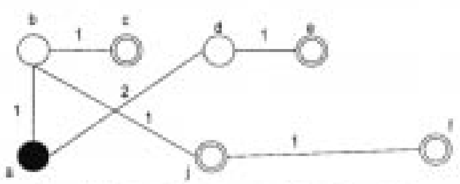


Fig. 4. The final routing tree for $D = \{c, e, f, j\}$.

4. Multicast routing protocols in practice

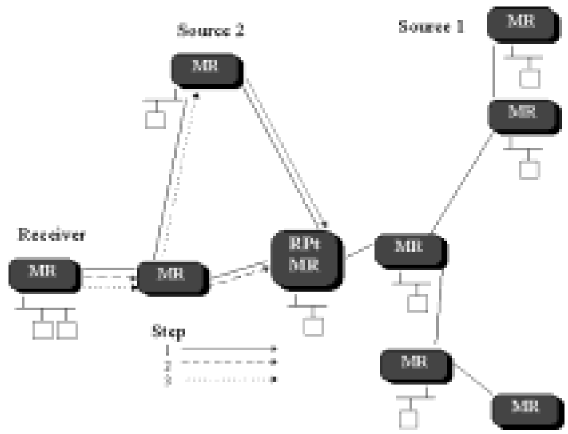
IP Multicat routing protocols generally follow two basic approaches, depending on the distribution of multicast group members throughtout network. The first approach is based on the assumption that the multicast group members are densely distributed throughout the network and bandwidth is plentiful, i.e. almost hosts on the network belong to the group. So-called "dense-mode" multicast routing protocols rely on specific flooding of the network with multicast traffic to set up and maintain the spanning tree. Dense-mode routing protocols include Distance Vector Multicast Routing Protocol DVMRP, Multicast Open Shortest Path First MOSPF, and Protocol-Independent Multicast –Dense Mode (PIM-DM).

The second approach to multicast routing is based on the assumption that the multicast group members are sparsely distributed throughout the network and bandwidth is not necessarily widely available, for example across many regions of Internet. The sparse-mode does not imply that the group has a few members, just that they are widely dispersed. In this case flooding would unnecessarily waste network bandwidth. Sparse-mode routing protocols include Core Based Trees (CBT) and Protocol Independent Multicast – Sparse Mode (PIM-SM).

Sparse-Mode PIM-SM

The Protocol Independent Multicast--Sparse Mode (PIM-SM) architecture:

- maintains the traditional IP multicast service model of receiver-initiated membership;
- uses explicit joins that propagate hop-by-hop from members' directly connected routers toward the distribution tree.
- builds a shared multicast distribution tree centered at a Rendezvous Point, and then builds source-specific trees for those sources whose data traffic warrants it.
- is not dependent on a specific unicast routing protocol.
- uses soft-state mechanisms to adapt to underlying network conditions and group dynamics.



Steps shown:

1. The sender at Source 2 registers at the Rendezvous Point Multicast Router RPt
2. A receiver joins at Rpt; there is now a bigger shared tree
3. The receiver is receiving lots of data from Source 2. The receiver sends an explicit join to Source 2 to construct a shortest path route.

Dynamic multicast membership changes

In many multicast applications, multicast participants are free to leave or join a multicast session dynamically. It is important to ensure that any change of multicast membership will not affect the traffic on the current connection and the traffic after the change remains near optimal in terms of network cost.

An example of group management protocol is IGMP Internet Group Management Protocol) which has become a standard of group management for Internet multicast. It is used by multicast routers to learn

For reducing leave latency IGMP version 2 uses Host-Membership Leave message, that is, a host must explicitly inform the router when it leaves a group.

5. Quality of Service-based Multicast routing

In the current Internet model, IP multicast routing is used to set-up the multicast tree spanning the sender and receivers. Quality of Service (QOS) is added on top of the selected route by using RSVP signalling, resource reservation, admission control and scheduling. This technique has the advantage of decoupling multicast routing from resource allocation.

But the main disadvantage is that the QOS requirements of an application cannot influence the choice of the multicast router.

An integrated QOS-based multicast routing protocol may be the right thing to do when an application has QOS requirements.

There has been some work in the area of QOS-based multicast routing algorithms [2], [3] but the are needs more practical and comprehensive schemes.

6. Tunnelling as a Transition Strategy for IP Multicast Routing.

Tunnelling in the context of multicast refers to the encapsulation of multicast packets in an IP datagram (i.e. unicast packet)to route through parts of an internetwork, such as the Internet, that don't support multicast routing. The most well-known demonstration of multicast tunneling is used in MBONE with DVMRP. The encapsulation is added on entry to a tunnel and stripped off on exit from a tunnel. Tunnels are useful as a transition strategy to achieving full native IP Multicast deployment.

References

[1] W. Stallings High-speed networks, 1998

[2] V.P. Kompella, J.C. Pasquale, and G.C Polyzos. Multicast routing for multimedia communication, IEEE/ACM Trans. Networking 1 (3) (1993) 286-292.

[3] V.P. Kompella, J.C. Pasquale, and G.C Polyzos. Multicast routing for multimedia communication, Proceedings of IEEE INFOCOM, Pages 20788-2085, 1992.

[4] Christian Huitema, Routing in the Internet, Prentice Hall, 1995.

[5] T.H. Cormen, C.E. Leiserson, R.L. Rivest, Introduction to Algorithms, MIT Press, Cambridge, MA, 1992.

IETF RFCs

See http://ds.internic.net/rfc/

RFC 1112 *Host Extensions for IP Multicasting*

RFC 1075 *Distance Vector Multicast Routing Protocol*

RFC 1584 *Multicast Extensions to OSPF*

IETF Drafts

ftp://ietf.org/internet-drafts/

Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification

[draft-ietf-idmr-pim-sm-spec-09.txt]

Internet Croup Management Protocol Version 2

*[draft-ietf-idmr-igmp-v2-08.txt]