

Efficient Resource Allocation for Wireless Multicast

De-Nian Yang, *Member, IEEE*, and Ming-Syan Chen, *Fellow, IEEE*

Abstract—In this paper, we propose a bandwidth-efficient multicast mechanism for heterogeneous wireless networks. We reduce the bandwidth cost of an Internet Protocol (IP) multicast tree by adaptively selecting the cell and the wireless technology for each mobile host to join the multicast group. Our mechanism enables more mobile hosts to cluster together and leads to the use of fewer cells to save the scarce wireless bandwidth. Besides, the paths in the multicast tree connecting to the selected cells share more common links to save the wireline bandwidth. Our mechanism supports the dynamic group membership and offers mobility of group members. Moreover, our mechanism requires no modification to the current IP multicast routing protocols. We formulate the selection of the cell and the wireless technology for each mobile host in the heterogeneous wireless networks as an optimization problem. We use Integer Linear Programming to model the problem and show that the problem is NP-hard. To solve the problem, we propose a distributed algorithm based on Lagrangean relaxation and a network protocol based on the algorithm. The simulation results show that our mechanism can effectively save the wireless and wireline bandwidth as compared to the traditional IP multicast.

Index Terms—Heterogeneous wireless networks, multicast.

1 INTRODUCTION

THE success of wireless and mobile communications in the 21st century has resulted in a large variety of wireless technologies such as second and third-generation cellulars, satellite, Wi-Fi, and Bluetooth. The heterogeneous wireless networks combine various wireless networks and provide universal wireless access. The leading wireless companies in some countries have operated networks with multiple wireless technologies, such as T-Mobile in the United States, British Telecom in the United Kingdom, Orange Telecom in France, NTT DoCoMo in Japan, and Chunghwa Telecom in Taiwan. The number of such companies would increase because the standards for operators to provide seamless services in networks with multiple wireless technologies have been proposed by the Third-Generation Partnership Project (3GPP) [1] and Unlicensed Mobile Access (UMA) [2]. In addition, users in the heterogeneous wireless networks are usually covered by more than one cell to avoid connection drop and service disruption. More mobile terminals in the wireless networks are likely to own multiple wireless technologies. Therefore, the heterogeneous wireless networks provide the mobile hosts with many choices for the cells and wireless technologies to access the Internet.

Multicast is an efficient way for one-to-many and many-to-many communications. Each multicast group owns a set of members, and each member can be a sender or a receiver of the group. The sender in a multicast group delivers data in a multicast tree to all receivers of the group. Current

Internet Protocol (IP) multicast routing protocols adopt the shortest path trees for data delivery [3], [4], [5], [6], [7]. The path from the root of the shortest path tree to each member must be the shortest path in the network. In other words, the routing of the shortest path tree is fixed once the root and all group members have been determined. As a consequence, the bandwidth consumption in an IP multicast tree will not be able to be reduced in wired networks.

In this paper, we first comment that the bandwidth consumption in the shortest path tree can be reduced in the heterogeneous wireless networks because the routing of the shortest path tree here is more flexible. The shortest path tree in the heterogeneous wireless networks consists of two parts. The first one is composed of the cell and the wireless technology chosen by each mobile host. The second one is comprised of the wired links that connect the root of the tree and the chosen cells. Therefore, we can change the routing of the shortest path tree by selecting different cells and wireless technologies for the mobile hosts to reduce the bandwidth consumption. Consider the scenario in Fig. 1 as an example, where mobile hosts A, B, C, and D are the members of the multicast group. The example presents three different shortest path trees to serve the four mobile hosts. The first one uses a WiMax cell to serve the four mobile hosts. The second one uses a Universal Mobile Telecommunications System (UMTS) cell to serve mobile hosts A and B and two Wi-Fi cells to serve mobile hosts C and D. The third one uses four Wi-Fi cells to serve the four mobile hosts. Therefore, this example shows that the routing of the shortest path tree in the heterogeneous wireless networks is not unique.

To the best of our knowledge, there is no related work about the selection of the cell and the wireless technology for each mobile host to build a bandwidth-efficient multicast tree in the heterogeneous wireless networks. Most previous works for mobile multicast in the heterogeneous

• The authors are with the Department of Electrical Engineering, National Taiwan University, #1 Roosevelt Rd. Sec. 4, Taipei, Taiwan 106, ROC. E-mail: {dnyang, mschen}@cc.ee.ntu.edu.tw.

Manuscript received 24 Sept. 2006; revised 25 Dec. 2006; accepted 13 Feb. 2007; published online 24 July 2007.

For information on obtaining reprints of this article, please send e-mail to: tmc@computer.org, and reference IEEECS Log Number TMC-0245-0906. Digital Object Identifier no. 10.1109/TMC.2007.70739.

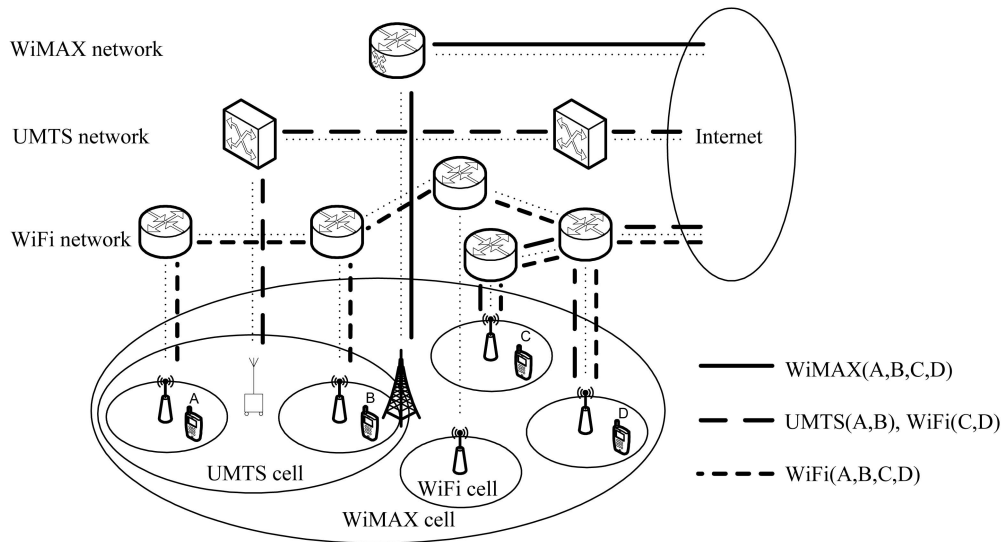


Fig. 1. An example that provides three different multicast trees by selecting different cells and wireless technologies for mobile hosts.

wireless networks focus on the efficient mechanisms to provide seamless handover between different networks [8] [9], [10], [11], [12] and the related security issues [13]. In addition, for video services, the network selection of cellular networks or Digital Video Broadcast - Handheld (DVB-H) for mobile users has been addressed [14]. Previous works also address the protocol design, reliable multicast, and other practical issues for homogeneous wireless networks [15], [16], [17], [18], [19]. Arabiah and Aljadhai [20] find a low-cost multicast tree, instead of the shortest path tree, in homogeneous wireless networks. A new member reduces the cost of the tree by connecting to the closest member and reduces the handoff delay by preestablishing multicast paths to all neighboring cells. However, resource allocation among heterogeneous wireless networks has not been addressed in the previous works. We believe that it is an important issue because current ISPs tend to operate multiple wireless networks and multiradio handsets and PDAs are appearing in the markets. Consequently, in this paper, we propose a mechanism for reducing the bandwidth consumption in the shortest path tree by adaptively selecting the cell and the wireless technology for each mobile host in the heterogeneous wireless networks. The feature distinguishes our work from others.

Explicitly, we formulate in this paper the selection of the cell and the wireless technology for each mobile host as an optimization problem, which is denoted as the Cell and Technology Selection Problem (CTSP) in the heterogeneous wireless networks for multicast communications. The problem is to select the cell and the wireless technology for each group member to minimize the total bandwidth cost of the shortest path tree. We design a mechanism, which includes an Integer Linear Programming (ILP) formulation, a distributed algorithm, and a network protocol, to solve the CTSP. We use ILP to formulate the CTSP, and the network operator can use our ILP formulation to find the optimal solution for network planning. We show that CTSP is NP-hard, which, in turn, justifies the necessity of designing efficient algorithms for suboptimal solutions. We devise an algorithm LAGRANGE, which is

based on Lagrangean relaxation [21] on our ILP formulation. We adopt the Lagrangean relaxation in our algorithm, instead of other optimization techniques, due to the following reasons: First, our algorithm decomposes the original problem into multiple subproblems such that each subproblem can be solved by each member and base station individually. In other words, the algorithm can be implemented in a distributed manner, and the important merit of the LAGRANGE algorithm enables us to design a network protocol accordingly. Second, the algorithm adapts to the change of the group membership and the mobility of group members. The algorithm iteratively reduces the bandwidth consumption according to the current group membership and the location of group members. Third, the algorithm provides the lower bound on the total bandwidth cost of the optimal shortest path tree, where the optimal shortest path tree is the shortest path tree with the optimal selection of the cell and the wireless technology for each member. For the multicast group with a large number of members, the lower bound obtained by our algorithm provides the benchmark for comparing with any algorithm for the problem since using the ILP formulation to find the total bandwidth cost of a large optimal shortest path tree is computationally infeasible.

This network protocol can be regarded as a rerouting mechanism. Note that rerouting mechanisms have been designed for unicast communication in backbone IP networks [22], circuit-switched networks [23], optical networks [24], and satellite networks [25] to reduce the bandwidth consumption. Our approach differs from the existing ones in the methods for finding a new routing according to the current one. Our approach is based on Lagrangean relaxation, which is a global optimization technique that iteratively improves the solution toward the globally optimal solution. However, most of the previous rerouting methods improve each part of the solution locally. Fortz and Thorup [22] adjust the cost of each link in shortest path routing according to the load of the link. Wong et al. [23] substitute a direct circuit-switched path with an alternate longer path. Lee and Li [24] move some links in a congested

wavelength path to the fibers with more available wavelengths. Donner et al. [25] choose nearby satellites and links of a congested or failure satellite to reroute a multiprotocol label switching (MPLS) path.

Each mobile host in our mechanism may need to switch to another cell or technology to reduce the total bandwidth cost of the shortest path tree. Therefore, our mechanism requires sophisticated handover protocols such as in the related works [26], [27], [28]. Note that our mechanism enables each mobile host to choose, either automatically or manually, the cell and the wireless technology. When some mobile hosts manually choose the cells and the wireless technologies, a partial shortest path tree spanning these mobile hosts is given, and our mechanism reduces the total bandwidth used in the tree by adaptively connecting other mobile hosts to the existing partial tree.

Overall, the contributions of this paper and the features of our mechanism are manifold:

- For each wireless technology, our mechanism reduces the number of cells used in the shortest path tree. Our algorithm clusters the mobile hosts such that nearby mobile hosts tend to use the same cell. Therefore, we can reduce the wireless bandwidth consumption even when the operator owns only one wireless technology. Our mechanism also optimizes the resource allocations for the operators with multiple wireless technologies. For a set of nearby mobile hosts, our mechanism uses a single larger cell or multiple smaller cells to serve these mobile hosts depending on the number of mobile hosts, the location of each mobile host, and the bandwidth cost of each wireless technology.
- Our mechanism is flexible since the bandwidth cost of each link and each cell can be assigned with no restriction. For example, we can concentrate on minimizing the wireless bandwidth if we assign a zero cost to each wired link and the cost model is suitable for the network with abundant wired bandwidth such as the optical network. Also, the flexible cost model enables the network operators to balance the load of both wireless cells and wired links. The network operators can increase the cost of a link or cell when the link or cell is congested [22].
- Our mechanism is transparent to the IP multicast routing protocols. The shortest path tree is created by joining the multicast group with the IP multicast routing protocols after each member has selected the cell and the wireless technology according to our mechanism. We thereby require no modification on the current IP multicast routing protocols.
- Our protocol supports the dynamic group membership. Our protocol reduces the total bandwidth cost according to the current group membership. When some mobile hosts join or leave a multicast group, each mobile host of the multicast group can adaptively initiate a horizontal or vertical handover to reduce the bandwidth consumption in the current shortest path tree.

The rest of the paper is organized as follows: Section 2 describes our assumption, presents our ILP formulation,

and shows that CTSP is NP-hard. We propose the LAGRANGE algorithm, which is based on Lagrangean relaxation, in Section 3 and our protocol based on the algorithm in Section 4. Section 5 presents our experimental results. Finally, we conclude our paper in Section 6.

2 PROBLEM DESCRIPTION

In this paper, we consider CTSP in the heterogeneous wireless networks for multicast communications. The problem is to select the cell and the wireless technology for each group member to minimize the total bandwidth cost of the shortest path tree. The total bandwidth cost of the shortest path tree consists of the total wireless bandwidth cost of the selected cells and the total wireline bandwidth cost of the shortest path tree spanning the root and each selected cell. In the following, we first describe our assumption and notation in Section 2.1. We then propose our ILP formulation and show that the CTSP is NP-hard in Section 2.2.

2.1 Assumption and Notation

In this paper, we assume that every wireless cell has the multicast capability. That is, the base station of the cell can send a single multicast packet to all mobile hosts in the cell instead of sending an individual packet to each mobile host. In addition, some members in a multicast group can be located in the wired network, but we focus on only the shortest path tree in the heterogeneous wireless networks, which could be a subtree of the whole multicast tree spanning all members. The reason is that the routing of the subtree spanning the members in the wired network is fixed. Therefore, the shortest path tree in the rest of the paper means the subtree in the heterogeneous wireless networks and the root of the subtree is the common gateway in the heterogeneous wireless networks. If each wireless network owns an individual gateway, our mechanism can still solve the CTSP by using a virtual common gateway to connect with the gateway in each wireless network and assigning a zero cost to each wireline link between the gateways. Note that we assume that the path between the root of a tree and a mobile host via each wireless network is determined by the multicast protocol in the network and is given in our problem.

The mobile hosts considered in this paper are the members of a multicast group. A cell *covers* a mobile host if the mobile host is within the transmission range of the base station of the cell. Let a cell be a *candidate cell* if the cell covers at least one mobile host. A node or link x is *downstream* to another node or link y in the shortest path tree if y is on the path from the root of the tree to x . A subtree that is *downstream* to a link e contains link e and every node and link that is downstream to e in the shortest path tree. For simplicity, the selection of the cell for each mobile host means the selection of both the cell and the wireless technology in the rest of the paper.

The notation in this paper is summarized as follows:

- C : the set of cells in the heterogeneous wireless networks.
- E : the set of links in the shortest path from each candidate cell to the root of the tree.
- M : the set of mobile hosts in the network.

- M_c : the set of mobile hosts covered by cell c , $c \in C$.
- C_m : the set of cells covering mobile host m , $m \in M$, $C_m \subseteq C$. If mobile host m selects cell c manually, we let the set C_m contain only a cell c .
- C_u : the set of cells that are downstream to node u in the shortest path tree $C_u \subseteq C$.
- E_c : the set of links in the shortest path from cell c to the root of the tree $E_c \subseteq E$.
- E_u : the set of links that are downstream to node u in the shortest path tree $E_u \subseteq E$.
- $e_{u,v}$: the link from node u to v , $e_{u,v} \in E$.
- b_c : the bandwidth cost of cell c , $c \in C$.
- $b_{u,v}$: the bandwidth cost of link $e_{u,v}$, $e_{u,v} \in E$.
- $c_{u,v}$: the cell with the base station v connected to link $e_{u,v}$, $c_{u,v} \in C$, $e_{u,v} \in E$.
- r : the root of the shortest path tree.

2.2 ILP

We use ILP to model CTSP. The ILP formulation can find the optimal shortest path tree in the heterogeneous wireless networks with any existing commercial software. Our ILP formulation has the following variables:

- $\pi_{m,c}$: a binary variable. $\pi_{m,c}$ is 1 if mobile host m selects cell c , $m \in M$, $c \in C_m$.
- σ_c : a binary variable. σ_c is 1 if cell c is used in the shortest path tree $c \in C$.
- $\lambda_{u,v}$: a binary variable. $\lambda_{u,v}$ is 1 if link $e_{u,v}$ is used in the shortest path tree $e_{u,v} \in E$.

The objective function of our ILP formulation is given as follows:

$$\min \sum_{c \in C} b_c \times \sigma_c + \sum_{e_{u,v} \in E} b_{u,v} \times \lambda_{u,v}.$$

The constraints of our ILP formulation are given as follows:

$$\begin{aligned} \sum_{c \in C_m} \pi_{m,c} &= 1, \forall m \in M, \\ \pi_{m,c} &\leq \sigma_c, \forall m \in M, \forall c \in C_m, \\ \sigma_c &\leq \lambda_{u,v}, \forall c \in C, \forall e_{u,v} \in E_c. \end{aligned}$$

Our constraints adopt a bottom-up manner. We first let each mobile host choose a cell. Once a cell is selected by any mobile host, the cell joins the shortest path tree. The first constraint guarantees that each mobile host selects one cell. For each mobile host m , the constraint enforces that $\pi_{m,c}$ of exactly one cell c is 1. The second constraint enforces that a cell is used in the shortest path tree if it is selected by any mobile host. If $\pi_{m,c}$ of any mobile host m and any cell c is 1, the inequality guarantees that σ_c must also be set as 1, and even the case that $\pi_{\hat{m},c}$ of another member \hat{m} is zero does not contradict the constraint. The third constraint states that a link is used in the shortest path tree if it is on the path from any selected cell to the root of the tree. When σ_c is 1 for a cell c , the inequality enforces $\lambda_{u,v}$ to be set as 1 for every link $e_{u,v}$ between c and the root of the tree. In addition to the above constraints, the problem has the constraints enforcing that $\pi_{m,c}$, σ_c , and $\lambda_{u,v}$ are all binary variables. We regard a set of selected cells that obey the above constraints as a *feasible solution* to the CTSP.

We show that the CTSP in the heterogeneous wireless networks for multicast communications is NP-hard because the *Minimum Set Cover* problem [29] is a special case of the CTSP problem. In Minimum Set Cover, each set is assigned a cost and covers some elements. The problem is to select the sets with the minimum total cost such that every element is covered by at least one selected set. Therefore, Minimum Set Cover is the same as the CTSP if we connect each cell in the CTSP directly to the root with a zero-cost wireline link, where each cell and mobile host in the CTSP are just the set and element in Minimum Set Cover, respectively.

3 DESIGN OF LAGRANGE ALGORITHM

In this section, we propose an algorithm for CTSP. The LAGRANGE algorithm is based on Lagrangean relaxation on our ILP formulation proposed in Section 2. The LAGRANGE algorithm has the following advantages:

- The algorithm can be implemented in a distributed manner. Each mobile host owns a cost for each covering cell and selects the cell with the smallest cost. The wireless networks compute and update the cost in a distributed manner to reduce the total bandwidth cost of the shortest path tree. No centralized server is required to maintain the group membership, the network topology, and the location of each mobile host. Therefore, the algorithm is easier to be integrated with the current IP multicast service model and protocols.
- The algorithm iteratively reduces the total bandwidth cost of the shortest path tree according to the current group membership and the set of cells covering the mobile hosts. In other words, the algorithm adapts to the dynamic join and leave of mobile hosts in a multicast group and the mobility of members.
- The algorithm provides a lower bound on the total bandwidth cost of the optimal solution to the CTSP. The lower bound can be used for comparing with the solution obtained by any algorithm for the problem.

The algorithm relaxes a constraint of our ILP formulation and transfers CTSP into the Lagrangean Relaxation Problem (LRP). The LRP owns a new objective function with the Lagrange multipliers and fewer constraints such that we can decompose the LRP into multiple subproblems, where each subproblem can be solved in a distributed manner. The members in our algorithm collaboratively construct the shortest path tree according to the solutions to the subproblems. Besides, the cost of each cell for each member is updated iteratively to reduce the total bandwidth cost of the shortest path tree according to the current group membership and the locations of members. Therefore, the algorithm is suitable for protocol design. In the rest of this section, we describe how we can solve CTSP:

- Transfer CTSP into the LRP.
- Decompose the LRP into multiple subproblems and solve each subproblem respectively.

- Select the cell and the wireless technology for each member according to the solutions to the subproblems.
- Reduce the total bandwidth cost of the shortest path tree by iteratively updating the cost of each cell for each mobile host.

3.1 Decomposing and Solving the LRP

The algorithm relaxes the second constraint in the ILP formulation to transfer CTSP into the LRP, and the objective function of the LRP is given as follows:

$$\begin{aligned} & \min \sum_{c \in C} b_c \times \sigma_c + \sum_{e_{u,v} \in E} b_{u,v} \times \lambda_{u,v} + \sum_{m \in M} \sum_{c \in C_m} \mu_{m,c} (\pi_{m,c} - \sigma_c) \\ & = \min \left[\sum_{m \in M} \sum_{c \in C_m} \mu_{m,c} \pi_{m,c} \right] \\ & \quad + \left[\sum_{c \in C} \left(b_c - \sum_{m: c \in C_m} \mu_{m,c} \right) \sigma_c + \sum_{e_{u,v} \in E} b_{u,v} \times \lambda_{u,v} \right], \end{aligned}$$

where $\mu_{m,c}$ is the Lagrange multiplier, $\mu_{m,c} \geq 0$, $\forall m \in M$, $\forall c \in C_m$. The Lagrange multiplier $\mu_{m,c}$ is the cost of cell c for mobile host m . The Lagrange multipliers connect the subproblems such that we can iteratively improve the solution to CTSP. The LRP includes the first and the third constraints. Compared with the objective function of the CTSP, the objective function of the LRP owns a new term corresponding to the relaxed constraint. Intuitively, for any feasible solution to the LRP that contradicts the relaxed constraint, namely, $\pi_{m,c} > \sigma_c$, the objective function punishes the solution with a larger objective value. Besides, any feasible solution to CTSP must also act as a feasible solution to the LRP since the set of constraints of the LRP is a subset of the constraints of CTSP. Therefore, when we adopt the optimal solution to CTSP as the feasible solution both to the LRP and CTSP, the objective value of the LRP must be not more than the objective value of CTSP because the new term in the objective function of the LRP must be nonpositive. Therefore, the objective value of the optimal solution to the LRP must be not more than the objective value of the optimal solution to CTSP. In other words, the objective value of the optimal solution to the LRP provides a lower bound to CTSP, which is the total bandwidth cost of the optimal shortest path tree.

We solve the LRP by decomposing the LRP into two subproblems. We divide the objective function and the constraints of the LRP into two parts, where each subproblem owns one part of the objective function and constraints. The variables in the two subproblems are mutually independent such that we can solve each subproblem individually, and the solution to the LRP is just the combination of the solutions to the two subproblems.

The objective function of the first subproblem is given as follows:

$$\min \sum_{m \in M} \sum_{c \in C_m} \mu_{m,c} \pi_{m,c}.$$

The first subproblem has the following constraint:

$$\sum_{c \in C_m} \pi_{m,c} = 1, \forall m \in M.$$

In the subproblem, each cell c is associated with a cost $\mu_{m,c}$ for each mobile host m . The optimal solution to the first subproblem is to find the cell with the minimum cost for each mobile host m . The runtime of the algorithm for the first subproblem is thereby $O(|M||C|)$. In the LAGRANGE algorithm, the cost $\mu_{m,c}$ for cell c is stored in each mobile host m , and each mobile host can thereby find the cell with the minimum cost individually.

The objective function of the second subproblem is given as follows:

$$\min \sum_{c \in C} \left(b_c - \sum_{m: c \in C_m} \mu_{m,c} \right) \sigma_c + \sum_{e_{u,v} \in E} b_{u,v} \times \lambda_{u,v}.$$

The second subproblem has the following constraint:

$$\sigma_c \leq \lambda_{u,v}, \forall c \in C, \forall e_{u,v} \in E_c.$$

In the subproblem, each cell c is associated with a profit $\sum_{m: c \in C_m} \mu_{m,c}$. If we choose a cell c , we acquire the profit but pay the wireless bandwidth cost b_c and wireline bandwidth cost $b_{u,v}$ of each link $e_{u,v}$ on the shortest path from c to the root. However, the wireline bandwidth cost $b_{u,v}$ for each link $e_{u,v}$ can be shared by all downstream selected cells. Therefore, the objective function of the second problem is to minimize the *net cost* of all selected cells in the shortest path tree that spans all candidate cells, and we have to find the best trade-off to select the cells.

Let $\delta_{u,v}$ denote the minimum net cost of the subtree that includes link $e_{u,v}$ and the subtree rooted at v . To find the minimum net cost of the whole shortest path tree, we consider each link of the shortest path tree in the bottom-up manner. For the link $e_{u,v}$ that connects to cell $c_{u,v}$, where $c_{u,v}$ denotes the cell with the base station v connected to link $e_{u,v}$, the minimum net cost $\delta_{u,v}$ is given as follows:

$$\delta_{u,v} = \min \left\{ 0, b_{c_{u,v}} + b_{u,v} - \sum_{m: c_{u,v} \in C_m} \mu_{m,c_{u,v}} \right\},$$

where the zero net cost corresponds to the case that cell $c_{u,v}$ is not selected, and net cost $b_{c_{u,v}} + b_{u,v} - \sum_{m: c_{u,v} \in C_m} \mu_{m,c_{u,v}}$ corresponds to the case that cell $c_{u,v}$ is selected. Therefore, net cost $\delta_{u,v}$ is guaranteed to be nonpositive. Afterward, for each of the upstream links $e_{u,v}$, we derive the minimum net cost $\delta_{u,v}$ from the minimum net cost of each downstream link $\delta_{v,w}$:

$$\delta_{u,v} = \min \left\{ 0, b_{u,v} + \sum_{w: e_{v,w} \in E} \delta_{v,w} \right\}.$$

If $\delta_{u,v}$ is negative, the above equation allows only some downstream cells and links of $e_{u,v}$ to be selected. The following proves that we can find the minimum net cost of the shortest path tree in $O(|E|)$ time:

Theorem 1. *The minimum net cost of the shortest path tree spanning all candidate cells can be obtained in $O(|E|)$ time.*

Proof. For any two child nodes x and y of the root r in the shortest path tree, the selection of the cells in the subtrees

rooted at x and y are independent. Therefore, the following equation holds:

$$\begin{aligned} & \min \sum_{c \in C} \left(b_c - \sum_{m: c \in C_m} \mu_{m,c} \right) \sigma_c + \sum_{e_{u,v} \in E} b_{u,v} \times \lambda_{u,v} \\ & = \min \sum_{v: e_{r,v} \in E} \left[\sum_{c \in C_v} \left(b_c - \sum_{m: c \in C_m} \mu_{m,c} \right) \sigma_c + b_{r,v} \times \lambda_{r,v} \right. \\ & \quad \left. + \sum_{e_{u,v} \in E_v} b_{u,v} \times \lambda_{u,v} \right], \end{aligned}$$

where the value in bracket is just the net cost $\delta_{r,v}$. The theorem follows. \square

After we obtain the minimum net cost of the shortest path tree, we can find the selected cells in the second subproblem. All cells in the subtree corresponding to a link $e_{u,v}$ are not selected if net cost $\delta_{u,v}$ is not negative. Therefore, each candidate cell c is selected in the second subproblem if the net cost $\delta_{u,v}$ of every link $e_{u,v}$ in the shortest path from c to the root of the tree is negative.

Since the variables to the first and second subproblems of the LRP are mutually independent, the solution to the LRP is just the combination of the solutions to the two subproblems. In the following, we explain how the LAGRANGE algorithm finds the solution to CTSP according to the solution to the LRP such that the algorithm can be implemented in a distributed manner.

3.2 Finding and Improving the Solution to the CTSP

Although the second subproblem selects the cells to minimize the net cost, the selected cells may not be feasible to CTSP because each mobile host is not guaranteed to be covered by a cell that is selected in the second subproblem. In other words, variables σ_c and $\lambda_{u,v}$ cannot represent the routing of the multicast tree because the second constraint is relaxed. In our algorithm, therefore, we select the cells according to the cells selected in the first subproblem. Let $\tilde{\sigma}_c$ be a binary variable that represents if the LAGRANGE algorithm selects cell c . Variable $\tilde{\sigma}_c$ is 1 if there exists at least one mobile host m that selects cell c in the first subproblem. In other words, variable $\pi_{m,c}$ is 1 in the first subproblem. Let $\tilde{\lambda}_{u,v}$ be the binary variable that represents if the LAGRANGE algorithm uses link $e_{u,v}$ in the shortest path tree. Variable $\tilde{\lambda}_{u,v}$ is 1 if link $e_{u,v}$ is on the shortest path from any selected cell c to the root, with $\tilde{\sigma}_c$ being 1. The advantage of the LAGRANGE algorithm is that the cell selection is performed only by the mobile host and is transparent to the IP multicast routing protocols. In other words, the shortest path tree is created by each member joining the multicast group with the IP multicast routing protocols after each member selects the cell according to our algorithm. We thereby require no modification on the current IP multicast routing protocols.

Each member m in the LAGRANGE algorithm selects the cell c according to the cost $\mu_{m,c}$, the Lagrange multiplier, of the cell in the first subproblem. We adjust the cost iteratively with the subgradient algorithm [21] and the solutions to the two subproblems of the LRP. Let $W(\boldsymbol{\mu})$ denote the objective function of the LRP in Section 3.1,

where $\boldsymbol{\mu} = (\mu_{m,c}, \forall m \in M, \forall c \in C_m)$. The subgradient corresponding to the optimal solution of the LRP is denoted as $\nabla W(\boldsymbol{\mu}) = (\partial W(\boldsymbol{\mu}) / \partial \mu_{m,c}, \forall m \in M, \forall c \in C_m)$, where

$$\frac{\partial W(\boldsymbol{\mu})}{\partial \mu_{m,c}} = \pi_{m,c} - \sigma_c.$$

The feasible solution to CTSP obtained by the LAGRANGE algorithm depends on the cost $\mu_{m,c}$ and the subgradient $\partial W(\boldsymbol{\mu}) / \partial \mu_{m,c}$ indicates the direction of adjusting $\mu_{m,c}$ to find the better feasible solution to CTSP $\forall m \in M, \forall c \in C_m$. The LAGRANGE algorithm increases or decreases the cost at each iteration. The LAGRANGE algorithm increases $\mu_{m,c}$ when $\pi_{m,c} - \sigma_c$ is positive. In other words, mobile host m selects cell c , but the cell is not selected in the second subproblem in this case. In contrast, the LAGRANGE algorithm decreases $\mu_{m,c}$ when $\pi_{m,c} - \sigma_c$ is negative. In other words, mobile host m does not select cell c , but the cell is selected in the second subproblem in this case.

We explain the adjustment of the costs in an intuitive way as follows: Although the solution to the second subproblem may not find the feasible shortest path tree in CTSP, the second subproblem provides an insight to find the shortest path tree with low wireless and wireline bandwidth costs. The second subproblem tends to select the cells that cover more mobile hosts to save the wireless bandwidth because the cell tends to own a larger profit, where the profit of each cell c is $\sum_{m: c \in C_m} \mu_{m,c}$. Besides, to save the wireline bandwidth, the second subproblem tends to select the cells such that the shortest paths from the cells to the root share more common wireline links. Therefore, if mobile host m does not select cell c , but the cell is selected in the second subproblem, we decrease $\mu_{m,c}$ such that the cell tends to own the lowest cost and to be chosen by m afterward. Therefore, the cost $\mu_{m,c}$, which is the Lagrange multiplier in the LRP, plays an important role to select the cells and find the bandwidth-efficient shortest path tree.

Fig. 2 shows the details of our algorithm. Initially, the Lagrange algorithm assigns a unit cost to each cell for each member in step 1, and each member can thereby select any cell. Afterward, our algorithm iteratively reduces the total bandwidth cost of the shortest path tree. At each iteration, our algorithm first finds the solution to the first subproblem in step 2. The algorithm then finds the solution to the second subproblem in steps 3 and 4. The algorithm then adjusts the cost of each cell for each member in step 5 such that we can find the shortest path tree with a smaller bandwidth cost at the next iteration. The LAGRANGE algorithm stops when the number of iterations η is larger than a threshold N , when our algorithm can no longer adjust the cost, or when the difference of the total bandwidth cost of the obtained shortest path tree and the lower bound on the total bandwidth cost of the optimal shortest path tree is within a threshold θ . The parameter γ in step 5 is the normalization constant. The parameter ε in step 5 is a parameter that dominates the modification of the cost at each iteration in the subgradient algorithm. With a larger ε , the shortest path tree improves faster, but the obtained shortest path tree tends to consume more bandwidth than the obtained shortest path tree with a

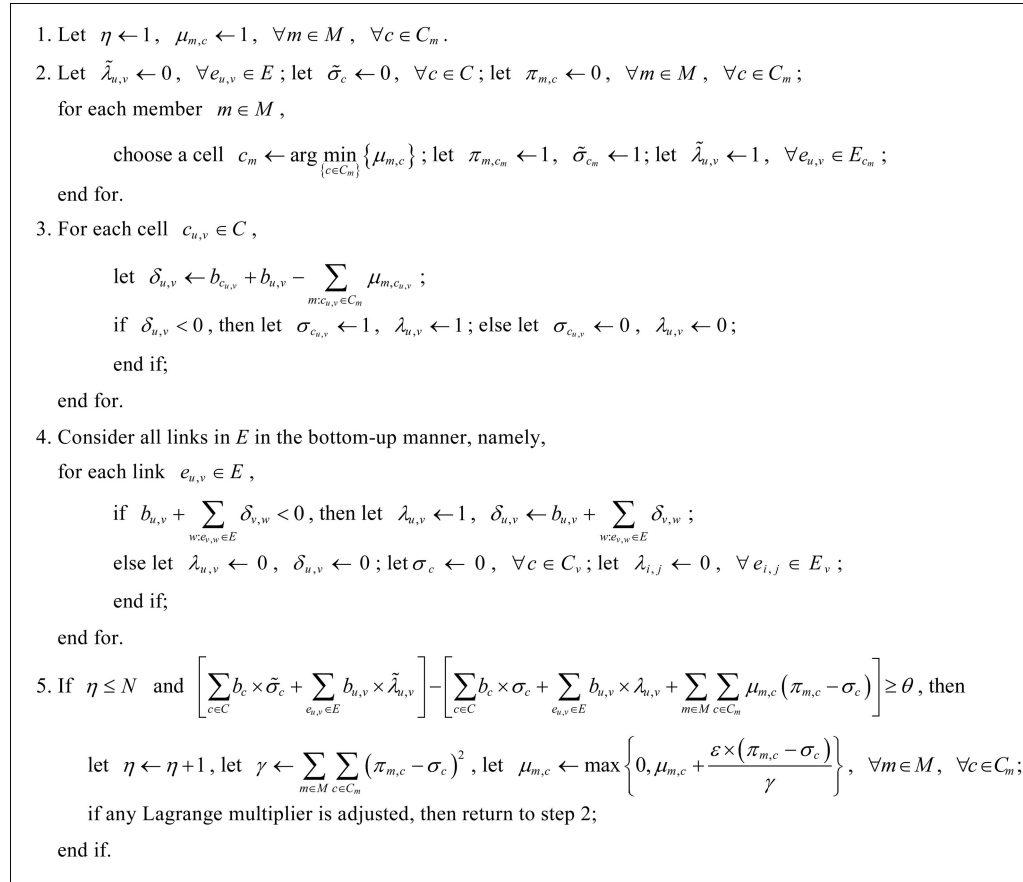


Fig. 2. The LAGRANGE algorithm.

smaller ε . In this paper, we thereby reduce ε as the improvement of the shortest path tree becomes smaller.

Consider Fig. 3 as an example of the LAGRANGE algorithm. Fig. 3a shows the network topology with five routers, six cells, and seven mobile hosts. The wireline and wireless costs are both 1. A mobile host is covered by the cell if it connects to the cell with a wireless link. Fig. 3b shows the first iteration of the LAGRANGE algorithm. The value beside each wireless link is the cost of the cell for the mobile host at the beginning of the iteration. A wireless link is represented by a solid line if the mobile host chooses the cell in the first subproblem. A cell is represented by a solid circle if the cell is selected in the second subproblem. The value beside each wireline link is the net cost in the subtree containing the link and the downstream subtree in the second subproblem. A wireline link is represented by a solid line if the link is adopted in the shortest path tree of CTSP for data delivery. Therefore, every cell is selected in the first subproblem, no cell is selected in the second subproblem, and the shortest path tree of the CTSP adopts all wireline links in Fig. 3b. Because no cell is selected in the second subproblem, the LAGRANGE algorithm increases the cost of a cell for the mobile host if the mobile host chooses the cell in the first subproblem, and Fig. 3c shows the new cost of each cell for each mobile host.

Each of the mobile hosts H_2, H_3, H_4, H_5 , and H_6 in Fig. 3c chooses another cell in the first subproblem because the new cell owns a lower cost. The new shortest path tree of the CTSP induces a smaller bandwidth cost in Fig. 3c. It

does not use cell C_2 and the wireline link from R_2 to C_2 because no mobile host selects cell C_2 in the first subproblem. Cells C_1 and C_3 are selected in the second subproblem. At the end of the iteration, the LAGRANGE algorithm decreases the costs of C_3 for mobile hosts H_2 and H_3 because the two mobile hosts do not select the cell in the first subproblem, but the cell is selected in the second subproblem. Fig. 3d shows the new cost of each cell for each mobile host.

The new shortest path tree induces a smaller bandwidth cost in Fig. 3d. It does not use cell C_4 and the wireline links from R_3 to C_4 . Before the LAGRANGE algorithm finds the new cost of each cell for each mobile host, mobile host H_3 hands over from C_4 to C_2 , mobile host H_5 moves out the coverage area of C_4 , and mobile host H_7 leaves the multicast group in Fig. 3e. Note that the shortest path tree of the CTSP needs to use cell C_2 and the wireline link from R_2 to C_2 because mobile host H_3 selects cell C_2 in the first subproblem. However, after the LAGRANGE algorithm updates the cost of each cell for each mobile host, mobile host H_3 selects cell C_3 in the first subproblem, and the new shortest path tree discards cell C_2 and the wireline link from R_2 to C_2 . Moreover, the new tree also saves the wireless bandwidth of cell C_6 and the wireline bandwidth of the link from R_5 to C_6 because mobile host H_7 chooses another cell in the second subproblem.

The above example shows that the LAGRANGE algorithm can reduce both the wireline and wireless bandwidth

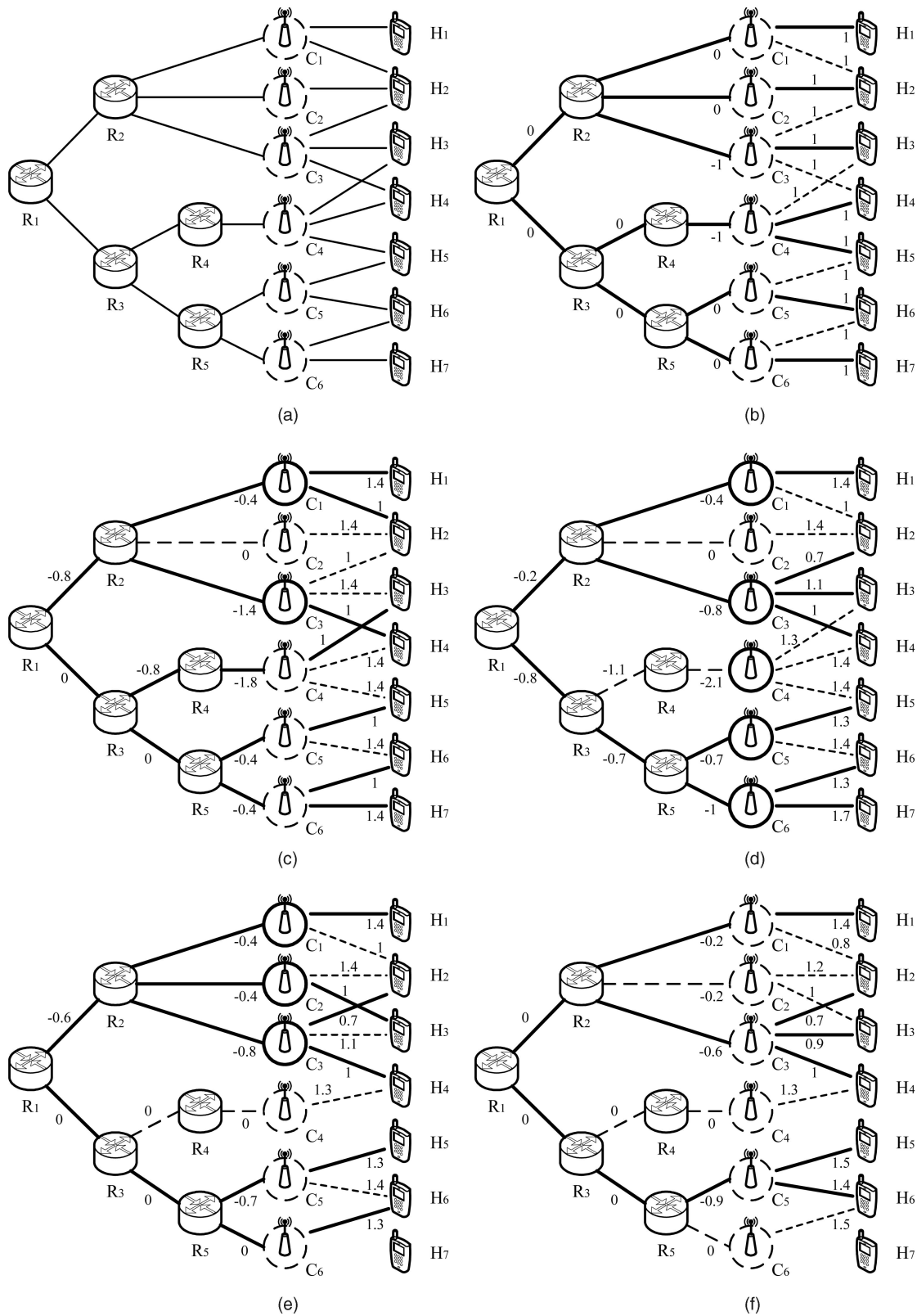


Fig. 3. An example of the LAGRANGE algorithm. (a) Network topology. (b) First iteration. (c) Second iteration. (d) Third iteration (before handoff and leave). (e) Third iteration (after handoff and leave). (f) Fourth iteration.

of the shortest path tree in the wireless networks. This example also indicates that the LAGRANGE algorithm adapts to the change of the group membership and the mobility of members. Each mobile host in the algorithm is

able to individually select the cell since the cell selected by each mobile host is the one with the lowest cost among all covering cells independent of the cells that cover other mobile hosts. Therefore, the LAGRANGE algorithm is

suitable to be implemented in a distributed manner. In the next section, we present the proposed protocol based on the algorithm.

4 PROTOCOL DESIGN

In this section, we propose a distributed protocol based on the LAGRANGE algorithm in Section 3. Our protocol is transparent to the current IP multicast routing protocols. Each mobile host individually selects the cell with the lowest cost to join the multicast group with the multicast routing protocols, and we regard the shortest path tree for data delivery as the *data tree* of the multicast group. To iteratively reduce the cost of the data tree, our protocol needs to solve the second subproblem of the LRP. Our protocol builds a *control tree* to solve the second subproblem in a distributed manner, where each router and base station in the control tree maintains a *node agent* and *cell agent*. Initially, the control tree spans every candidate cell. Afterward, our protocol incrementally prunes the control tree to reduce the protocol overhead.

In the following, we explain the information stored in each agent in Section 4.1, the control messages in Section 4.2, and the operations of our protocol in Section 4.3.

4.1 State

The information stored in each agent is a *soft state* to guarantee the robustness of the protocol. A soft state stored in an agent needs to be refreshed periodically by neighbor agents or covering mobile hosts. Therefore, when any incident link or adjacent agent fails, the node agent removes the state to ensure that no idle state remains in the agent.

Each node in a multicast tree stores the required information for its parent node and child nodes to build the tree in a distributed manner, which is identical to the standard multicast routing protocols. Each node agent stores the following states: 1) multicast group address, 2) the address of the parent node agent in the control tree, and 3) the bandwidth cost of the link that connects the node agent and the parent node agent. For each child agent in the control tree, the node agent stores the address of the child agent and a Join Timer. We use the Join Timer to maintain the soft state for the child agent. When the timer times out, the node agent removes the state corresponding to the child agent.

In addition to the above information, each cell agent stores the following information: 1) the bandwidth cost of the cell, 2) a Control Flag, which is either TRUE or FALSE, indicating whether the cell is selected in the second subproblem of the LRP in Section 3, and 3) a Data Flag, which is either TRUE or FALSE, indicating whether the base station of the cell is in the data tree. For each mobile host covered by the cell, the cell agent also stores the following information: 1) the address of the mobile host, 2) the cost of the cell for the mobile host, namely, the Lagrange multiplier, and 3) a Join Timer. When the timer times out, the cell agent removes the state corresponding to the mobile host.

4.2 Control Messages

Our protocol has the following control messages: Join, Join_Ack, Leave, Request, Reply, and Inform. We introduce each control message as follows:

- *Join*. Each mobile host or node agent sends a Join message to a cell agent or the parent node agent to join the control tree. Each mobile host or node agent also periodically sends a Join message to refresh the soft state.
- *Join_Ack*. Each cell agent or node agent returns a Join_Ack message to confirm the Join message after it receives a Join message. The Join_Ack message sent by a cell agent also contains the Data Flag and the cost of the cell for the mobile host. Each mobile host may select another cell after it acquires the cost for the cell from the cell agent.
- *Leave*. Each mobile host sends a Leave message to each covering cell agent when it decides to leave the multicast group or when it is no longer covered by the cell. Each cell agent sends a Leave message to the parent node agent if it has no covered mobile host or if it is pruned in the control tree. Each node agent sends a Leave message to the parent node agent if it has no child agent or if it is pruned in the control tree.
- *Request, Reply, and Inform*. Our protocol uses the three messages to update the cost of each cell in a distributed manner. The root of the control tree periodically initiates an update procedure by sending a Request message along the tree to each cell agent. The Reply message finds the net cost of the control tree. Each node agent sends an Inform message to the downstream cells if it obtains a zero net cost. The Inform message determines the Control Flag of each cell agent.

4.3 Operations

We describe the protocol operations as follows:

1. *Join a multicast group*. When a mobile host decides to join a multicast group, it sends a Join message to the cell agent of each cell that covers the mobile host. If the mobile host receives a Join_Ack message from any cell agent, with Data Flag being TRUE, it selects the cell to join the multicast group. Otherwise, it can select any cell to join the group. When a cell agent receives a Join message from a new mobile host, it creates a new state for the mobile host. If the cell agent has not joined the control tree, it sends a Join message toward the root. The Join message propagates upstream until it reaches the root or any node agent that is in the control tree.
2. *Hand over to a new cell*. When a mobile host hands over to a new cell, it sends a Join message to the new cell and a Leave message to the original cell. The Leave message may propagate via the new cell if the mobile host is no longer covered by the original cell. A cell agent removes the state for the mobile host after it receives the Leave message from the mobile host. If the Leave message fails to reach the original cell agent, the cell agent can still remove the state for

the mobile host after the Join timer for the mobile host times out.

3. *Update the cost of each cell.* The root of the control tree periodically sends a Request message along the control tree to each cell agent. After the message arrives at a cell agent, the cell agent first calculates the net cost according to the LAGRANGE algorithm in Section 3. If the net cost is negative, the cell agent then sets the Control Flag as true and sends a Reply message with the net cost to its parent node agent; otherwise, it sets the Control Flag as FALSE and sends a Reply message with a zero net cost to its parent node agent. The cell agent then updates the cost for each covered mobile host.

Each node agent first finds the net cost according to the net cost included in the Reply message from every child node agent. The node agent then sends a Reply message with the obtained net cost to its parent node agent. Besides, if the node agent obtains a zero net cost, it sends an Inform message to each child node agent that sends a Reply message with a negative net cost. The Inform message propagates downstream in the similar way to each cell agent that sends a Reply message with a negative net cost. In this case, each cell agent then changes the Control Flag to FALSE and updates the cost for each covered mobile host. Consider Fig. 3c as an example. Node R_3 obtains a zero net cost and sends an Inform message to R_4 . The Inform message propagates downstream to C_4 . Therefore, cell C_4 knows that it is not selected in the second subproblem of the LRP and sets the Control Flag as FALSE.

4. *Prune the control tree.* Our protocol incrementally prunes the control tree to reduce the overhead of the protocol. When a cell agent or node agent obtains a zero net cost for a period of time, it sends a Leave message to the parent node agent. A node agent removes the state of the group and leaves the control tree if it receives a Leave message from every child agent. Therefore, the prune procedure enables fewer node agents to store the states and send the messages of our protocol. Consider Fig. 3b as an example. Cells C_1 , C_2 , C_5 , and C_6 and node R_4 send the Leave messages upstream because each of them obtains a zero net cost. Nodes R_3 and R_5 can remove the states of the group and leave the control tree because each of them receives the Leave message from every child agent. Nodes R_2 and R_4 still need to store the state of the group because each of them has a child agent with a negative net cost. However, node R_2 can remove the states for C_1 and C_2 because the two cell agents send the Leave messages. Note that the prune procedure still maintains the correctness of our algorithm because any node agent finds the same net cost when a child agent sends no Reply message or a Reply message with a zero net cost. Besides, a cell agent or node agent rejoins the control tree once it obtains a negative net cost. Therefore, our protocol reduces the protocol overhead without sacrificing the bandwidth efficiency of the data tree.
5. *Leave the multicast group.* Each mobile host sends a Leave message to a cell agent when it decides to

leave the multicast group. Each cell agent leaves the control tree if it covers no mobile host. Each node agent leaves the control tree if it has no child agent.

5 EXPERIMENTAL STUDIES

In this section, we present our simulation results. To the best of our knowledge, there is no related algorithm for CTSP in the previous works. Therefore, we compare the LAGRANGE algorithm with two other algorithms that can represent the reasonable user behaviors. In the first algorithm RAND, each mobile host randomly selects a cell. In the second algorithm LOCAL, each mobile host locally selects the wireless technology with the minimum bandwidth cost because the mobile host tends to spend the least monetary cost in this case. The mobile host selects the cell with the minimum distance to the base station if there is more than one cell with the minimum bandwidth cost. Moreover, we also compare the solution obtained by our algorithm with the optimal solution obtained by CPLEX with our ILP formulation in small wireless networks. In large wireless networks, we compare the solution obtained by our algorithm with the lower bound on the total bandwidth cost of the optimal shortest path tree, where we find the lower bound from the solution of the LRP in Section 3.

To test the performance of our algorithm in different scenarios, we change the following parameters:

1. *Group size.* The group size is the number of members, namely, mobile hosts, in a multicast group. We change the group size to test the scalability of our algorithm and protocol.
2. *Transmission range of a base station.* For each wireless technology, the size of the overlapping area of adjacent cells is different when the transmission range of a based station changes.
3. *Bandwidth cost of each link and cell.* The network operators can assign a larger bandwidth cost to a wireless cell rather than a wireline link. The network operators can also give a larger bandwidth cost to a congested link or cell to balance the traffic load in the networks. Besides, we also consider that every wireline link is assigned a zero cost to represent the case that the network operators concern only the wireless bandwidth consumption.

We measure 100 samples in each scenario. The performance metrics in our simulation are listed as follows:

1. *Total bandwidth cost of the data tree and the control tree.* The data tree is the shortest path tree for data delivery, and the control tree is the shortest path tree in our protocol to solve the second subproblem of the LRP in a distributed manner.
2. *Number of links and cells in the data tree and the control tree.* The number of control messages and the number of nodes storing the agent of our protocol are proportional to the number of links and cells in the control tree.

In our simulation, we adopt the probability-based mobility model that is widely used in the previous works [8] [9], [10], [19]. We distribute all mobile hosts uniformly at

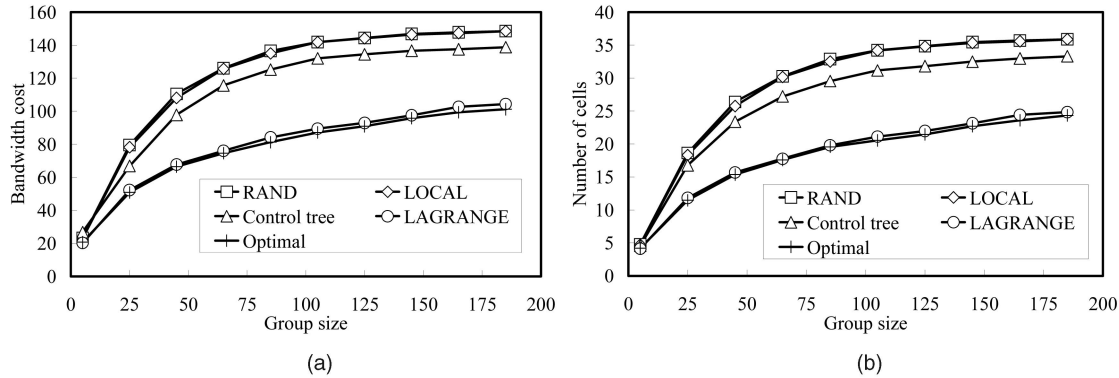


Fig. 4. Simulation results of small wireless networks. (a) Total bandwidth cost. (b) Number of cells in the tree.

random in the networks initially. Afterward, each mobile host decides to stay or move at each minute. If the mobile host decides to move, it randomly chooses the direction and the speed between 0 and 2 km per minute. We adopt the join/leave dynamics measured at MBone [30]. The arrival of a new member is modeled as a Poisson process with a rate of 0.5 per minute, and the membership duration is exponentially distributed with a rate of 0.5 per minute. At the end of each minute, we find the set of cells covering each mobile host and then obtain the shortest path trees with the RAND and LOCAL algorithms. For the LAGRANGE algorithm, we run one iteration every 1 minute to test the algorithm in the scenario with the dynamic group membership and mobility of members. To prune the control tree incrementally, each cell agent or node agent in our protocol sends a Leave message to the parent node agent after the obtained net cost remained zero for 5 minutes. The simulation time takes 1,000 minutes. At the end of each simulation, we average the performance metrics measured at each minute.

In the following, we present our simulation results for small wireless networks in Section 5.1 and the results for large wireless networks in Section 5.2. We then explain the transient behavior of the LAGRANGE algorithm in Section 5.3.

5.1 Results for Small Wireless Networks

We first compare the solutions obtained by the LAGRANGE algorithm with the optimal solutions obtained by our ILP formulation with CPLEX [31]. We simulate only small wireless networks because solving large ILP problems is computationally infeasible. The network is in a 25 km × 25 km service area and has 36 hexagon cells. The base stations of every adjacent nine cells are connected to a router, and each router is connected to the gateway. The bandwidth cost of each cell and link is 1 and 3. Fig. 4 presents the total bandwidth cost and the number of cells used in the data tree and the control tree. The number of links used in the data tree and control tree is similar to Fig. 4b. Fig. 4 shows that our algorithm outperforms both RAND and LOCAL. Our algorithm saves about 40 percent of bandwidth cost. Besides, the total bandwidth cost obtained by our algorithm is very close to the optimal solution. Although our protocol maintains a control tree for

each group, the total bandwidth cost of the control tree is smaller than the total bandwidth cost of the data trees obtained by RAND and LOCAL. The reason is that our protocol incrementally prunes the control tree to reduce the size of the tree.

5.2 Results for Large Wireless Networks

Fig. 5 shows the simulation results for large heterogeneous wireless networks. We adopt the hierarchical networks in the Georgia Tech Internetwork Topology Models (GT-ITM) network generator [32] to represent the hierarchical architecture in the current wireless networks. There are three wireless networks in a 50 km × 50 km service area with parameters (144, 16, 9, 1, 2), (64, 16, 4, 2, 4), and (4, 4, 0, 5, 15), where the five parameters in each parentheses are the number of hexagon cells, the number of cells connected to each second-level router, the number of second-level routers connected to each first-level router, the bandwidth cost of each link, and the bandwidth cost of each cell. In the first and second networks, the first-level routers connect to the gateway. In the third network, the second-level routers connect to the gateway directly, and there is no first-level router. Fig. 5b increases the transmission range of the base station of each cell to 1.3 times the original transmission range. Figs. 5c and 5d change the bandwidth cost of each wireline link to zero. Since finding the optimal solution with the ILP formulation in large networks is computationally infeasible, here, we use the lower bound on the total bandwidth cost of the optimal shortest path tree provided by the LRP in Section 3. The total bandwidth cost of the optimal shortest path tree must be larger than the lower bound but smaller than the cost of the solution obtained by any algorithm. Therefore, Fig. 5 implies that the solutions obtained by our algorithm are close to the optimal solutions. Besides, the LOCAL algorithm outperforms RAND in Fig. 5 because mobile hosts in LOCAL tend to select the wireless technologies with smaller bandwidth costs. Therefore, the shortest path tree in LOCAL tends to use fewer wireless technologies.

According to Figs. 5a and 5b, the total bandwidth cost of a data tree obtained by our algorithm decreases as the transmission ranges of the base stations increase. The reason is that each cell covers more mobile hosts such that our algorithm clusters the mobile host and requires fewer

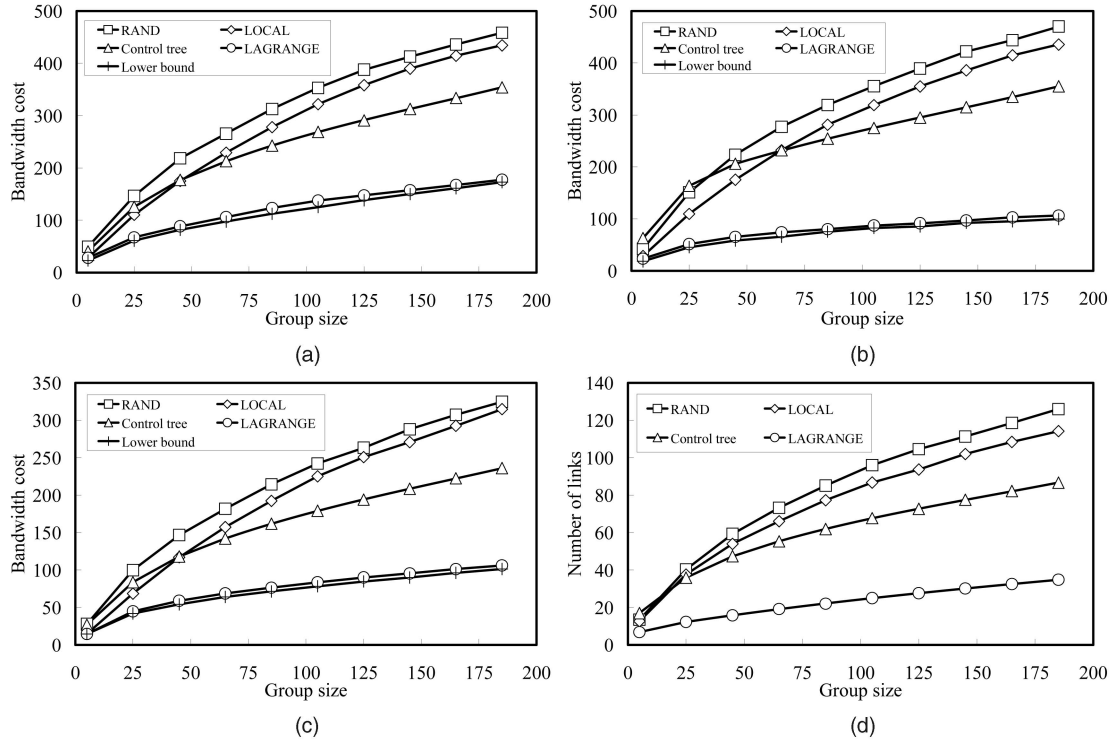


Fig. 5. Simulation results of large wireless networks. (a) Original scenario. (b) Larger transmission range. (c) Zero bandwidth cost for each link. (d) Zero bandwidth cost for each link.

wireless cells and wireline links. However, the results of RAND and LOCAL are the same for different transmission ranges of the base stations. A mobile host in LOCAL still selects the cell in which the base station is closest, even though there are more cells covering the mobile host. The RAND and LOCAL algorithms cannot reduce the total bandwidth cost of a data tree because each mobile host individually and independently selects the cell, ignoring the possibility of sharing the bandwidth cost with adjacent mobile hosts.

Figs. 5c and 5d demonstrate that our algorithm outperforms RAND and LOCAL in the case that the network operators concern only the wireless bandwidth. Therefore, we believe that each cell in our algorithm can support more multicast groups because each group uses much fewer cells. Moreover, Fig. 5d shows that our algorithm uses less wireline bandwidth, even though we only minimize the consumption of the wireless bandwidth. The reason is that there are fewer cells in a data tree such that we need fewer links to connect to the cells.

5.3 Transient Behavior of the LAGRANGE Algorithm

Fig. 6 presents the transient behavior of the LAGRANGE algorithm with 65 members in the multicast group. We change the probability that each mobile host decides to move at each iteration. Fig. 6a first shows the total bandwidth cost after each iteration of the algorithm when no mobile host moves. Fig. 6a indicates that the total bandwidth cost of the LAGRANGE algorithm approaches the lower bound on the total bandwidth cost of the optimal shortest path tree, and the control tree is pruned iteratively.

Our algorithm, at some iterations, generates the shortest path trees with slightly larger bandwidth costs than the trees in the previous iterations. The reason is that our algorithm, which is based on Lagrangean relaxation, searches the slightly worse solutions to avoid trapping in locally optimal solutions [21]. Figs. 6b, 6c, and 6d change the frequency of each mobile host to move to a new location. The average bandwidth cost of a data tree slightly increases when a mobile host moves more frequently. However, Figs. 5a and 6d show that the total bandwidth costs of the data tree and the control tree are still less than the total bandwidth cost of the data tree generated by RAND and LOCAL.

Our algorithm converges toward the optimal solution iteratively. The convergence time is correlated to the number of iterations and the time between two iterations, where the latter one can be assigned by the network operators. Therefore, given the mobility frequency of users, the network operators can set an appropriate period of time between two iterations for our algorithm. Ideally, the network operator can set a short period of time such that our algorithm finds the best solution before a mobile host hands over. However, this approach induces a large amount of overhead, since all iterations are performed within a short period of time. Therefore, the trade-off between the quality of the solution and the time between two intervals are shown in Fig. 6. The figure with a higher mobility corresponds to a larger interval between two iterations in our algorithm.

Fig. 6 shows that our protocol incrementally prunes the control tree. The bandwidth costs of the control trees are identical to the bandwidth costs of data trees in Figs. 6a and 6b when our algorithm obtains the best solutions. In Figs. 6c

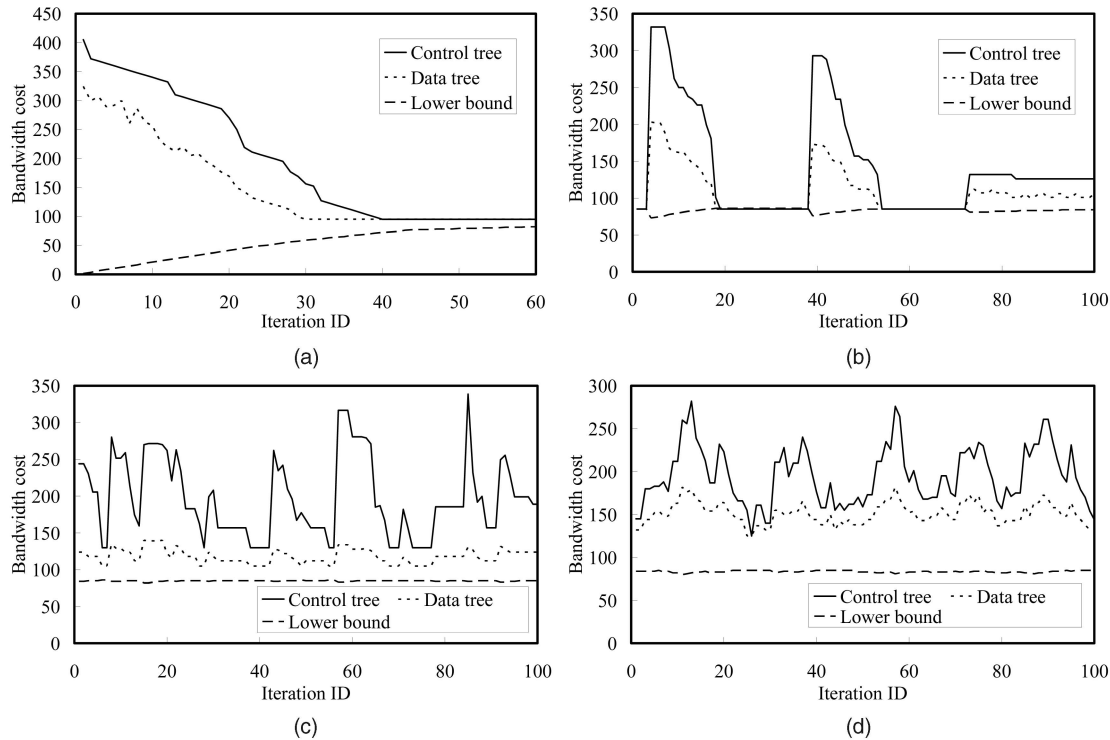


Fig. 6. Transient behavior of the LAGRANGE algorithm with different mobility. (a) Probability = 0 percent. (b) Probability = 0.1 percent. (c) Probability = 0.5 percent. (d) Probability = 2 percent.

and 6d, the bandwidth costs of control trees decrease iteratively before a mobile host hands over. However, the costs increase after a handoff because our algorithm temporarily needs a larger control tree to reduce the cost of the data tree in a distributed manner. In addition, the bandwidth cost is different from the bandwidth consumption. The bandwidth cost is proportional to the number of links and cells in a multicast tree. Since our protocol induces only several control messages in each link and cell at each iteration, the bandwidth consumption in a control tree is much smaller as compared to the data tree.

6 CONCLUSION

In this paper, we have proposed a new mechanism for reducing the total bandwidth cost of the IP multicast tree by adaptively selecting the cell and the wireless technology for each mobile host. We model the selection of the cell and the wireless technology for each mobile host as an optimization problem. We use ILP to formulate the optimization problem and show that the problem is NP-hard. The network operator can use the ILP formulation to find the optimal solution for network planning in small wireless networks. We design an algorithm based on Lagrangean relaxation and devise a distributed protocol based on the algorithm. Our algorithm iteratively reduces the total bandwidth cost of the shortest path tree. Our protocol supports the dynamic group membership and mobility of members. Moreover, our protocol requires no modification on the current IP multicast routing protocols. Our simulation results show that our mechanism can effectively save the network bandwidth compared with the traditional IP multicast.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Council of Taiwan under Contract NSC93-2752-E-002-006-PAE. The authors would like to thank the anonymous referees for their helpful comments to improve this paper.

REFERENCES

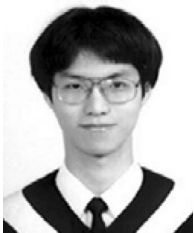
- [1] 3GPP System to Wireless Local Area Network (WLAN) Interworking, 3GPP TS 23.234, 2007.
- [2] *Unlicensed Mobile Access (UMA)*, <http://www.umatechnology.org/index.htm>, 2007.
- [3] D. Waitzman, C. Partridge, and S. Deering, *Distance Vector Multicast Routing Protocol*, IETF RFC 1075, 1988.
- [4] J. Moy, *Multicast Extensions to OSPF*, IETF RFC 1584, 1994.
- [5] D. Estrin et al., *Protocol-Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification*, IETF RFC 2117, 1997.
- [6] A. Ballardie, *Core-Based Trees (CBT Version 2) Multicast Routing Protocol Specification*, IETF RFC 2189, 1997.
- [7] S. Bhattacharyya, *An Overview of Source-Specific Multicast (SSM)*, IETF RFC 3569, 2003.
- [8] T.G. Harrison, C.L. Williamson, W.L. Mackrell, and R.B. Bunt, "Mobile Multicast (MoM) Protocol: Multicast Support for Mobile Hosts," *Proc. ACM MobiCom*, pp. 151-160, 1997.
- [9] C.R. Lin and K.-M. Wang, "Mobile Multicast Support in IP Networks," *Proc. IEEE INFOCOM*, vol. 3, pp. 1664-1672, 2000.
- [10] Y. Wang and W. Chen, "Supporting IP Multicast for Mobile Hosts," *ACM Mobile Networks and Applications*, vol. 6, no. 1, pp. 57-66, Jan. 2001.
- [11] J.-R. Lai, W. Liao, M.-Y. Jiang, and C.-A. Ke, "Mobile Multicast with Routing Optimization for Recipient Mobility," *Proc. IEEE Int'l Conf. Comm. (ICC '01)*, vol. 5, pp. 1340-1344, 2001.
- [12] M. Hossain, A.K. Elhakeem, and W. Hamouda, "Handoff Latency Improvement Using Multicasting Schemes in Heterogeneous Networks," *Proc. 16th IEEE Int'l Symp. Personal Indoor and Mobile Radio Comm. (PIMRC '05)*, vol. 4, pp. 2766-2770, 2005.
- [13] Y. Sun, W. Trappe, and K. Liu, "A Scalable Multicast Key Management Scheme for Heterogeneous Wireless Networks," *IEEE/ACM Trans. Networking*, vol. 12, no. 4, pp. 653-666, Aug. 2004.

- [14] L. Huang, K.A. Chew, and R. Tafazolli, "Network Selection for One-to-Many Services in 3G-Broadcasting Cooperative Networks," *Proc. 61st IEEE Vehicular Technology Conf. (VTC '05-Spring)*, vol. 5, pp. 2999-3003, 2005.
- [15] M. Hauge and Ø. Kure, "Multicast in 3G Networks: Employment of Existing IP Multicast Protocols in UMTS," *Proc. Fifth ACM Int'l Workshop Wireless Mobile Multimedia (WoWMoM '02)*, 2002.
- [16] R. Rummler and H. Aghvami, "End-to-End IP Multicast for Software Upgrades of Reconfigurable User Terminals within IMT-2000/UMTS Networks," *Proc. IEEE Int'l Conf. Comm. (ICC '02)*, vol. 1, pp. 502-506, 2002.
- [17] S.K. Palat, I.N. Weerasekera, and A. Casati, "Multicasting in UMTS," *Proc. Third IEEE Int'l Conf. 3G Mobile Comm. Technologies*, pp. 96-101, 2002.
- [18] U. Mudugamuwa, M. Karaliopoulos, R. Tafazolli, and B. Evans, "Reliable Multicast Transport and Power Scheduling for MBMS Delivery over 3G Mobile Satellite Systems," *Proc. 59th IEEE Vehicular Technology Conf. (VTC '04-Spring)*, vol. 5, pp. 2836-2841, 2004.
- [19] R. Rummler, Y.W. Chung, and A.H. Aghvami, "Modeling and Analysis of an Efficient Multicast Mechanism for UMTS," *IEEE Trans. Vehicular Technology*, vol. 54, no. 1, pp. 350-365, Jan. 2005.
- [20] T. Alrabiah and A. Aljadhah, "Low-Cost Multicast Routing in Wireless Mobile Networks," *Proc. IEEE Wireless Comm. and Networking Conf. (WCNC '00)*, vol. 3, pp. 1467-1471, 2000.
- [21] G.L. Nemhauser and L.A. Wosley, "Integer and Combinatorial Optimization," *Wiley-Interscience Series in Discrete Math. and Optimization*, 1999.
- [22] B. Fortz and M. Thorup, "Optimizing OSPF/IS-IS Weights in a Changing World," *IEEE J. Selected Areas in Comm.*, vol. 20, no. 4, pp. 756-767, May 2002.
- [23] E. Wong, A. Chan, and T. Yum, "Analysis of Rerouting in Circuit-Switched Networks," *IEEE/ACM Trans. Networking*, vol. 8, no. 3, pp. 419-427, June 2000.
- [24] K.-C. Lee and V.O.K. Li, "A Wavelength Rerouting Algorithm in Wide-Area All-Optical Networks," *IEEE J. Lightwave Technology*, vol. 14, no. 6, pp. 1218-1229, June 1996.
- [25] A. Donner, M. Berioli, and M. Werner, "MPLS-Based Satellite Constellation Networks," *IEEE J. Selected Areas in Comm.*, vol. 22, no. 3, pp. 438-448, Apr. 2004.
- [26] K. Pahlavan et al., "Handoff in Hybrid Mobile Data Networks," *IEEE Personal Comm.*, vol. 7, no. 2, pp. 34-47, Apr. 2000.
- [27] J. McNair, I.F. Akyildiz, and M. Bender, "An Inter-System Handoff Technique for the IMT-2000 System," *Proc. IEEE INFOCOM*, vol. 1, pp. 208-216, 2000.
- [28] J. McNair and F. Zhu, "Vertical Handoffs in Fourth-Generation Multinetwork Environments," *IEEE Wireless Comm.*, vol. 11, no. 3, pp. 8-15, June 2004.
- [29] M.R. Gary and D.S. Johnson, *Computer and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, 1979.
- [30] K.C. Almeroth and M.H. Ammar, "Multicast Group Behavior in the Internet's Multicast Backbone (MBone)," *IEEE Comm. Magazine*, vol. 35, no. 6, pp. 124-129, June 1997.
- [31] *CPLEX Mathematical Programming Optimizer*, <http://www.ilog.com/products/cplex/>, 2008.
- [32] E.W. Zegura, K.L. Calvert, and M.J. Donahoo, "A Quantitative Comparison of Graph-Based Models for Internet Topology," *IEEE/ACM Trans. Networking*, vol. 5, no. 6, pp. 770-783, Dec. 1997.



Ming-Syan Chen received the BS degree in electrical engineering from the National Taiwan University, Taipei, and the MS and PhD degrees in computer, information, and control engineering from the University of Michigan, Ann Arbor, in 1985 and 1988, respectively. He is currently a professor jointly appointed by the Electrical Engineering Department, Computer Science and Information Engineering Department, and Graduate Institute of Communication Engineering (GICE), National Taiwan University. He was a research staff member at IBM T.J. Watson Research Center, Yorktown Heights, New York, from 1988 to 1996. In addition to serving as a program committee member in many conferences, he served as an associate editor of the *IEEE Transactions on Knowledge and Data Engineering* (TKDE) from 1997 to 2001, is currently on the editorial board of the *Very Large Data Base (VLDB) Journal*, *Knowledge and Information Systems*, and *International Journal of Electrical Engineering*, and was a distinguished visitor of the IEEE Computer Society for the Asia-Pacific from 1998 to 2000 and from 2005 to 2007 (invited twice). He served as the international vice chair for INFOCOM 2005, the program chair of the Sixth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2002), a program cochair of the Fourth International Conference on Mobile Data Management (MDM 2003), the program vice chair of the 24th IEEE International Conference on Data Engineering (ICDE 2008), the Eighth IEEE International Conference on e-Commerce Technology/Third IEEE International Conference on Enterprise Computing, e-Commerce and e-Services (CEC/EEE 2006), ICDE 2006, the 25th International Conference on Distributed Computing Systems (ICDCS 2005), the 32nd International Conference on Parallel Processing (ICPP 2003), and the 28th International Conference on Very Large Data Bases (VLDB 2002). He was a keynote speaker on Web data mining at the 1999 International Computer Congress (ICC) and a tutorial speaker on Web data mining at the Sixth International Conference on Database Systems for Advanced Applications (DASFAA 1999) and on parallel databases at ICDE 1995. He was also a guest coeditor of the *IEEE Transactions on Knowledge and Data Engineering* special issue for data mining in December 1996. His research interests include database systems, data mining, mobile computing systems, and multimedia networking. He has published more than 220 papers in his research areas. He is the holder of or has applied for 18 US patents and seven ROC patents in data mining, Web applications, interactive video playout, video server design, and concurrency and coherency control protocols. He is a recipient of the National Science Council (NSC) Distinguished Research Award, the Pan Wen Yuan Distinguished Research Award, the Teco Award, and the K.-T. Li Research Penetration Award for his research work and the Outstanding Innovation Award from IBM for his contribution to a major database product. He has also received numerous awards for his research, teaching, inventions, and patent applications. He is a fellow of the IEEE and ACM.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.



De-Nian Yang received the BS and PhD degrees from the Department of Electrical Engineering, National Taiwan University, Taipei, in 1999 and 2004, respectively. He is currently a postdoctoral researcher for the military service in the Department of Electrical Engineering, National Taiwan University. His research interests include network planning, multicasting, and quality of service (QoS) in wireless networks. He is a member of the IEEE.