Shared Multicast Trees And The Center Selection Problem: A Survey *

H.F. Salama D.S. Reeves Y. Viniotis
ECE Department and CSC Department
North Carolina State University
{hfsalama,reeves,candice}@eos.ncsu.edu

Abstract

MC routing protocols currently being standardized, PIM and CBT, propose the use of shared center-based trees, but they do not specify how to select the centers of these trees. We survey previous work on the center selection problems in the fields of operations research and communication networks.

1 Introduction

Recent advances in optical fiber and switch technologies have resulted in a new generation of high-speed networks that can achieve speeds of up to a few gigabits per second, along with very low bit error rates. In addition, the progress in audio, video, and data storage technologies has given rise to new distributed real-time applications. These applications may involve multimedia, e.g., videoconferencing which requires low end-to-end delays, or they may be distributed control applications requiring high transmission reliability. Quality of service (QoS) parameters are used to express the applications' requirements which must be guaranteed by the underlying network. Distributed applications will utilize future networks, and in many cases they will involve multiple users. Hence the increasing importance of multicast (MC) communication protocols which are able to manage the network resources efficiently and to satisfy the QoS requirements of each individual application. Active research on MC communication involves among others: connection establishment [1], routing [2, 3, 4], group addressing techniques [5, 6, 7], and error handling [8].

In the past, very few network applications involved multiple users and none of them had QoS requirements. In addition, the bandwidth requirements of most applications were very modest. Thus simple MC routing algorithms were sufficient to manage the network bandwidth. In many cases MC trees were simply constructed by the superposition of multiple unicast paths. The situation is different,

*This work was supported in part by the Center for Advanced Computing and Communication at North Carolina State University, and by AFOSR grants F49620-92-J-0441DEF and F49620-96-1-0061.

however, for the emerging real-time applications and their stringent QoS requirements. The delay constraint, i.e., the upper bound on end-to-end delay, is an important QoS requirement, because most real-time applications, and the interactive ones in particular, are delay-sensitive. A number of new MC routing algorithms designed specifically to meet the delay requirements of these real-time applications were proposed during the past few years. These are called delay-constrained MC routing algorithms.

All routing algorithms and protocols, that are used in current networks, use simple cost metrics such as the number of hops. However, most real-time applications make extensive use of the underlying network's resources. Hence the need for more sophisticated cost metrics to enable efficient management of these resources.

The increasing importance of MC communication has triggered intensive efforts aimed at designing standard MC routing protocols. These protocols construct MC trees that can be classified into two categories: source-specific trees (MOSPF [9], DVMRP [10], and PIM [11]) and shared trees (CBT [12] and PIM[1] [11]). A protocol that implements source-specific trees constructs a separate MC tree for each MC source. Each such tree is rooted at a source node and spans all destinations of the corresponding MC session. On the other hand, only one (or only a few) shared tree(s) is used by all sources to transmit each source's traffic stream to the destinations of a given MC session. For each shared tree, one node is selected to be its center. The center of a shared tree is assigned specific responsibilities that will be discussed in section 2.

The problem of constructing source-specific MC trees, both unconstrained and delay-constrained, has been extensively studied over the years. We surveyed and evaluated different algorithms for constructing source-specific MC trees in [13, 14, 4, 15]. In this report, we survey previous work on shared MC trees and the center selection problem.

2 Shared Trees and the Center Selection Problem

Several MC routing protocols propose the use of one single MC tree (or only a few MC trees as will be discussed later) per MC session. All sources transmitting to that MC session distribute their traffic over that tree. The use of shared trees is motivated by the fact that less overhead is required to construct and administer one shared tree per MC session than to construct and maintain a source-specific multicast tree for every source transmitting to that session. For example, when a new source joins an already existing MC session, it does not have to construct an entire (source-specific) tree that spans all members of that MC session. It merely has to find a path to connect itself to the existing shared tree. As a result, a source node does not have to keep an explicit list of the members of the MC session it is transmitting to. Similarly when a new member joins an already existing MC, it does not have to join the (source-specific) trees of all sources transmitting to that session. It simply has to find a path to connect itself to the existing shared tree. As a result, there is no need to keep an explicit list of the sources of a MC session at each member node. Eliminating the need for explicit source lists and explicit member lists, results in good scalability for the shared MC trees.

The use of shared MC trees has some disadvantages, however. The first disadvantage of shared

[1] PIM has two modes: a dense mode (DM) that uses source-specific trees, and a sparse mode (SM) that allows the use of either source-specific trees or shared trees.

trees as compared to source-specific trees is high traffic concentration. Traffic streams from different sources share the links of the shared tree which results in high traffic concentration on these links [2]. Another disadvantage is that the end-to-end delays along shared trees are longer than the corresponding delays when source-specific trees are used [16].

A shared tree does not have to have a center. However, designating a node to be the center of a shared tree facilitates several session management operations. If a node is selected to be the center of a shared tree, it performs all the operations performed by the other nodes in the tree, but in addition it may have other responsibilities. For example, it may be responsible for expanding the tree when a new source/destination joins the MC session, and it may be responsible for collecting traffic from all sources and processing that traffic before multicasting it to all destinations. Finding the optimal shared tree is difficult. Therefore, results from previous work suggest to start by selecting a node to be the center (the center selection problem) and then constructing a shared tree around that node. A node is selected to be a center based on the expected quality of the shared tree to be constructed around that node.

MC routing protocols that use shared trees propose the use of more than one center per MC session, with each center administering its own shared tree. This guarantees a higher level of fault tolerance, because if one center fails, nodes on its shared tree may be able to join the shared tree of another center. Another advantage of using multiple shared trees per MC session is that, in some cases, a single shared tree, can not satisfy the end-to-end delay requirements of all members of a MC session while multiple shared trees can. A third advantage is that multiple shared trees may be used to achieve different QoS levels. In this case each member of the MC session joins the shared tree which offers the QoS level that best satisfies its requirements.

Two types of centers have been proposed in the literature: administrative centers and distribution centers. An administrative center of a shared tree is assigned administrative responsibilities only. It is not responsible, however, for collecting the traffic streams from all sources and then multicasting them to all members of the MC session. The traffic from a source is forwarded to each MC session member over the shortest path from the source to that member along the shared tree. Such a path does not necessarily have to pass through the center. This approach is adopted in PIM-SM [11] and CBT [12]. The administrative responsibilities of a center, as specified by these protocols, include: managing the join/leave operations of members and sources., and advertizing the status of the MC session.

In PIM-SM [11], a center is called the rendezvous point (RP). Each RP administers a shared tree. A source node, or a receiver node, joins such a shared tree through the shortest path from itself to the RP. PIM-SM does not specify how to choose the RPs. In CBT [12], a center is called the core. Cores are placed by hand based on the topological distribution of the group membership at session initiation time. Sources and receivers join a existing shared tree in a similar manner as has been described above for PIM-SM's shared trees. Both PIM-SM and CBT permit the center to be any node in the network, not necessarily a member of the MC session.

The other class of centers is distribution centers. Traffic from any source must be transmitted to the distribution center first. The center may execute some application-specific operations on the arriving traffic. Then it multicasts it to all group members. In addition, distribution centers also perform administrative tasks. The construction of the MC tree is simple in this case: the shortest path from each source to the center and then the minimum Steiner tree (or delay-constrained minimum Steiner tree) from the center to all group members. This scheme was proposed for a class of videoconferences in [17]. The authors selected the node with the minimum sum of shortest paths costs to all participants in the videoconference to be the center. Thus the center did not have to be a group member. The amount of traffic that traverses the network in case of distribution centers is larger than the traffic that traverses the network if administrative centers are used instead. Therefore, most previous work on shared trees, in communication networking context, has focused on the selection of administrative centers and the construction of the corresponding shared trees which are commonly denoted as center-based trees. Distribution centers and the corresponding centralized distribution trees are advantageous only for specific classes of real-time applications (e.g., some videoconferencing applications).

3 Literature Survey

In this section, we survey previous work and performance studies of shared trees and the center selection problem. MC routing protocols that implement shared trees were already discussed briefly in section 2.

3.1 Definitions from Graph Theory

A graph $G = (V, E)$ is defined over a set of nodes V and a set of links E interconnecting the nodes. A walk in a network is an alternating sequence of nodes and links $v_0, e_1, v_1, e_2, v_2, \ldots, v_{k-1}, e_k, v_k$ such that every $e_i = (v_{i-1}, v_i) \in E$, $1 \leq i \leq k$. A walk is a path $P(v_0, v_k)$ if all its nodes are distinct. If the graph is unweighted, then the length of the path is equal to the number of links (hops) on the path. If the graph is weighted, i.e., it is a network, then the length of a path is equal to the sum of the weights of the links on that path. The graph theoretic terms defined below are useful when addressing the center selection problem [18]. They apply to both undirected and directed graphs.

The **eccentricity** of a node is the length of the shortest path to the node farthest away from that node. The **center** of a graph is the node with minimum eccentricity. The **radius** of a graph is the minimum eccentricity at any node in the graph. It occurs at the center. The **diameter** of a graph is the maximum eccentricity at any node in the graph. The **median** of a graph is the node for which the sum of the shortest distances to all other nodes in the graph is minimum. A graph may have more than one center and more than one median.

A group is a set of nodes $G = \{v_1, \ldots, v_k\} \subset V$. The center of a group is the node with the minimum shortest distance to the group member farthest away. The median of a group is the node for which the sum of the shortest distances to all other nodes in the graph is minimum. The center of a group and the median of a group do not have to be group members.

Breadth-first search can be used for computing centers and medians of unweighted graphs. For weighted graphs shortest path algorithms can be used to achieve the same goal.

3.2 Operations Research

Many variations of the center selection problem in networks have been investigated in operations research. Problems considering only one center and problems for selecting multiple centers were studied. Emergency facility location [19], hazardous facility location [20], and defensive facility location to minimize competition [21] are just a few examples of the problems studied. There is an entire research area called **Location Research** [22]. Most location research problem define the weight of a link as the delay along that link. Location problems in networks can be classified into two main categories:

- The minisum location problem: finds one (or more than one) median for a given network.
- The minimax location problem: finds one (or more than one) center for a given network.

Location research problems consider situations in which all services either originate from the center or terminate at the center. This means that the center (or media) is either the only source or the only destination[2].

An interesting version of the multiple center problem was presented in [23]: *locate a minimum number of emergency facilities such that the maximum delay from the closest emergency facility to any node in the network is less than a given value.* This is a minimax problem. The authors propose a zero-one integer programming solution.

3.3 Communication Networks

Recent work on the center selection problem is motivated by the PIM and CBT protocols. Both protocols use shared center-based trees for multicasting, but neither of them proposes an efficient method for center selection.

The use of shared multicast trees was first proposed by David Wall [16] in 1981. He investigated the problems of broadcasting and selective broadcasting (multicasting) with emphasis on broadcasting. The only link weight function he considered was link delay. Wall proposed different criteria to be optimized in a center-based tree, e.g., the average end-to-end delay or the maximum end-to-end delay. He also described the optimum center-based tree corresponding to each criterion he proposed. The optimum solutions are difficult and in some cases even *NP*-complete. To avoid this excessive complexity, Wall proposed three heuristic algorithms to select the center of a broadcast (or a MC) session. The first heuristic selects as the center the node whose shortest path tree spanning all members has the least maximum delay. The second heuristic selects as the center the node whose shortest path tree spanning all members has the least average delay. The third heuristic selects as the center the node whose shortest path tree spanning all members has the minimum diameter. In all three cases, the proposed shared tree is the shortest path tree rooted at the selected center.

In addition, Wall compared the broadcast delays along center-based trees to the delays if source-specific shortest path trees were used instead. The following is an important result from his work.

[2] The communication networks' shared MC tree problem is different. Traffic may originate from multiple sources and it must be delivered to all destinations.

Select a random node. Construct the shortest path broadcast tree rooted at that node. Use the constructed tree as a center-based shared tree that carries traffic from any node to all other nodes. The maximum delay between any two nodes along that tree is at most two times the maximum delay achieved if each node uses its own source-specific shortest path tree to broadcast its traffic instead.

Finally, Wall proposed a distributed implementation scheme for his three heuristics. His approach consists of two stages. First, each node computes its criterion assuming that all necessary information is locally available. Then each node broadcasts its criterion value. The values are compared distributedly and a center is selected.

In 1993, Wei and Estrin [2] compared source-specific shortest path MC trees to center-based MC trees using simulation. One center-based tree is compared to the corresponding source-specific trees. The authors assigned a link delay and a link cost to each link. Two groups of center-based trees are studied. Trees of the first group minimize tree cost while the trees of the other group minimize delay. The authors applied the following criteria for evaluating the quality of the different trees: the maximum end-to-end delay between any two nodes, the average end-to-end delay, the total tree cost, and the resulting traffic concentration on the network links.

The heuristics Wall proposed permit the center to be located at any node in the network. Wei and Estrin proposed heuristics that are identical to Wall's heuristics except that they only allow group members to be chosen as centers for a MC session. This reduces the amount of computation required to select the center.

Wei and Estrin's simulation results show that source-specific trees achieve smaller delays and less traffic concentration than center-based trees. However the costs of the center-based trees are less than those of the corresponding source-specific trees. Their results also show that forcing the center of a MC session to be a member of the MC group does not affect either the delay or the cost of the center-based trees.

Shukla et al. [24] proposed a protocol for constructing MC trees in asymmetrically loaded (directed) networks. Their proposal allows the use of either source-specific trees or center-based trees. The authors defined the cost of a link as a function of both the length of the link and the number of streams utilizing that link. They proposed a center selection protocol based on a tournament. A simplified version of the tournament works as follows. Each receiver is paired with a source in a decreasing order of distances. The node at the middle of the shortest path between each pair is the winner of that pair. All winners are then paired together, and the next group of winners is computed, and so on until only one winning node remains. That node is selected to be the center of the shared MC tree.

Thaler and Ravishankar [25] proposed two distributed center selection protocols, and two versions of each protocol. The two versions differ in the amount of information, about the MC session and the state of the network, each of them requires. Either one of the proposed protocols can be applied to select $n \geq 1$ centers for a given MC session. However, these protocols do not attempt to distribute the n centers evenly throughout the network. The proposed protocols allow the centers to migrate from one node to another dynamically as the group membership changes or the load on the network changes. The authors evaluate their proposed center selection algorithms and most of the previously proposed algorithms using simulation.

References

[1] M. Huang, I. Frisch, and C.-H. Chow, "Nondistributed Multiparty Connection Establishment for Broadband Networks," in *Proceedings of the 1992 IEEE International Conference on Communication (ICC'92)*, 1992.

[2] L. Wei and D. Estrin, "A Comparison of Multicast Trees and Algorithms," Tech. Rep. USC-CS-93-560, Computer Science Department, University of Southern California, September 1993.

[3] C. Noronha and F. Tobagi, "Evaluation of Multicast Routing Algorithms for Multimedia Streams," in *Proceedings of IEEE International Telecommunications Symposium*, August 1994.

[4] H. Salama, D. Reeves, Y. Viniotis, and T.-L. Sheu, "Evaluation of Multicast Routing Algorithms for Real-Time Communication on High-Speed Networks," in *Proceedings of the Sixth IFIP Conference on High Performance Networking (HPN '95)*, pp. 27–42, Chapman and Hall, September 1995.

[5] S. Deering and D. Cheriton, "Multicast Routing in Datagram Internetworks and Extended LANs," *ACM Transactions on Computer Systems*, vol. 8, pp. 85–110, May 1990.

[6] A. Segall, T. Barzilai, and Y. Ofek, "Reliable Multiuser Tree Setup with Local Identifiers," *IEEE Journal on Selected Areas in Communications*, vol. 9, pp. 1427–1439, December 1991.

[7] A. Eleftheriadis, S. Pejhan, and D. Anastassiou, "Address Management and Connection Control for Multicast Communication Applications," in *Proceedings of IEEE INFOCOM '95*, pp. 386–393, 1995.

[8] B. Kim, *Performance Analysis of k-Reliable Multicast ARQ Protocols*. PhD thesis, North Carolina State University, Department of Electrical and Computer Engineering, 1995.

[9] J. Moy, "MOSPF, Analysis and Experience." Internet RFC 1585, May 1994.

[10] D. Waitzman, C. Partridge, and S. Deering, "Distance Vector Multicast Routing Protocol." Internet RFC 1175, November 1988.

[11] S. Deering *et al.*, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification." Internet Draft, September 1995. Available at ftp://ftp.ietf.cnri.reston.va.us/internet-drafts/draft-ietf-idmr-pim-spec-02.txt.

[12] A. Ballardie, "Core Based Tree (CBT) Multicast: Protocol Specification." Internet Draft, February 1996. Available at ftp://ftp.ietf.cnri.reston.va.us/internet-drafts/draft-ietf-idmr-cbt-spec-04.txt.

[13] H. Salama, Y. Viniotis, D. Reeves, and T.-L. Sheu, "Multicast Routing Algorithms for High-Speed Networks," Tech. Rep. TR 29.1883, IBM, May 1994.

[14] H. Salama, D. Reeves, Y. Viniotis, and T.-L. Sheu, "Comparison of Multicast Routing Algorithms for High-Speed Networks," Tech. Rep. TR 29.1930, IBM, September 1994.

[15] H. Salama, D. Reeves, and Y. Viniotis, "Evaluation of Multicast Routing Algorithms for Real-Time Communication on High-Speed Networks." submitted to the IEEE Journal on Selected Areas in Communications.

[16] D. Wall, *Mechanisms for Broadcast and Selective Broadcast*. PhD thesis, Stanford University, June 1980.

[17] Y.-W. Leung and T.-S. Yum, "Optimum Connection Paths for a Class of VideoConferences," in *Proceedings of the IEEE International Conference on Communications (ICC) '91*, pp. 859–865, 1991.

[18] F. Buckley and F. Harary, *Distance in Graphs*. Addison-Wesley Publishing Company, 1990.

[19] P. Koselar and W. Walker, "An Algorithm for the Dynamic Relocation of Fire Companies," *Operations Research*, vol. 22, pp. 249–274, March–April 1974.

[20] G. List, P. Mirchandani, M. Turnquist, and K. Zografos, "Modeling and Analysis for Hazardous Materials Transportation: Risk Analysis, Routing Scheduling and Facility Location," *Transportation Science*, vol. 25, pp. 100–114, May 1991.

[21] D. Trietsch, "Optimal Multifacility defensive Location on Planes with Rectilinear Distances," *Networks*, vol. 23, pp. 517–523, September 1993.

[22] G. Handler and P. Mirchnadani, *Location on Networks: Theory and Algorithms*. The MIT Press, 1979.

[23] C. Toregas, R. Swain, C. ReVelle, and L. Bergman, "The Location of Emergency Service Facilities," *Operations Research*, vol. 19, pp. 1363–1373, October 1971.

[24] S. Shukla and E. Boyer, "Multicast Tree Construction in Network Topologies with Asymmetric Link Loads," Tech. Rep. NPS-EC-94-012, Naval Postgraduate School, Department of Electrical and Computer Engineering, September 1994.

[25] D. Thaler and C. Ravishankar, "Distributed Center-Location Algorithms: Proposals and Comparisons," in *Proceedings of IEEE INFOCOM '96*, pp. 75–84, 1996.