An Optimal Distributed Algorithm for Joint Resource Allocation and Routing in Node-based Wireless Networks

Chun Zhang[a], Jim Kurose[a], Yong Liu[b] and Don Towsley[a]

[a]Dept. of Computer Science,
University of Massachusetts,
Amherst, MA 01003
{czhang, kurose, towsley}@cs.umass.edu

[b]Dept. of Electrical & Computer Engineering
Polytechnic University
Brooklyn, NY 11201
yongliu@poly.edu

Abstract

In wireless data networks, the routing of data depends on the link capacities which, in turn, are determined by the allocation of communication resources (such as power, frequency, and time slots) to the links. In this paper, we propose a distributed algorithm for individual node to allocate resource to its incoming and outgoing links and establish routing tables to minimize global cost by generalizing Gallager's result [1]. We focus on so-called *node-based wireless data networks* in which each node has fixed set of resources to be allocated to its incoming and outgoing links, the link capacity is a concave and increasing function of resources allocated to it, and the global cost is the sum of all node costs, each of which is a convex and increasing function of resource allocation variables of the node, and link data rates on the incoming and outgoing links of the node. The algorithm is applied independently at each node: it iteratively updates the local resource allocation based on link data rates, and then updates the routing table based on information communicated between neighbor nodes about the marginal global cost to each destination. The marginal global cost is computed through the sensitivity analysis of the node cost with respect to the link data rates. For stationary input traffic, the global cost converges, with successive updates of the resource allocation and routing tables, to the minimum over all resource allocation and routing assignments. Through simulation over a TDMA wireless data network, we show that the parameters of our proposed algorithm can be tuned to speed up the optimization convergence.

Index Terms

Wireless Data Networks, Resource Allocation, Routing Optimization, Mathematical Programming/Optimization

I. Introduction

In wireless data networks, efficient resource usage (where resources may include power, frequency, and time slots) is of great importance. One way of achieving resource-efficient communication is to move from optimizing routing in isolation to optimally coordinating routing and resource allocation. In this paper, we propose a distributed algorithm by generalizing Gallager's result [1] which only focus on routing optimization problem.

Routing optimization is used to find a set of routes, i.e., the set of paths along which packets are forwarded in order to optimize a well-defined objective function. The routing optimization problem has been well studied [1] [2] in wired networks with fixed link capacities: the network cost (delay, or worst case link utilization) is normally a convex function of link data rates, and the routing optimization problem is formulated and solved as a convex multi-commodity network flow problem. In contrast, link capacities in wireless networks are not necessarily fixed, but instead can be adjusted by the allocation of communication resources (e.g., power, frequency, or time slots) among the various links. A change in resource allocation thus changes the link capacities, and thus impacts the routing decision. Therefore, the resource allocation problem and routing problem are coupled through link capacities, and the overall performance and resource usage of the wireless network can be improved by simultaneous optimization of resource allocation and routing.

Recently, the joint resource allocation and routing optimization problem has been one of the most intensively studied areas. Solution approaches may be roughly classified as being static [3] [4] [5], dynamic [6] [7], or quasi-static. In the static case, the input to the joint optimization problem (such as the node locations, link properties, overall resource, and traffic demand) is fixed. Because the static case is essentially a network design problem, it can be solved off-line, centrally. With dynamic resource allocation and routing, control decisions are made according to the instantaneous states of the networks (such as queues, as well as available resources). A principal challenge here is that conveying (instantaneous) state information to nodes for use in the decision-making process. Quasi-static resource allocation lies between the extremes of static and dynamic cases. Here, the state of the operational data network evolves over time - new source-destination pairs may establish data transmission sessions and old sessions may terminate. Over a longer time scales, nodes or links may fail, new nodes and links may be added. The resource allocation and routing must be changed over the evolving network to satisfy the changing demands. Our interest here will be on *distributed algorithms for quasi-static resource allocation and routing*, i.e., in algorithms in which each node makes its own resource allocation decision and constructs its own routing tables based on periodic updating information from neighbor nodes. The various pros and cons of centralized versus distributed approaches are well known [1] [8].

In this paper, we consider a class of wireless data networks known as *node-based wireless data networks*, in which each node has a fixed set of resources (power and time slot fractions are normally node-level fixed resource) to be allocated to its incoming and outgoing links. Additionally, in *node-based wireless data networks*, link capacity is a concave and increasing function of resources allocated to it. Node resource allocation determines the capacities on all wireless links, upon which nodes route traffic from source to destination for all traffic demands. Our goal is to find an allocation of resources and a feasible set of routes that minimizes the global network cost (such as overall energy consumption rate) among all possible resource allocation and routing combinations.

In a wired network, the network resources are link capacities. Each link incur higher performance cost (delay, and link utilization) as it consumes more resource (capacity) to carry traffic. In contrast, in *node-based wireless data networks*, the network resources are power, frequency, time-slots etc. Each node incurs higher resource cost (such as overall energy consumption rate) as the node consumes more resources in transmitting and receiving data. We characterize a *node cost* by a convex and increasing function of its resource variables, and the data rates on its incoming and outgoing links. The global network cost is the summation of all node costs.

We extend [1], which only focuses on routing optimization problem with fixed link capacities, to take resource allocation into account as well. We achieve the joint optimality in two steps: in the first step, for a set of routes ϕ, we find its associated optimal resource allocation $h^*(\phi)$ which gives lowest global network cost $D^*(\phi)$; in the second step, we find the set of routes with minimum network cost

$$\phi^* = \underset{\phi}{\operatorname{argmin}}\ D^*(\phi)$$

It is thus straightforward to see that the solution to the joint optimization problem is $(h^*(\phi^*), \phi^*)$. In *node-based wireless data networks*, for fixed set of routes, resource allocations on all the nodes are independent with each other. Therefore, the first step can be solved by each node *independently*. However, similar as in wired network, routing decisions on all the nodes are coupled through the network. The second step does require cooperation between nodes. In [1], the routing optimization for wired network is driven by the marginal link costs. For our problem, the routing optimization is driven by the marginal node costs, which can be calculated by nodes through sensitivity analysis of their optimal node costs with respect to the data rates of their incoming and outgoing links.

The framework of our algorithm is shown in Figure 1. The joint resource allocation and routing optimization is achieved through node's collective routing optimization, and independent local resource optimization for current flow. Based on current flow rates, each node independently optimizes its resource allocation, and computes marginal node cost with respect to incoming and outgoing link data rates through sensitivity analysis. The marginal node cost then drives the global routing optimization similar to [1]. We show that for stationary input traffic, our generalized solution converges to optimal resource allocation and routing with successive updates of the node resource allocation decision and routing tables. Through simulation over a TDMA wireless data network, we show that the parameters of our proposed algorithm can be tuned to speed up the optimization convergence.

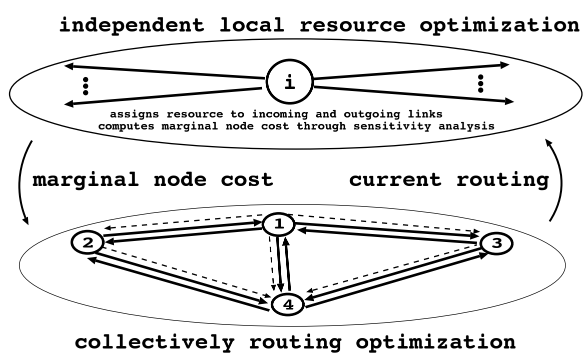

Fig. 1. An algorithm for optimal resource allocation and routing

The paper is organized as follows. Section II describes the routing optimization problem for wired networks with fixed capacities. In section III, we present resource model for the *node-based wireless network* that supports the data network, and demonstrate examples that fit our framework. Based on the network model and resource model, we formalize the joint resource allocation and routing optimization problem in section IV. In section V, for fixed the routing, we solve the resource allocation problem, and calculate the marginal cost with respect to the link data rates. In section VI, we generalize Gallager's distributed algorithm to search for the set of optimal routes. We investigate issues such as faster algorithm convergence through simulation in section VII. In section VIII, we conclude the paper.

II. Routing Optimization with Fixed Capacities

The routing optimization problem with fixed capacities has been well studied in wired networks [1] [2], and the model is rather standard: the topology is represented by a directed graph, the demand is modeled as a multi-commodity flow, and the global cost is the sum of link costs, each of which is a convex and increasing function of link data rate on it.

A. Network Topology

We represent the data network by a directed graph $G = (V, E)$ where V is the set of nodes, and E the set of links. Let a node in V be represented by an integer from $1, 2, \ldots, n$, and a link in E, from node i to node k, be represented by (i, k). For node i, we use $L_I(i)$ to denote the set of links that terminates at it, and $L_U(i)$ the set of links that emanates from it. We call $L(i) = L_I(i) \cup L_U(i)$ as the set of links adjacent to node i. For link (i, j), we use c_{ij} to denote its capacity.

B. Multi-commodity Network Flows

Let $r_i(j)$ be the expected traffic, in $bits/s$, entering the network at node i and destined for node j. Let $t_i(j)$ be the total expected traffic at node i destined for node j. Thus $t_i(j)$ includes both $r_i(j)$ and traffic from other nodes that is routed through i for destination j. Finally let $\phi_{ik}(j)$ be the fraction of the $t_i(j)$ that is routed over link (i, k). Since $t_i(j)$ is the sum of the input traffic and traffic routed to i from other nodes,

$$t_i(j) \quad = \quad r_i(j) + \sum_l t_l(j)\phi_{li}(j) \tag{1}$$

Equation (1) implicitly expresses the conservation of flow at each node: the traffic rate into a node for a given destination is equal to the traffic rate out of the node for that destination. Next, we define ϕ precisely to ensure the equation (1) has a unique solution of t given r and ϕ.

Definition : A routing variable set ϕ for network $G = (V, E)$ is a set of nonnegative numbers $\phi_{ik}(j)$, $i, k, j \in V$, satisfying the following conditions.

1) $\phi_{ik}(j) = 0$ if $i = j$, or $(i, k) \notin E$,
2) $\sum_k \phi_{ik}(j) = 1$ if $i \neq j$,
3) $\forall i, j (i \neq j) \in V$, there is a routing path from i to j, which means there is a sequence of nodes, $i, k, l, \ldots, m, j$ such that $\phi_{ik}(j) > 0$, $\phi_{kl}(j) > 0$, $\ldots$, $\phi_{mj}(j) > 0$.

Theorem 2.1: Let a network $G = (V, E)$ have input set r and routing variable set ϕ. Then the set of equations (1) has a unique solution for t. Each component $t_i(j)$ is nonnegative and continuously differentiable as a function of r and ϕ.

Proof: Proved in [1]. Included in Appendix A for completeness. ■

Now let $f_{ik}(j)$ be the expected traffic rate destined to j , in $bits/s$, on link (i, k), and f_{ik} be the aggregated expected traffic rate, on link (i, k). We have,

$$f_{ik}(j) = t_i(j)\phi_{ik}(j) \tag{2}$$

$$f_{ik} = \sum_j f_{ik}(j) \tag{3}$$

Clearly, a feasible set of flows f must satisfy the capacity constraints.

$$f_{ik} \leq c_{ik}, \quad (i, k) \in E \tag{4}$$

The explicit flow conservation at each node using flow variable set f is given as follows. For $f_{ik}(j) \geq 0$, $(i, k) \in E, j \in V$,

$$\sum_{(i,k) \in L_U(i)} f_{ik}(j) - \sum_{(l,i) \in L_I(i)} f_{li}(j) = r_i(j), \quad i \neq j \tag{5}$$

C. Routing Optimization Problems with Fixed Link Capacities

In case of data network with fixed link capacities, one of the most common cost functions used in the literature is the sum of link costs, each of which is a convex and increasing function of link data rate on it. Let D be the overall cost, and D_{ij} be the cost of link (i, j), we have

$$D = \sum_{(i,j)\in E} D_{ij}(f_{ij}) \tag{6}$$

Using flow variable set f as control variables, the optimization is formalized as a convex optimization problem.

Problem Formulation over f (7)

Given: network $G = (V, E)$, link capacity set c, input demand set r.
Minimize: cost D.
Constraints:
1) Flow conservation. See (5).
2) Capacity feasibility constraints. See (4).

Using routing variable set ϕ as control variables. [1] proposed a distributed algorithm to achieve the same minimized network cost.

Problem Formulation over ϕ (8)

Given: network $G = (V, E)$, link capacity set c, input demand set r.
Minimize: cost D.
Constraints:
1) Route constraints. Demand r is implemented by routing variables set ϕ.
2) Capacity feasibility constraints. See (4).

In this paper, however, we are interested in distributed solution for the case of quasi-static networks for joint resource allocation and routing problem. we are going to generate [1] to take resource allocation into consideration as well.

III. Communication Resource Model and Assumptions

In wireless data networks, the capacities of individual links depend on the media-access scheme and selection of certain critical communication resources (such as power, frequency, or time-slot fractions) allocated to individual links. We refer to these critical communication resources collectively as *resource variables*. We assume that the media-access scheme is fixed, but the resource allocation can be adjusted at our will. The resource variables are themselves limited by various resource constraints (such as limits on the total transmit power at each node). Next, we first introduce the communication resource model in *node-based wireless networks*. Then, we describe three examples that fit in this communication resource model.

A. Communication Resource Model of Node-based Wireless Networks

In *node-based wireless networks*, the communication resource model has three key properties.

Fixed resources on nodes: Each node i has fixed set of communication resources denoted by u_i. In an operational network, power and time slot (fractions) are normally node-level fixed resources. Frequency allocation can be dynamic. However, in this paper, we will focus on the case where communication resources are fixed at each node.

Resource allocation to adjacent links: Each node can only allocate its set of resources to its incoming and outgoing links. This translates to that a node only consumes its resources if it transmits or receives data. Let $h_{kl}(i)$, $(k, l) \in L(i)$ be a vector of communication resources allocated to adjacent link (k, l) by node i. The resource feasibility constraint states that the total amount of resources allocated by node i must be less than or equal to its fixed set of resources u_i, we have

$$\sum_{(k,l)\in L(i)} h_{kl}(i) \quad \leq \quad u_i \tag{9}$$

Concave capacity of allocated resources: The sending capacity of link (k, l) is a concave and increasing function of set of resources allocated to it by sender node k, and the receiving capacity of link (k, l) is a concave and increasing function of set of resources allocated to it by receiver l. The capacity of link (k, l) is limited by its sending capacity and receiving capacity. Let κ^s_{kl} and κ^r_{kl} denote the sending capacity and receiving capacity of link (k, l). We have,

$$c_{kl} \quad = \quad \min(\kappa^s_{kl}(h_{kl}(k)), \kappa^r_{kl}(h_{kl}(l))) \tag{10}$$

where κ^s_{kl} and κ^r_{kl} are concave and increasing functions of allocated set of resources.

In general, the capacity c_{ik} depends not only on its own allocated set of resources, but also on set of communication resources allocated to other links in the network (due to interference). However, in this paper, we will focus on the interference free case. Certain joint resource allocation and routing optimization problems (including three examples we show next) directly fits the interference free model. Furthermore, some problems with more complex interference considerations (such as problems in [9]) was reduced and solved as an interference free problem.

B. Examples of the Resource Model of Node-Based Wireless Networks

Certain joint resource allocation and routing optimization problems directly fits the resource model of *node-based wireless networks*. Here, we will only illustrate how the Gaussian channels with FDMA, and Gaussian channels with TDMA using omni-directional/directional antennas, fit into this framework.

1) Gaussian Channel With FDMA: In the Gaussian channel with FDMA, each node i is preassigned disjoint frequency W_i, and power P_i. For simplicity, we assume that for data transmission over link (i, j), only the sending node i is required to allocate set of resources (power $P_{ij}(i)$, and frequency $W_{ij}(i)$) for communication to take place. The received power at node j is $\sigma_{ij} P_{ij}(i)$ where σ_{ij} is the path loss function on link (i, j). The receiving node j is also subject to independent additive white Gaussian noises (AWGNs) with power spectral densities N_0^j. Based on the Shannon capacity formula, the capacity of link (i, j) is a concave and increasing function of the resource variables $(P_{ij}(i), W_{ij}(i))$.

$$c_{ij} = W_{ij} \log_2\left(1 + \frac{\sigma_{ij} P_{ij}(i)}{N_0^j W_{ij}(i)}\right) \tag{11}$$

The communication variables at node $i \in V$ are constrained by its overall resources limits,

$$\sum_{(i,j)\in L_U(i)} P_{ij}(i) \le P_i \qquad \sum_{(i,j)\in L_U(i)} W_{ij}(i) \le W_i \tag{12}$$

2) Gaussian Channel with TDMA using Omni-Directional Antennas: In the Gaussian channel with TDMA using Omni-directional antennas, node i is preassigned disjoint frequency W_i, and power P_i. In the TDMA case, the resource variables are time-slots rather than frequency or power. Similarly to the FDMA case, we assume that for data transmission over link (i, j), only the sending node i is required to allocate resources (time slots) to link (i, j). This means that a node can parallelly receive packets from multiple neighbors at any time slot, but it can only send packet to one neighbor node at each time slot. Let $\tau_{ij}(i)$ be the time-slot fraction allocated to link (i, j) by node i. The achieved capacity c_{ij} is a linear (hence, convex) function of $\tau_{ij}(i)$,

$$c_{ij} = \tau_{ij}(i) W_i \log_2\left(1 + \frac{\sigma_{ij} P_i}{N_0^j W_i}\right) \tag{13}$$

Each node overall utilization for transmitting can not exceed 100%,

$$\sum_{(i,j)\in L_U(i)} \tau_{ij}(i) \le 1 \tag{14}$$

3) Gaussian Channel with TDMA using Directional Antennas: In the case of communication over a network in which each node is equipped with a single directional antenna, a node can only send packets or receive packet at any time slot: to send packets through link (i, j), both the sending node i and receiving node j are required to allocate resource (time slot). Let P_i be the power at node i, and W be the network wide frequency. Let $\tau_{ij}(i)$ be the time-slot allocated to link (i, j) by the sending node i, and $\tau_{ij}(j)$ the time-slot allocated to link (i, j) by the receiving node j. Assuming that directional antennas eliminate interference between any pair of links, the achieved capacity c_{ij} is then a concave and increasing function of $\tau_{ij}(i)$ and $\tau_{ij}(j)$,

$$c_{ij} = \min(\kappa^s_{ij}(\tau_{ij}(i)), \kappa^r_{ij}(\tau_{ij}(j))) \tag{15}$$

$$= \min(\tau_{ij}(i), \tau_{ij}(j)) W \log_2\left(1 + \frac{\sigma_{ij} P_i}{N_0^j W}\right) \tag{16}$$

At node $i \in V$, the overall node utilization for both transmitting and receiving can not exceed 100%. i.e.,

$$\sum_{(i,j)\in L_U(i)} \tau_{ij}(i) + \sum_{(i,j)\in L_I(i)} \tau_{ij}(i) \le 1 \tag{17}$$

However, as shown by [10], the above condition does not guarantee a feasible solution. Instead, a sufficient condition is given in [10] as follows.

$$\sum_{(i,j)\in L_U(i)} \tau_{ij}(i) + \sum_{(i,j)\in L_I(i)} \tau_{ij}(i) \leq 2/3 \tag{18}$$

We thus take $u_i = 2/3$ as fixed node resource (time-slot) for this case.

IV. Joint Optimization on Resource Allocation and Routing

A model for joint optimization on resource allocation and routing over the *node-based wireless data networks* can be derived by combining the network flow model and the resource model described in the previous two sections. This model reflects how the link capacities depend on the allocation of communications resources, and how the overall optimal network cost be achieved by joint optimization on resource allocation and routing.

A. Network Cost in Node-based Wireless Networks

In *node-based wireless networks*, the overall network cost is the sum of all node costs, each of which is a convex and increasing function of resource variables of the node, and link data rates on the incoming and outgoing links of the node. The node cost reflects the cost at the node level, while the network cost reflects the overall cost. Let D_i be the cost of node i, we have,

$$D_i = D_i^f(\{f_{kl}|(k,l) \in L(i)\}) + D_i^h(\{h_{kl}(i)|(k,l) \in L(i)\}) \tag{19}$$

where D_i^f is a convex and increasing function of link data rates of incoming and outgoing links of node i, and D_i^h a convex and increasing function of resource variables of the node i.

Let D denote the overall network cost, we have

$$D = \sum_{i\in V} D_i \tag{20}$$

Note that D is a convex and increasing function of flow rate set f and resource variables h.

B. Joint Resource Allocation and Routing Optimization Problem

In *node-based wireless networks*, the joint resource allocation and routing problem can be formulated as a convex optimization problem using resource variables h and flow variables f as control variables. Let the network cost $D^{h,f}$ be a function of resource variable set h and flow variable set f, we formalize the joint optimization problem as follows.

Problem Formulation over h, f (21)

Given: network $G = (V, E)$, node resource set u, input demand set r.
Minimize: $D^{h,f}$
Constraints:
1) Flow conservation. See (5).
2) Capacity feasibility constraints. See (4).
3) Resource feasibility constraints. See (9).

Since $D^{h,f}$ is a convex function over h and f, and the constraints define a convex set over h and f. The joint resource allocation and routing optimization problem is a convex optimization problem. This implies that it can be solved globally and efficiently using centralized algorithms from convex optimization literature [11] [3]. However, in this paper, we are interested in a distributed solution to the above problem. Therefore, similar to [1], we use routing variables set ϕ (rather than flow rate variables f) and resource variable set h as control variables in formalizing the problem. Let the network cost $D^{h,\phi}$ be a function of resource variable set h and flow variable set ϕ, we have,

Problem Formulation over h, ϕ (22)

Given: network $G = (V, E)$, node resource set u, input demand set r.
Minimize: cost $D^{h,\phi}$.
Constraints:
1) Route constraints. Demand r is implemented by routing variables ϕ.
2) Capacity feasibility constraints. See (4).

3) Resource feasibility constraints. See (9).

We use a node-based penalty function to implicitly satisfy the resource feasibility constraints. Let b_i denote a vector of allocated resources of node i. The penalty function at node i, denoted by D_i^P, is an increasing function of allocated resources b_i, and approaches infinity when any kind of allocated resource m of node i, $b_i(m)$, goes into the fixed amount resource m of node i, $u_i(m)$.

$$\forall m, \quad \lim_{b_i(m)\to u_i(m)} D_i^P = \infty \tag{23}$$

The overall penalty function D^P is the sum of all node penalty functions.

$$D^P = \sum_{i\in V} D_i^P \tag{24}$$

Let A_i be the objective function at node i, and A the overall objective function. We have,

$$A_i = D_i + D_i^P \tag{25}$$

$$A = \sum_{i\in V} A_i = D + D^P \tag{26}$$

Similar to the representation of D using $D^{h,f}$ and $D^{h,\phi}$, we use $A^{h,f}$ to denote A a function of h and f, and $A^{h,\phi}$ a function of h and ϕ.

Adding penalty function D^P may deviate from optimizing the original cost function D. However, in *node-based wireless networks*, the node-based penalty function may prevent a node resource from being completely depleted in favor of global cost. The remaining resource may be used to better accommodate the changing demand, or be used for fast recovery in the case of node or link failures.

C. An Example of the Joint Resource Allocation and Routing Problem

Now we give an example of joint resource allocation and routing problem. In wireless data networks where nodes are energy constrained, minimizing the energy consumption (power) is very important. One way to achieve energy-aware routing is to minimize the overall energy consumption rate through joint resource allocation and routing.

The cost at node i, D_i, is the energy consumption rate at node i. For simplicity, we assume that the energy is only consumed at the transmission side. Therefore,

$$D_i = \sum_{(i,j)\in L_U(i)} \tau_{ij}(i) P_{ij}(i) \tag{27}$$

where $\tau_{ij}(i)$ and $P_{ij}(i)$ are time-slots and power allocated to link (i,j) by node i.

The penalty at node i, D_i^P, is a function of allocated resources b_i. For the three examples we given in section III, we choose the penalty function as follows.

Gaussian Channel with FDMA

$$D_i^P = \frac{\alpha_i^p}{P_i - \sum_{(i,j)\in L_U(i)} P_{ij}} + \frac{\alpha_i^w}{W_i - \sum_{(i,j)\in L_U(i)} W_{ij}} \tag{28}$$

Gaussian Channel with TDMA (Omni-directional Antenna)

$$D_i^P = \frac{\alpha_i^\tau}{1 - \sum_{(i,j)\in L_U(i)} \tau_{ij}(i)} \tag{29}$$

Gaussian Channel with TDMA (Directional Antenna)

$$D_i^P = \frac{\alpha_i^\tau}{\frac{2}{3} - \sum_{(i,j)\in L_U(i)} \tau_{ij}(i) - \sum_{(j,i)\in L_I(i)} \tau_{ji}(i)} \tag{30}$$

where α_i^p, α_i^w, and α_i^τ are positive parameters to adjust the degree of punishment.

V. RESOURCE ALLOCATION FOR FIXED ROUTING

Now we solve the joint resource allocation and routing optimization problem for *node-based wireless networks*. In this section, we solve this optimization problem in the case of fixed routing (thus fixed data flow). With fixed set of routes, we are able to solve the global resource allocation problem via independent local resource optimization at each node. Furthermore, through sensitivity analysis, each node locally calculates the marginal node cost with respect to link data rates. As we shall see later, the marginal node cost serves a key role in searching for the optimal set of routes.

A. Node-Level Resource Optimization Problem

In *node-based wireless networks*, the node costs among individual nodes are coupled together through routing variables. However, with fixed set of routes, resource allocations on all nodes are independent with each other. Therefore, the global optimal resource allocation can be achieved by local resource optimization at individual nodes. Assuming that with fixed route set ϕ, each node can always estimate the flow rate set f on incoming and outgoing links, each node then solves the following convex resource allocation problem.

Problem Formulation over h at node i (31)

Given: resource vector u_i, flow rates $f_{kl}, (k,l) \in L(i)$.
Minimize: $A_i^{h,f}$ (or $A_i^{h,\phi}$)
Constraints:
1) Capacity feasibility constraints. See (4).

For given set of routes ϕ and its resulting flow rate f, let $A_i^f(f)$ (or $A_i^\phi(\phi)$) denote the optimal $A_i^{h,f}$ (or $A_i^{h,\phi}$) at node i. Summing up the optimization result at all nodes, we get globally minimized objective function $A^f(f)$ (or $A^\phi(\phi)$).

$$A^f(f) = \sum_{i \in V} A_i^f(f), \quad A^\phi(\phi) = \sum_{i \in V} A_i^\phi(\phi) \tag{32}$$

As we mentioned earlier, the optimization problem of $A^{h,f}$ is a convex optimization problem. From [11], we know that the optimization problem over $A^f(f)$ is also a convex optimization problem. More specifically, $A^f(f)$ is a convex and increasing function over a convex set of flow sets satisfying the capacity feasibility constraints and the resource feasibility constraints. This convex set of flow sets is a closure of $|E|$ planes (each of which corresponds to $f_{ij} = 0$, $(i,j) \in E$), and an envelope imposed by resource constraints. The shape of the envelope depends on the link capacity functions as a concave and increasing function of resources allocated to them, and fixed set of resources of all nodes. [1] We use F to denote the convex set of flow sets satisfying the capacity feasibility constraints and the resource feasibility constraints, and F_∞ the set of flow sets on the envelope imposed by resource constraints. Due to the punishment of penalty function D^P, the flow sets f_∞ on the resource boundary F_∞ satisfies $\lim_{f \to f_\infty^-} A^f(f) = \infty$. (i.e.,$f_\infty$ must be implemented by fully consuming at least a kind of resource at certain node). The penalty function D^P keeps f from approaching envelope F_∞ when $A^f(f)$ is minimized.

B. Node Sensitivity Analysis

While solving the local resource optimization problem for the fixed flow (31), each node i also locally calculates the marginal node cost with respect to the link data rates $\partial A_i^f(f)/\partial f_{kl}$, $(k,l) \in L(i)$ through sensitivity analysis [11].

From (10), we see that for link $(k,l) \in E$, c_{kl} is determined by resource allocated by both sender k and receiver l. An increase of data flow f_{kl} requires the increase of capacity c_{kl}. This, in turn, requires that both sender k and receiver l to allocate additional resources to link (k,l), incurring higher cost A_k^f and A_l^f. Therefore, the marginal global cost with respect to the link data rate of link (k,l), $\partial A^f(f)/\partial f_{kl}$ is calculated as the sum of marginal node cost over two end nodes k and l.

$$\frac{\partial A^f(f)}{\partial f_{kl}} = \frac{\partial A_k^f(f)}{\partial f_{kl}} + \frac{\partial A_l^f(f)}{\partial f_{kl}} \tag{33}$$

Note that the marginal global cost $\partial A^f(f)/\partial f_{kl}$ can be derived through local communication between end nodes k and l.

[1] In case of wired network with fixed capacities, the envelope is given by E planes (which correspond to $f_{ij} = c_{ij}$, $(i,j) \in E$).

VI. Searching for Optimal Routing

In previous section, for given fixed set of routes, nodes can achieve the optimal resource allocation through independent node-level resource optimization, and calculate the marginal global cost through local sensitivity analysis and communication between neighbor nodes. Now, we focus on routing optimization. i.e., an algorithm adjusts routing variables to converge to the optimal set of routes. Next, we propose our solution by generalizing Gallager's result [1]. We first generalize [1]'s necessary and sufficient condition for optimal set of routes. Then we generalize [1]'s distributed algorithm.

A. Necessary and Sufficient Conditions for Optimal Cost

Now we generalize [1]'s necessary and sufficient conditions to minimize A^ϕ over all feasible sets of routes. Similar to [1], we compute the partial derivatives of A^ϕ with respect to the inputs r and the routing variables ϕ as follows.

$$\frac{\partial A^\phi(\phi)}{\partial r_i(j)} = \sum_k \phi_{ik}(j)\left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)}\right] \tag{34}$$

$$\frac{\partial A^\phi(\phi)}{\partial \phi_{ik}(j)} = t_i(j)\left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)}\right] \tag{35}$$

The existence and uniqueness of $\partial A^\phi(\phi)/\partial r_i(j)$ and $\partial A^\phi(\phi)/\partial \phi_{ik}(j)$ is given by the following theorem.

Theorem 6.1: Let a network have input set r and routing variable set ϕ, and let each marginal link cost $\frac{\partial A^f(f)}{\partial f_{ik}}$ be continuous in $f_{ik}, (i,k) \in E$. Then the set of equations (34) has a unique set of solutions for $\frac{\partial A^\phi(\phi)}{\partial r_i(j)}$. Furthermore, (35) is valid and both $\frac{\partial A^\phi(\phi)}{\partial r_i(j)}$ and $\frac{\partial A^\phi(\phi)}{\partial \phi_{ik}(j)}$ for $i \neq j$, $(i,k) \in E$ are continuous in r and ϕ.

Proof: See Appendix B. ■

Using Lagrange multipliers for the constraint $\sum_k \phi_{ik}(j) = 1$, and taking into account the constraint $\phi_{ik}(j) \geq 0$, the necessary conditions for a minimum of A^ϕ with respect to ϕ are, $\forall i \neq j$, $(i,k) \in E$,

$$\frac{\partial A^\phi(\phi)}{\partial \phi_{ik}(j)} \begin{cases} = \lambda_{ij} & \phi_{ik}(j) > 0 \\ \geq \lambda_{ij} & \phi_{ik}(j) = 0. \end{cases} \tag{36}$$

This states that for given i, j, all links (i,k) for which $\phi_{ik}(j) > 0$ must have the same marginal cost $\partial A^\phi(\phi)/\partial \phi_{ik}(j)$, and that this marginal cost must be less than or equal to $\partial A^\phi(\phi)/\partial \phi_{ik}(j)$ for the links on which $\phi_{ik}(j) = 0$. However, as shown by [1], (36) is not a sufficient condition to minimize A^ϕ even for the routing optimization problem in wired networks.

Given i, j in (35), if $t_i(j) = 0$, then $\forall k$, we have $\partial A^\phi(\phi)/\partial \phi_{ik}(j) = 0$. This means that, if node i is not on any route carrying the traffic destined to j, the above necessary conditions would be automatically satisfied. Thus, we hypothesize that (36) would be sufficient to minimize A^ϕ if the factor $t_i(j)$ were removed from the condition.

Theorem 6.2: Let F be a convex and compact set of flow sets, which is enclosed by $|E|$ planes (each of which corresponds to $f_{ij} = 0$, $(i,j) \in E$), and a boundary envelope F_∞. Assume that A^f is convex and continuously differentiable for $f \in F - F_\infty$, Let Ψ be the set of ϕ for which the resulting set of flow rates f are in the above convex and bounded set $F - F_\infty$. Then (36) is necessary for ϕ to minimize A^ϕ over Ψ and (37), for all $i \neq j$, $(i,k) \in E$, is sufficient.

$$\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \geq \frac{\partial A^\phi(\phi)}{\partial r_i(j)} \tag{37}$$

Proof: See Appendix C. ■

B. A Distributed Algorithm for Routing Optimization

Based on the above sufficient condition, we now develop a gradient-based algorithm by generalizing [1]. Each node i must incrementally decrease those routing variables $\phi_{ik}(j)$ for which the marginal cost $\partial A^f(f)/\partial f_{ik} + \partial A^\phi(\phi)/\partial r_k$ is large, and increase those for which it is small. The algorithm breaks into three parts: a protocol between nodes to calculate the marginal costs, an algorithm for calculating the routing updates and modifying the routing variables, and a protocol for forecasting the flow rates of next iteration and allocating resources to support them. We discuss the protocol to calculate the marginal costs first.

Assume that each node i can estimate the link traffic rate f_{kl} for each incoming and outgoing link $(k,l) \in L(i)$. Based on this information, node i calculates $\partial A_i^f(f)/\partial f_{kl}, (k,l) \in L(i)$ through sensitivity analysis. Then, for each pair of nodes i, j with common link (i,j), node j sends $\partial A_j^f(f)/\partial f_{ij}$ to node i. Upon receiving it from node j, node i computes $\partial A^f(f)/\partial f_{ij}$ using (33).

In order to see how node i can calculate $\partial A^\phi(\phi)/\partial r_i(j)$. Define node m to be downstream from node i (with respect to destination j) if there is a routing path from i to j passing through m (i.e., a path with positive routing variables on each link). Similarly, we define i as upstream from m if m is downstream from i. A routing variable set ϕ is loop free if for each destination j, there is no i, $m(i \neq m)$ such that i is both upstream and downstream for m. The protocol used for an update, now, is as follows: for each destination node j, each node i waits until it has received the value $\partial A^\phi(\phi)/\partial r_k(j)$ from each of its downstream neighbors $k \neq j$. The node i then calculates $\partial A^\phi(\phi)/\partial r_i(j)$ from (34) (using the convention that $\partial A^\phi(\phi)/\partial r_j(j) = 0$) and broadcasts this to all of its neighbors. It is easy to see that this procedure is free of deadlocks if and only if ϕ is loop free.

We shall later define small but important detail that has been omitted so far in the updating protocol between nodes: a small amount of additional information is necessary for the algorithm to maintain loop freedom. It turns out to be necessary, for each destination j and each node i, to specify a set $B_i(j)$ of blocked node k for which $\phi_{ik}(j) = 0$ and the algorithm is not permitted to increase $\phi_{ik}(j)$ from 0. We first define and discuss the algorithm and then define the sets $B_i(j)$.

The algorithm Γ, on each iteration, maps the current routing variable ϕ into a new set $\phi^1 = \Gamma(\phi)$. The mapping is defined as follows. For $k \in B_i(j)$,

$$\phi^1_{ik}(j) = 0, \quad \Delta_{ik}(j) = 0. \tag{38}$$

For $k \notin B_i(j)$, define

$$a_{ik}(j) = \frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} - \min_{m \notin B_i(j)} \left[\frac{\partial A^f(f)}{\partial f_{im}} + \frac{\partial A^\phi(\phi)}{\partial r_m(j)} \right] \tag{39}$$

$$\Delta_{ik}(j) = \min[\phi_{ik}(j), \eta a_{ik}(j)/t_i(j)] \tag{40}$$

where η is a scale parameter of Γ to be discussed later. Let $k_{min}(i,j)$ be a value of m that achieves the minimization in (40). Then

$$\phi^1_{ik}(j) = \begin{cases} \phi_{ik}(j) - \Delta_{ik}(j) & k \neq k_{min}(i,j) \\ \phi_{ik}(j) + \sum_{k \neq k_{min}(i,j)} \Delta_{ik}(j) & k = k_{min}(i,j). \end{cases} \tag{41}$$

The algorithm reduces the fraction of traffic sent on non-optimal links and increases the fraction on the best link. The amount of reduction, given by $\Delta_{ik}(j)$, is proportional to $\alpha_{ik}(j)$, with the restriction that $\phi^1_{ik}(j)$ cannot be negative. In turn $\alpha_{ik}(j)$ is the difference between the marginal cost to node j using link (i,k) and using the best link. Note that as the sufficient condition (37) is approached, the changes get small, as desired. The amount of reduction is also inversely proportional to $t_i(j)$. The reason for this is that the change in link traffic is related to $\Delta_{ik}(j)t_i(j)$. Thus when $t_i(j)$ is small, $\Delta_{ik}(j)$ can be changed by a large amount without greatly affecting the marginal cost. Finally the changes depend on the scale factor η. For η very small, convergence of the algorithm is guaranteed, as shown in Theorem 6.3, but rather slow. As η increases, the speed of convergences increases but the danger of no convergence increases.

We now complete the definition of algorithm Γ by defining the block sets $B_i(j)$. See [1] for further reasoning on how this definition guarantees the loop free properties.

Definition: The set $B_k(j)$ is the set of nodes k for which both $\phi_{ik}(j) = 0$ and k is blocked relative to destination j. A node k is blocked relative to j if k has a routing path to j containing some link (l,m) for which $\phi_{lm}(j) > 0$, and $\partial A^\phi(\phi)/\partial r_l(j) \leq \partial A^\phi(\phi)/\partial r_m(j)$, and

$$\phi_{lm}(j) \geq \eta \left[\frac{\partial A^f(f)}{\partial f_{lm}} + \frac{\partial A^\phi(\phi)}{\partial r_m(j)} - \frac{\partial A^\phi(\phi)}{\partial r_l(j)} \right] \Big/ t_i(j) \tag{42}$$

The protocol required for a node i to determine the set $B_i(j)$ is as follows. Each node l, when it calculates $\partial A^\phi(\phi)/\partial r_l(j)$, determines, for each downstream m, if $\phi_{lm}(j) > 0$, and $\partial A^\phi(\phi)/\partial r_l(j) \leq \partial A^\phi(\phi)/\partial r_m(j)$, and satisfy (42). If any downstream neighbor satisfies these conditions, node l adds a special tag to its broadcast of $\partial A^\phi(\phi)/\partial r_l(j)$. The node l also adds the special tag if the received value $\partial A^\phi(\phi)/\partial r_m(j)$ from any downstream m contained a tag. In this way all nodes upstream of l also send the tag. The set $B_i(j)$ is then the set of nodes k for which the received $\partial A^\phi(\phi)/\partial r_k(j)$ was tagged.

Theorem 6.3: Let F be a convex and compact set of flow sets, which is enclosed by $|E|$ planes (each of which corresponds to $f_{ij} = 0$, $(i,j) \in E$), and a boundary envelope F_∞. Assume that A^f is a convex and increasing function for $f \in F - F_\infty$ and that $\forall f_\infty \in F_\infty$, $\lim_{f \to f_\infty^-} A^f = \infty$. For every positive number A_0, if ϕ^0 satisfies $A^\phi(\phi^0) \leq A_0$, then with scale factor $\eta = [Mn^8]^{-1}$,

$$\lim_{m\to\infty} \Gamma(\phi^m) = \min_{\phi}(A^\phi(\phi)) \tag{43}$$

where

$$M = \max_{(l_1,m_1),(l_2,m_2)\in E} \max_{f:A^f(f)\leq A_0} \frac{\partial^2 A^f}{\partial f_{l_1 m_1(f)} \partial f_{l_2 m_2}}(f) \tag{44}$$

Proof: See Appendix D. ■

.

Note that η depends on some upper bound A_0 to A^ϕ; this is natural, since when the link data rates f are very close to the capacities nodes can provide, small changes in the link data rates cause large changes in the cost. The proof uses a ridiculously small value of η to guarantee convergence under all conditions. In next section, we simulate over a TDMA network to identify practical values for η.

Finally, we discuss the protocol for forecasting the flow rates of next iteration and allocating resources to support the updated traffic. Assume that each node i can estimate the demand rate set $r_i(j)$ entering from i. First, for each destination node j, each node i signals the downstream nodes under ϕ^1 (which is the set of routes for next iteration) so that each node k gets a list of of upstream nodes under ϕ^1. Second, for each destination node j, each node i waits until it has received the forecasted value $f^1_{li}(j)$ from each of its upstream node l under ϕ^1. For each downstream node k under ϕ^1, each node i then calculates $f^1_{ik}(j)$ from (1)(2) and sends it to k. Each node i also calculates forecasted f^1_{kl}, $(k,l) \in L(i)$ from (3). Based on the forecasted link data rates of incoming and outgoing links f^1, nodes allocate and optimize resource allocation as we have discussed in section V-A.

We have proposed a distributed algorithm for routing optimization. Note that in each iteration, the resource allocation is also optimized through local independent resource optimization at all nodes. Combining the collective routing optimization, and independent local resource optimization at all nodes, we have achieved the optimal cost over all feasible resource allocation and routing combinations.

VII. Numerical Examples

In this section, we present numerical results of the proposed joint resource and routing optimization problem on a synthetic wireless network. We illustrate through this particular example how the choice of step-size scale factor η affect convergence speed. It will become clear that, in practice, it is possible to choose a η much larger than the value used in the proof of Theorem 6.3 to expedite the convergence. Furthermore, we will discuss the impact of the added penalty function on the original optimization problem.

A. Simulation over TMDA Networks with Directional Antennas

We implemented the proposed algorithm to minimize the overall energy consumption rate in a TDMA wireless network where nodes are equipped with direction antennas. TDMA wireless networks with direction antennas have been introduced in section III-B.3, and the joint resource allocation and routing problem to minimize the overall energy consumption rate has been formulated in section IV-C.

Figure 2 illustrates a synthetic network with 30 nodes and 254 links (127 bi-directed links). We place nodes uniformly on the unit square $[0,1] \times [0,1]$. Two nodes i, j can communicate to each other if the distance between them y_{ij} is within the threshold 0.35. For data communication over link (i,j), the path loss is $\sigma_{ij} = (y_0/y_{ij})^2$, where $y_0 = \min_{(k,l)\in E} y_{kl}$ is the reference distance, and the receiver j is subject to noise with power spectral density $N_0 = 0.1$. The network has frequency resource $W = 1$, and each node has power limit $P_i = 100$. Each node have fixed resources (time slots $u_i = 2/3$) to be allocated to its incoming and outgoing links. Let $\tau_{ij}(i)$ and $\tau_{ij}(j)$ be time-slots assigned to link (i,j) by node i and j respectively, from (16), the capacity of link (i,j) is,

$$c_{ij} = \min\left(\tau_{ij}(i), \tau_{ij}(j)\right) \log_2 \left(1 + 1000\left(\frac{y_0}{y_{ij}}\right)^2\right) \tag{45}$$

We generate traffic demand set r by randomly picking 32 source and destination pairs. The traffic rate between a source and destination pair is uniformly distributed within $[0.18, 0.55]$.

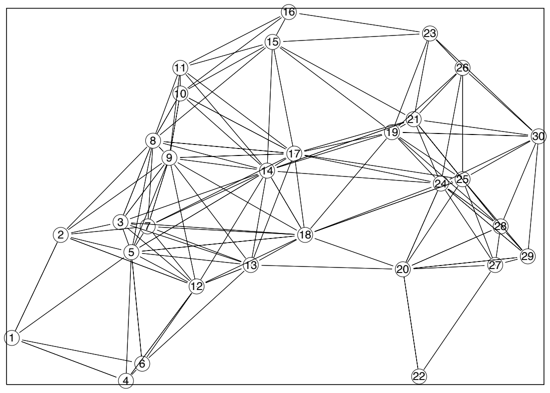

Fig. 2. Topology of a uniformly generated wireless network with 30 nodes and 127 bidirectional links

From (27), the energy consumption rate at node i is,

$$D_i = \sum_{(i,j)\in L(U)} 100\tau_{ij}(i).$$

Our goal is to minimize the overall network energy consumption rate ($D = \sum_{i\in V} D_i$). We use the penalty functions defined in (30), repeated as follows:

$$\begin{aligned} D_i^P &= \frac{\alpha_i^\tau}{\frac{2}{3} - \sum_{(i,j)\in L_U(i)} \tau_{ij}(i) - \sum_{(j,i)\in L_I(i)} \tau_{ji}(i)} \\ D^P &= \sum_{i\in V} D_i^P, \end{aligned}$$

where α_i^τ is a weight to adjust the degree of punishment on node i.

The initial set of resource allocation and routes are chosen using a baseline algorithm. Each node i fairly allocates all of its own set of resources to its incoming and outgoing links so that the resulting sending capacity for all outgoing links, and the receiving capacity for all incoming links are equal. i.e., $\forall(k,l) \in L(i)$,

$$\tau_{kl}(i)\log_2\left(1 + 1000\left(\frac{y_0}{y_{kl}}\right)^2\right) = \mathcal{C}_i, \tag{46}$$

where $\mathcal{C}_i$ is the fairly allocated capacity by node i to its all incoming and outgoing links. The resulting capacity of link (i,j) is thus

$$c_{ij} = \min\{\mathcal{C}_i, \mathcal{C}_j\}. \tag{47}$$

The initial set of routes is then computed using a distributed protocol like Open Shortest Path First (OSPF) [12] with the inverse of the capacities computed above as link weights.

B. Scale Factor η and Convergence

In the previous section, with a small scale factor η, we have shown that optimization algorithm Γ must eventually converge to the optimum for a network with stationary inputs, nodes, and node resources. The question of whether the algorithm can adapt fast enough to keep up with changing statistics is not obvious and deserves more study. Faster adaption may be achieved through more frequent optimization protocol, or using larger scale factor η. However, the frequent optimization protocols reduce the resource available for data transmission, and may reduce the accuracy of link data rate estimation. A large scale factor η may break the convergence of the algorithm. In this section, we numerically compare how the proposed algorithm converges to the optimum with different scale factors η.

With fixed penalty functions $\alpha_i^\tau = 10^{-4}$, $i \in V$, we choose four different scale factors $\eta = \{10^{-3}, 10^{-5}, 10^{-6}, 10^{-7}\}$ to run the algorithm, that minimizes the objective function $A = D + D^P$. Note that all four choices of η are much larger than the value given in Theorem 6.3. When $\eta = 10^{-3}$, the algorithm generates infeasible flow, and then fails the optimization. When η is chosen from $\{10^{-5}, 10^{-6}, 10^{-7}\}$, the algorithm does converge to the optimum as shown in Figure 3. We see that the optimization reduces the objective function A from the baseline 338.31 to 269.09, an improvement around 20% from the baseline solution. We also see that the convergence of the algorithm depends to a great extent on the value of η: a larger value of η may lead to faster convergence. From the figure we see that for $\eta = 10^{-5}, 10^{-6}, 10^{-7}$, the algorithm takes roughly 400, 4000 and 40000 iterations to reduce A within 5% of the optimal value (282.44). We also conducted simulations with other parameter settings, and get similar results.

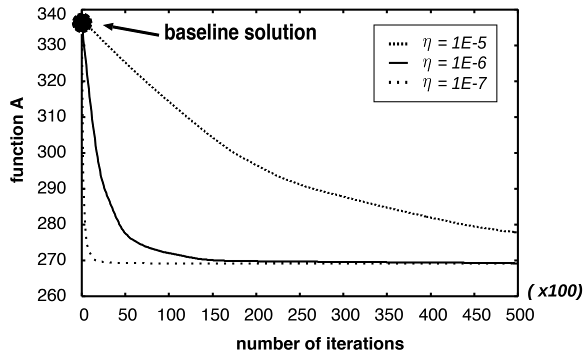

Fig. 3. Comparisons of convergence speed for different ηs

C. Penalty Functions and Optimization Result

In this paper, we use penalty function to enforce the resource feasibility constraints. The added penalty function in the objective function will deviate obtained solution from the real optimum of the original optimization problem without the penalty function. To examine the impact of the penalty function on the original optimal goal, we vary the degree of punishment by choosing α_i^τ from $\{10^{-4}, 10^{-2}, 1\}$.

Table I shows the optimization result for different values of α^τ as we set $\alpha_i^\tau = \alpha^\tau$, $i \in V$. As we expected, the larger α^τ, the more objective function A deviates from the original function D. When $\alpha^\tau = 10^{-4}$, the fraction of the penalty in the overall objective function is $D^P/A = 4 \times 10^{-5}$. The negligible penalty results in a solution D very close to the optimum, which is 69.07 less than the baseline case. In contrast, when $\alpha^\tau = 1$, the fraction of the penalty in the overall objective function is $D^P/A = 0.2$. In this case, the degree of punishment is not negligible. Consequently, D incurs higher cost, and the gain is reduced from 69.07 to 49.28 (a loss of gain by 29%). The result of $\alpha^\tau = 10^{-2}$ stays between the previous two cases. The fraction of the penalty in the overall objective function is $D^P/A = 3 \times 10^{-3}$. As the degree of punishment is still small, D is closer to the case of $\alpha^\tau = 10^{-4}$.

α^τ	$A = D + D^P$	D	D^P/A	$D - D(baseline)$
10^{-4}	269.09	269.08	4×10^{-5}	69.07
10^{-2}	271.61	270.90	3×10^{-3}	67.25
1	354.23	288.87	2×10^{-1}	49.28

TABLE I
OPTIMIZATION RESULT WITH DIFFERENT α^τS

We have seen how the degree of penalty influences the solution of the original problem. On the other hand, the node-based penalty function may lead to more balanced resource usage. We compare nodes utilization using three optimization results with $\alpha^\tau = 10^{-4}, 10^{-2}, 1$ respectively. (Reminds that a node utilization is at most $2/3$, see (18)). Figure 4 plots utilization of all 11 top-utilized nodes (above 25%) in any of the three optimization results. We clearly see that the node utilization gets more balanced as α^τ increases. The highest node utilization is reduced from 63% to 35% as we increases α^τ from 10^{-4} to 1.

Further exploring this tradeoff between the original cost function D and the balance of node resource usage by adjusting the degree of penalty may be an interesting problem. In the above examples, when $\alpha^\tau = 10^{-2}$, compared to $\alpha^\tau = 10^{-4}$, the gain is only slightly reduced by 3%, while the highest node utilization drops substantially from 63% to 49%.

VIII. CONCLUSION

In this paper, we have developed an optimal distributed algorithm for joint resource allocation and routing for so-called *node-based wireless networks*. Previous works [1] [8] have been focusing on distributed algorithms for routing optimization in wired networks where link capacities are fixed. We generalize them to take resource allocation into consideration as well. We

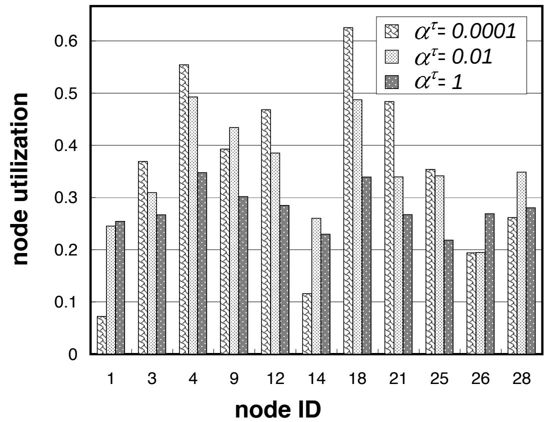

Fig. 4. Comparison of node utilization for different α^τs

have been able to prove the convergence of the algorithm. This is a significant theoretical improvement over the algorithms proposed for wired networks. We have demonstrated that our algorithm can be used to jointly optimize routing and resource allocation on a wide range of *node-based wireless data networks*, such as FDMA and TDMA type of wireless networks. We have also investigated issues such as faster algorithm adaptation to the change of network, and the impact of barrier method used in the algorithm to the original optimization goal.

Acknowledgment

This research was supported in part by the National Science Foundation under NSF grants ANI-9980552, ANI-0085848, and the Engineering Research Centers Program, award number EEC-0313747 001. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

References

[1] R. G. Gallager, "A minimum delay routing algorithm using distributed computation," in *IEEE Transaction on Communications*, january 1977, pp. 73–85.
[2] D. G. Cantor and M. Gerla, "Optimal routing in a packet-switched computer network," *IEEE Transactions on Computer*, vol. C-23, no. 10, pp. 1062–1069, 1974.
[3] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Transactions on Communications*, vol. 52, no. 7, pp. 1136–1144, 2004.
[4] R. Bhatia and M. Kodialam, "On power efficient communication over multi-hop wireless networks: Joint routing, scheduling and power control," in *INFOCOM*, 2004.
[5] R. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multi-hop wireless networks," in *INFOCOM*, 2003.
[6] A. Sankar and Z. Liu, "Maximum lifetime routing in wireless ad-hoc networks," in *INFOCOM*, 2004.
[7] M. J. Neely, "Energy optimal control for time varying wireless networks," in *INFOCOM*, 2005.
[8] D. Bertsekas and R. Gallager, *Data networks (2nd ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1992.
[9] R. Bhatia and M. Kodialam, "On power efficient communication over multihop wireless networks: Joint routing, scheduling and power control," in *IEEE Infocom*, 2004.
[10] M. Kodialam and T. Nandagopal, "Characterizing achievable rates in multi-hop wireless networks: The joint routing and scheduling problem," in *MobiCom*, 2003.
[11] S. Boyd, *Convex Optimization*. Cambridge University Press, 2004.
[12] J. Moy/IETF, "OSPF version 2(RFC 2328)," 1998.
[13] D. G. Luenberger, *Linear and Nonlinear Programming*. Kluwer Academic Publishers, 2003.

IX. Appendix A

Next is the proof for Theorem 2.1. It was proved in [1]. We include it for completeness.

Proof:

Without loss of generality, take the destination node j to be the nth of the n nodes and drop the argument j from (1),

$$t_i = r_i + \sum_{l=1}^{n-1} t_l \phi_{li}, \quad 1 \le i \le n. \tag{A1}$$

Summing both sides over i, we see that any solution to (A1) satisfies

$$t_n = \sum_i r_i. \tag{A2}$$

Temporarily let $\phi_{ni} = r_i/t_n$ and substitute this in (A1).

$$t_i = \sum_{l=1}^{n} t_l \phi_{li}. \tag{A3}$$

Any solution to (A2) and (A3) satisfies (A1) and vice versa. Let $\hat{\Phi}$ be the $n \times n$ matrix with components ϕ_{li}. $\hat{\Phi}$ is stochastic (i.e., $\phi_{li} \geq 0$ for all l, i and $\sum_i \phi_{li} = 1$ for all l) and (A3) is just the formula for steady-state probabilities in a Markov chain.

It is well know that if $\hat{\Phi}$ is irreducible, then (A3) has a unique solution, aside from a scale factor determined by (A2), and $t_i > 0$, $1 \leq i \leq n$. The matrix $\hat{\Phi}$ is irreducible; however, if for each i, k there is a path $i.l, m, \ldots, p, k$ such that $\phi_{il} > 0$, $\phi_{lm} > 0$, ..., $\phi_{pk} \geq 0$. If $r_i > 0$ for $1 \leq i \leq n-1$, then node n has a path to each i, $1 \leq i \leq n-1$. By the definition of routing variables, each i has a path to n and consequently $\hat{\Phi}$ is irreducible. Thus (A1) has a unique solution, with positive t_i, if $r_i > 0$ for $1 \leq i \leq n-1$.

Now let $t = (t_1, \ldots, t_{n-1})$, $r = (r_1, \ldots, r_{n-1})$and let Φ be the $n-1 \times n-1$ matrix with components $\phi_{li} (1 \leq i, l \leq n-1)$. Equation (A1) for $1 \leq i \leq n-1$ is then $t(I - \Phi) = r$. Since this equation has a unique solution for $r_i > 0$, $I - \Phi$ must have an inverse, and

$$t = r(I - \Phi)^{-1} \tag{A4}$$

Since the components of t are positive when the components of r are positive, components of t are nonnegative when the components of r are nonnegative. Differentiating (A4), we get the continuous function of Φ

$$\frac{\partial t_i}{\partial r_l} = [(I - \Phi)^{-1}]_{li} \tag{A5}$$

Using (A5) in (A4), the solution to (A1) is conveniently expressed, for any r, as

$$t_i = \sum_l \frac{\partial t_i}{\partial r_l} r_l \tag{A6}$$

Finally, differentiating (A1) with respect to ϕ_{km}, we get

$$\frac{\partial t_i}{\partial \phi_{km}} = \sum_{l=1}^{n-1} \frac{\partial t_l}{\partial \phi_{km}} \phi_{li} + t_k \delta_{im}$$

where $\delta_{im} = 1$ for $i = m$ and 0 and otherwise. For fixed k, m, this is the same of equations as (A1), so that the solution, continuous in ϕ, is

$$\frac{\partial t_i}{\partial \phi_{km}} = \frac{\partial t_i}{\partial r_m} t_k \tag{A7}$$

■

X. Appendix B

Next is the proof for Theorem 6.1.

Proof:

First we show that (34), repeated below with the destination node again taken to be n, has a unique solution.

$$\frac{\partial A^{\phi}(\phi)}{\partial r_i} = \sum_k \phi_{ik} \left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^{\phi}(\phi)}{\partial r_k} \right] \tag{B1}$$

Let $b_i = \sum_k \phi_{ik} \partial A^f(f)/\partial f_{ik}$ and let b the column vector $(b_1, \ldots, b_{n-1})$. Let $\nabla \bullet A^{\phi}$ be the column vector $(\partial A^{\phi}(\phi)/\partial r_1$, ..., $\partial A^{\phi}(\phi)/\partial r_{n-1})$ Then (B1) can be rewritten as

$$\nabla \bullet A^{\phi} = b + \Phi(\nabla \bullet A^{\phi}). \tag{B2}$$

We saw in the proof of Theorem 2.1 that $I - \Phi$ has a unique inverse with components given by (A5). Thus the unique solution to (B2) is

$$\frac{\partial A^{\phi}(\phi)}{\partial r_i} = \sum_l \frac{\partial t_l}{\partial r_i} \sum_m \phi_{lm} \frac{\partial A^f(f)}{\partial f_{lm}} \tag{B3}$$

$$= \sum_{l,m} \frac{\partial f_{lm}}{\partial r_i} \frac{\partial A^f(f)}{\partial f_{lm}} \tag{B4}$$

Differentiating $\partial A^{\phi}(\phi)$ directly with (2) and (3), we get the same unique solution, which, from Theorem 2.1, is continuous in ϕ.

Finally we calculate $\partial A^{\phi}(\phi)/\partial \phi_{ik}$ directly using (2) and (3),

$$\begin{aligned}
\frac{\partial A^{\phi}(\phi)}{\partial \phi_{ik}} &= \sum_{l,m} \frac{\partial A^f(f)}{\partial f_{lm}} \phi_{lm} \frac{\partial t_l}{\partial \phi_{ik}} + \frac{\partial A^f(f)}{\partial f_{ik}} t_i \\
&= t_i \Bigg[\sum_{l,m} \frac{\partial A^f(f)}{\partial f_{lm}} \phi_{lm} \frac{\partial t_l}{\partial r_k} \Bigg] + t_i \frac{\partial A^f(f)}{\partial f_{ik}} \\
&= t_i \Bigg[\frac{\partial A^{\phi}(\phi)}{\partial r_k} + \frac{\partial A^f(f)}{\partial f_{ik}} \Bigg] \qquad \text{(B5)}
\end{aligned}$$

■

We have used (A7) and (B3) to derive (B5), which is the same as (35). This is clearly continuous in ϕ given the continuity of t_i and $\partial A^{\phi}(\phi)/\partial r_i$, and the proof is complete.

XI. Appendix C

Next is the proof for Theorem 6.2.

Proof:

First we show that (36) is a necessary condition to minimize A^{ϕ} by assuming that ϕ does not satisfy (36). This means that there is some i, j, k and m such that

$$\phi_{ik}(j) > 0, \quad \frac{\partial A^{\phi}(\phi)}{\partial \phi_{ik}(j)} > \frac{\partial A^{\phi}(\phi)}{\partial \phi_{im}(j)} \tag{C1}$$

Since these derivatives are continuous, a sufficiently small increase in $\phi_{im}(j)$ and corresponding decrease in $\phi_{ik}(j)$ will decrease A^{ϕ}, thus establishing that ϕ does not minimize A^{ϕ}.

Next we show that (37), repeated below, is a sufficient condition to minimize A^{ϕ}.

$$\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^{\phi}(\phi)}{\partial r_k(j)} \geq \frac{\partial A^{\phi}(\phi)}{\partial r_i(j)}, \quad \text{all } i, j, k. \tag{C2}$$

Suppose that ϕ satisfies (C2) and has link data rates f and node data rates t. Let ϕ^* be any other set of routing variables with link data rates f^* and node data rates t^*. Define

$$f(\lambda) = (1 - \lambda) f + \lambda f^* \tag{C3}$$

$$A^f(\lambda) = A^f(f(\lambda)) \tag{C4}$$

Since A^f is a convex, non-decreasing function of the flow rate sets f, therefore $A^f(\lambda)$, is convex in λ, and hence

$$\left. \frac{dA^f(\lambda)}{d\lambda} \right|_{\lambda=0} \leq A^{\phi}(\phi^*) - A^{\phi}(\phi) \tag{C5}$$

Since ϕ^* is arbitrary, proving that $dA^f(\lambda)/d\lambda \geq 0$ at $\lambda = 0$ will complete the proof. From (C3) and (C4),

$$\left. \frac{dA^f(\lambda)}{d\lambda} \right|_{\lambda=0} = \sum_{i,k} \frac{\partial A^f(f)}{\partial f_{ik}} \left(f^*_{ik} - f_{ik} \right) \tag{C6}$$

We now show that

$$\sum_{i,k} \frac{\partial A^f(f)}{\partial f_{ik}} f^*_{ik} \geq \sum_{j,k} r_k(j) \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \tag{C7}$$

Note from (C2) that

$$\sum_k \frac{\partial A^f(f)}{\partial f_{ik}} \phi^*_{ik}(j) \geq \frac{\partial A^\phi(\phi)}{\partial r_i(j)} - \sum_k \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \phi^*_{ik}(j) \tag{C8}$$

Multiplying both sides of (C8) by $t^*_i(j)$, summing over i, j, and recalling that $f^*_{ik} = \sum_j t^*_i(j)\phi^*_{ik}(j)$ (see (2)) , we obtain

$$\sum_{i,k} \frac{\partial A^f(f)}{\partial f_{ik}} f^*_{ik} \geq \sum_{i,j} t^*_i(j) \frac{\partial A^\phi(\phi)}{\partial r_i(j)} - \sum_{i,j,k} t^*_i(j)\phi^*_{ik}(j) \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \tag{C9}$$

From (1), $\sum_i t^*_i(j)\phi^*_{ik}(j) = t^*_k(j) - r_k(j)$. Substituting this into the rightmost term of (C9) and canceling, we get (C7). Note that the only inequality used here was (C8), and that if ϕ is substituted for ϕ^*, this becomes an equality from the equation for $\frac{\partial A^\phi(\phi)}{\partial r_i(j)}$ in (34). Thus

$$\sum_{i,k} \frac{\partial A^f(f)}{\partial f_{ik}} f_{ik} = \sum_{j,k} r_k(j) \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \tag{C10}$$

Substituting (C7) and (C10) into (C6), we see that $dA^f(\lambda)/d\lambda \geq 0$ at $\lambda = 0$, completing the proof.

■

XII. Appendix D

We prove Theorem 6.3 through a sequence of seven lemmas. The first five establish the descent properties of the algorithm, the sixth establishes a type of continuity condition, showing that if ϕ does not minimize A^ϕ, the for any ϕ^* in a neighborhood of ϕ, $A^\phi(\Gamma^m(\phi^*)) < A^\phi(\phi)$ for some m. The seventh lemma is a global convergence theorem which does not require continuity in the algorithm Γ; Lemmas 12.6 and 12.7 together establish Theorem 6.3.

Let ϕ be an arbitrary set of routing variables satisfying $A^\phi(\phi) < A_0$ for some A_0. Let $\phi^1 = \Gamma(\phi)$ and let t, f, t^1, f^1 be the node and link data rates corresponding to ϕ and ϕ^1, respectively. Let f^λ, $(0 \leq \lambda \leq 1)$ be defined by $f^\lambda_{ik} = (1-\lambda) f_{ik} + \lambda f^1_{ik}$, and let

$$A^f(\lambda) = A^f(f(\lambda)) \tag{D1}$$

From the Taylor remainder theorem,

$$A^\phi(\phi^1) - A^\phi(\phi) = \left.\frac{dA^f(\lambda)}{d\lambda}\right|_{\lambda=0} + \frac{1}{2}\left.\frac{d^2 A^f(\lambda)}{d\lambda^2}\right|_{\lambda=\lambda^*} \tag{D2}$$

where λ^* is some number, $0 \leq \lambda^* \leq 1$. The continuity of the second derivative above will be obvious from the proof of Lemma 12.4, which upper bounds that term. The first three lemmas deal with $\frac{dA^f(\lambda)}{d\lambda}|_{\lambda=0}$.

Lemma 12.1:

$$\left.\frac{dA^f(\lambda)}{d\lambda}\right|_{\lambda=0} = \sum_{i,j,k} -\Delta_{ik}(j) a_{ik}(j) t^1_i(j) \tag{D3}$$

Proof: Using the definitions of $a_{ik}(j)$ and $\Delta_{ik}(j)$ in (39) and (40),

$$\begin{aligned}
\sum_k \Delta_{ik}(j) a_{ik}(j) &= \sum_{k \neq k_{min}(i,j)} [\phi_{ik}(j) - \phi^1_{ik}(j)] \left\{ \frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} - \min_{m \notin B_i(j)} \left[\frac{\partial A^f(f)}{\partial f_{im}} + \frac{\partial A^\phi(\phi)}{\partial r_m(j)} \right] \right\} \\
&= \sum_k [\phi_{ik}(j) - \phi^1_{ik}(j)] \left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \right] && \text{(D4)} \\
&= \frac{\partial A^\phi(\phi)}{\partial r_i(j)} - \sum_k \phi^1_{ik}(j) \left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \right] && \text{(D5)}
\end{aligned}$$

In (D4), we have used (41) to extend the sum over all k and in (D5), we have used (34). Multiplying both sides of (D5) by $t_i^1(j)$, summing, and using (1) and (2), we get

$$\begin{aligned}\sum_{i,j,k} \Delta_{ik}(j) a_{ik}(j) t_i^1(j) &= \sum_{i,j} t_i^1(j) \frac{\partial A^\phi(\phi)}{\partial r_i(j)} - \sum_{i,k} f_{ik}^1 \frac{\partial A^f(f)}{\partial f_{ik}} - \sum_{k,j} \left[t_k^1(j) - r_k(j)\right] \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \\ &= -\sum_{i,k} f_{ik}^1 \frac{\partial A^f(f)}{\partial f_{ik}} + \sum_{k,j} r_k(j) \frac{\partial A^\phi(\phi)}{\partial r_k(j)} \quad \text{(D6)} \\ &= \sum_{i,k} (f_{ik} - f_{ik}^1) \frac{\partial A^f(f)}{\partial f_{ik}} \quad \text{(D7)} \\ &= -\frac{dA^f(\lambda)}{d\lambda}\bigg|_{\lambda=0} \quad \text{(D8)}\end{aligned}$$

We have used (C10) to get (D7), and (D8) from (D1), completing the proof. ■

Lemma 12.2:

$$\frac{dA^f(\lambda)}{d\lambda}\bigg|_{\lambda=0} \le -\frac{1}{\eta(n-1)^3} \sum_{i,j} \Delta_i^2(j) t_i^2(j) \quad \text{(D9)}$$

where

$$\Delta_i(j) = \sum_k \Delta_{ik}(j) \quad \text{(D10)}$$

Proof: From the definition of $\Delta_{ik}(j)$ in (40), $-a_{ik}(j) \le -t_i(j)\Delta_{ik}(j)/\eta$. Substituting this into (D3) yields

$$\begin{aligned}\frac{dA^f(\lambda)}{d\lambda}\bigg|_{\lambda=0} &\le -\frac{1}{\eta} \sum_{i,j,k} \Delta_{ik}^2(j) t_i(j) t_i^1(j) \\ &\le -\frac{1}{(n-1)\eta} \sum_{i,j} \Delta_i^2(j) t_i(j) t_i^1(j) \quad \text{(D11)}\end{aligned}$$

where (D11) follows from Cauchy's inequality, $(\sum_k \alpha_k \beta_k)^2 \le (\sum \alpha_k^2)(\sum \beta_k^2)$, with $\alpha_k = 1$, $\beta_k = \Delta_{ik}(j)$, and the sum over $k \ne i$.

Now define $t_i^*(j)$ as the total flow at node i destined for j if the routing variables $\phi_{ik}(j)$ (for $k \ne k_{min}(i,j)$) are reduced by $\Delta_{ik}(j)$ but $\phi_{ik}(j)$ for $k = k_{min}(i,j)$ is not increased. Mathematically $t_i^*(j)$ satisfies

$$t_i^*(j) = \sum_l t_l^*(j)[\phi_{li}(j) - \Delta_{li}(j)] + r_i(j) \quad \text{(D12)}$$

This has a unique solution because of the loop freedom of ϕ. Subtracting (D12) from (1) results in

$$t_i(j) - t_i^*(j) = \sum_l [t_l(j) - t_l^*(j)]\phi_{li}(j) + \sum_l t_l^*(j)\Delta_{li}(j) \quad \text{(D13)}$$

From (A6), using $\sum_l t_l^*(j)\Delta_{li}(j)$ for $r_i(j)$,

$$t_i(j) - t_i^*(j) = \sum_l \frac{\partial t_i(j)}{\partial r_l(j)} \sum_k t_k^*(j)\Delta_{kl}(j) \quad \text{(D14)}$$

Since ϕ is loop-free, $\partial t_i(j)/\partial r_l(j) \le 1$. Also if $\partial t_i(j)/\partial r_l(j) > 0$, then l is upstream of i for destination j and $\phi_{il}(j)$ (and hence $\Delta_{il}(j)$) is zero. Thus

$$t_i(j) - t_i^*(j) \le \sum_l \sum_{k \ne i} t_k^*(j)\Delta_{kl}(j) = \sum_{k \ne i} t_k^*(j)\Delta_k(j) \quad \text{(D15)}$$

Multiplying the left side by $\Delta_i(j) \le 1$ preserves the inequality, yielding

$$t_i(j)\Delta_i(j) \le \sum_k t_k^*(j)\Delta_k(j) \quad \text{(D16)}$$

Since the right-hand side of (D14) is nonnegative, we also have $t_i(j)\Delta_i(j) \geq t_i^*(j)\Delta_i(j)$. We interrupt the proof now for a short technical lemma, which was proved by [1]. We include it here for completeness. The lemma will be used for further proof.

Lemma 12.3: Let α_i, $\beta_i(1 \leq i \leq m)$ be nonnegative numbers satisfying $\alpha_i \leq \sum_k \beta_k$; $\alpha_i \geq \beta_i$ for $1 \leq i \leq m$. Then

$$\sum_{i=1}^{m} \alpha_i \beta_i \geq \frac{1}{m^2} \sum_i \alpha_i^2 \tag{D17}$$

Proof:

$$\sum_i \alpha_i \beta_i \geq \sum_i \beta_i^2 \geq \frac{1}{m} (\sum \beta_i)^2 \tag{D18}$$

where we have used $\alpha_i \geq \beta_i$ and then Cauchy's inequality. Since $\sum \beta_i \geq \alpha_k$ for all k,

$$\sum_i \alpha_i \beta_i \geq \frac{1}{m} \alpha_k^2, \quad \text{for all } k. \tag{D19}$$

This implies (D17), completing the proof of Lemma (12.3). ■

Now let $\alpha_i = t_i(j)\Delta_i(j)$ and $\beta_i = t_i^*(j)\Delta_i(j)$. Since these terms are nonzero only for $i \neq j$, we can take $m = n-1$. Since the conditions of the lemma are satisfied for this choice,

$$\sum_i \Delta_i^2(j) t_i(j) t_i^*(j) \geq \frac{1}{(n-1)^2} \sum_i \Delta_i^2(j) t_i^2(j). \tag{D20}$$

Since $t_i^*(j) \leq t_i^1(j)$, we can substitute (D20) into (D11), getting (D9) and completing the proof of Lemma (12.2).

Lemma 12.4: Let M be an upper bound of $\frac{\partial^2 A^f(f^\lambda)}{\partial f^\lambda_{l_1 m_1} \partial f^\lambda_{l_2 m_2}}$ over all l_1, m_1, l_2, m_2 and over $0 \leq \lambda \leq 1$. Then for any λ, $0 \leq \lambda \leq 1$,

$$\frac{d^2 A^f(\lambda)}{d\lambda^2} \leq M(n+2)(n-1)^2 n^2 \sum_{j,k} \Delta_k^2(j) t_k^2(j) \tag{D21}$$

Proof: The bound M must exist because $\frac{\partial A^f(\lambda)}{\partial f^\lambda_{l_1 m_1} \partial f^\lambda_{l_2 m_2}}$ is a continuous function of λ over the compact region $0 \leq \lambda \leq 1$. Taking the second derivative, we get

$$\begin{aligned}
\frac{d^2 A^f(\lambda)}{d\lambda^2} &= \sum_{l_1, m_1} \sum_{l_2, m_2} \frac{\partial^2 A^f(\lambda)}{\partial f^\lambda_{l_1 m_1} \partial f^\lambda_{l_2 m_2}} (f^1_{l_1 m_1} - f_{l_1 m_1})(f^1_{l_2 m_2} - f_{l_2, m_2}) \\
&\leq \sum_{l_1 m 1} \sum_{l_2 m_2} M |f^1_{l_1 m_1} - f_{l_1 m_1}| |f^1_{l_2 m_2} - f_{l_2 m_2}| \\
&\leq \sum_{i,k} M |E| |f^1_{ik} - f_{ik}|^2 \\
&\leq \sum_{i,k} M n(n-1) |f^1_{ik} - f_{ik}|^2
\end{aligned} \tag{D22}$$

We now upper bound $|f^1_{ik} - f_{ik}|$ by first upper bounding $|t^1_i(j) - t_i(j)|$. As in the proof of Lemma 12.2, we have

$$\begin{aligned}
t_i^1(j) - t_i(j) &= \sum_l [t_l^1(j) - t_l(j)] \phi^1_{li}(j) + \sum_l t_l(j) [\phi^1_{li}(j) - \phi_{li}(j)] \\
&= \sum_l \frac{\partial t_i^1(j)}{\partial r_l(j)} \sum_k t_k(j) [\phi^1_{kl}(j) - \phi_{kl}(j)]
\end{aligned} \tag{D23}$$

Since $0 \leq \partial t_i^1(j) / \partial r_l(j) \leq 1$, we can upper bound this by

$$t_i^1(j) - t_i(j) \leq \sum_k t_k(j) \Delta_k(j)$$

We can lower bound (D23) in the same way, considering only terms in which $\phi^1_{kl}(j) - \phi_{kl}(j) < 0$, and this leads to

$$|t^1_i(j) - t_i(j)| \leq \sum_k t_k(j)\Delta_k(j) \tag{D24}$$

$$f^1_{ik} - f_{ik} = \sum_j [t^1_i(j) - t_i(j)]\phi^1_{ik}(j) + t_i(j)[\phi^1_{ik}(j) - \phi_{ik}(j)]$$

$$|f^1_{ik} - f_{ik}| \leq \sum_j \sum_l t_l(j)\Delta_l(j)\phi^1_{ik}(j) + \sum_j t_i(j)|\phi^1_{ik}(j) - \phi_{ik}(j)| \tag{D25}$$

The double sum in (D25) has at most $(n-1)^2$ nonzero terms ($j \neq i$, $l \neq j$) and the second sum at most $n-1$ terms. Using Cauchy's inequality on both terms together, we get

$$|f^1_{ik} - f_{ik}|^2 \leq n(n-1)\left\{\sum_{j,l} t^2_l(j)\Delta^2_l(j)[\phi^1_{ik}(j)]^2 + \sum_j t^2_i(j)[\phi^1_{ik}(j) - \phi_{ik}(j)]^2\right\}$$

$$\sum_k |f^1_{ik} - f_{ik}|^2 \leq n(n-1)\left\{\sum_{j,l} t^2_l(j)\Delta^2_l(j) + 2\sum_j t^2_i(j)\Delta^2_i(j)\right\} \tag{D26}$$

Summing over i and substituting the result in (D22),we get (D21) completing the proof. ■

Lemma 12.5: For given A_0, define

$$M = \max_{l_1,m_1,l_2,m_2} \max_{f:A^f(f)\leq A_0} \frac{\partial^2 A^f(f)}{\partial f_{l_1m_1}\partial f_{l_2m_2}}(f) \tag{D27}$$

$$\eta = [Mn^8]^{-1}. \tag{D28}$$

Then for all ϕ such that $A^\phi(\phi) \leq A_0$,

$$A^\phi(\phi^1) - A^\phi(\phi) \leq -\frac{1}{2\eta(n-1)^3}\sum_{i,j}\Delta^2_i(j)t^2_i(j). \tag{D29}$$

Proof: Temporarily let M be as defined in Lemma 12.4. Combining Lemma 12.2 and Lemma 12.4,

$$A^\phi(\phi^1) - A^\phi(\phi) \leq \left[-\frac{1}{\eta(n-1)^3} + \frac{Mn^2(n-1)^2(n+2)}{2}\right]\sum_{i,j}\Delta^2_i(j)t^2_i(j). \tag{D30}$$

For $\eta = [Mn^8]^{-1}$, the second term in brackets above is less than half the magnitude of the first term, yielding (D29). It follows that $A^\phi(\phi^1) \leq A^\phi(\phi) \leq A_0$. By convexity then $A^f(f^\lambda) \leq A_0$ for $0 \leq \lambda \leq 1$. Thus M as given in (D27) satisfies the condition on M in Lemma 12.4, completing the proof. ■

Lemma 12.6: Let the scale factor η satisfy (D28) for a given A_0 and let ϕ be an arbitrary set of routing variables that does not minimize A^ϕ and satisfies $A^\phi(\phi) \leq A_0$. Given this ϕ, there exists and $\epsilon > 0$ and an m, $1 \leq m \leq n$, such that for all ϕ^* satisfying $|\phi - \phi^*| < \epsilon$,

$$A^\phi(\Gamma^m(\phi^*)) < A^\phi(\phi) \tag{D31}$$

Proof: We consider three cases. The first is the typical case in which no blocking occurs and $A^\phi(\Gamma(\phi)) < A^\phi(\phi)$, the second is the case in which blocking occurs, and the third is the case in which $A^\phi(\Gamma(\phi)) = A^\phi(\phi)$.

Case 1: No blocking; $\Delta_i(j)t_i(j) > 0$ for some i, j. If no nodes are blocked for ϕ, then by the definition of blocking (42), there is a neighborhood of ϕ^* around ϕ for which no blocking occurs. In this neighborhood,

$$a_{ik}(j) = \left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)}\right] - \min_{1\leq m\leq n}\left[\frac{\partial A^f(f)}{\partial f_{im}} + \frac{\partial A^\phi(\phi)}{\partial r_m(j)}\right] \tag{D32}$$

which is continuous in ϕ. It follows from (40) that $\Delta_{ik}(j)$ is continuous in ϕ, and the upper bound to $A^\phi(\Gamma(\phi)) - A^\phi(\phi)$ in (D29) is continuous in ϕ. Since by assumption the bound in (D29) is strictly negative, there is a neighborhood of ϕ^* around ϕ for which

$$A^\phi(\Gamma(\phi^*)) - A^\phi(\phi^*) < -\frac{1}{4\eta(n-1)^3}\sum_{i,j}\Delta^2_i(j)t^2_i(j) \tag{D33}$$

where $\Delta_i(j)$ and $t_i(j)$ correspond to the given ϕ. Choose ϵ small enough so that (D33) is satisfied for $|\phi - \phi^*| < \epsilon$ and also so that

$$|A^\phi(\phi^*) - A^\phi(\phi)| \quad < \quad \frac{1}{4\eta(n-1)^3} \sum_{i,j} \Delta_i^2(j) t_i^2(j)$$

Combining this with (D33), we have (D31) for $m = 1$.

Case 2: Blocking occurs. For any ϕ, we can use (34) to lower bound $a_{ik}(j)$ by

$$a_{ik}(j) \geq \frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} - \frac{\partial A^\phi(\phi)}{\partial r_i(j)} \tag{D34}$$

$$\Delta_{ik}(j)t_i(j) \geq \min\left\{\phi_{ik}(j)t_i(j), \eta\left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} - \frac{\partial A^\phi(\phi)}{\partial r_i(j)}\right]\right\} \tag{D35}$$

The lower bounds above are continuous functions of ϕ. Since blocking occurs in ϕ, there is some i, j, k such that both

$$\frac{\partial A^\phi(\phi)}{\partial r_k(j)} - \frac{\partial A^\phi(\phi)}{\partial r_i(j)} \geq 0 \tag{D36}$$

and

$$\phi_{ik}(j)t_i(j) \geq \eta\left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi)}{\partial r_k(j)} - \frac{\partial A^\phi(\phi)}{\partial r_i(j)}\right] \tag{D37}$$

Combining (D35) to (D37)

$$\Delta_{ik}(j)t_i(j) \geq \eta \frac{\partial A^f(f)}{\partial f_{ik}} \tag{D38}$$

Since the right-hand side of (D35) is continuous in ϕ, there is a neighborhood of ϕ^* around ϕ for which

$$\Delta^*_{ik}(j)t^*_i(j) \geq \frac{\eta}{2}\frac{\partial A^f(f)}{\partial f_{ik}} \tag{D39}$$

Equation (D31), for $m = 1$, now follows in the same way as in case 1.

Case 3: $\Delta_{ik}(j)t_i(j) = 0$ for all i, j, k. Let Φ_3 be the set of ϕ for which $\Delta_{ik}(j)t_i(j) = 0$ for all i, j, k. Let $\phi^{(l)} = \Gamma^l(\phi)$ for the given ϕ and let $m \geq 2$ be the smallest integer such that $\phi^{(m-1)} \notin \Phi_3$. We first show that $m \leq n$. Note first that for any $\phi \in \Phi_3$, Γ changes $\phi_{ik}(j)$ only for i, j such that $t_i(j) = 0$ and thus the node data rates t and link data rates f cannot change. $\partial A^\phi/\partial r_i(j)$ can change, however, and as we shall see later, must change for some i, j if ϕ does not minimize A^ϕ.

Now consider $\phi^{(l)}$ ($0 \leq l \leq m-2$, where $\phi^{(0)}$ denotes the original ϕ). Since $\phi^{(l)} \in \Phi_3$, $\Delta^{(l)}_{ik}(j) > 0$ implies that $t_i(j) = 0$. From (40), $\phi^{(l)}_{ik}(j) = \Delta^{(l)}_{ik}(j)$ and $\phi^{(l+1)}_{ik}(j) = 0$. For a given i, j, all $\phi^{(l)}_{ik}(j)$ are reduced to 0 except for the k which minimizes $\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi^{(l)})}{\partial r_k(j)}$. Thus, using (34),

$$\frac{\partial A^\phi(\phi^{(l+1)})}{\partial r_i(j)} = \min_k\left[\frac{\partial A^f(f)}{\partial f_{ik}} + \frac{\partial A^\phi(\phi^{(l)})}{\partial r_k(j)}\right] \leq \frac{\partial A^\phi(\phi^{(l)})}{\partial r_i(j)} \tag{D40}$$

Since this equation is satisfied for all l, $0 \leq l \leq m-2$, we see that $\partial A^\phi(\phi^{(l)})/\partial r_i(j)$ can be reduced on iteration l only if $\partial A^\phi(\phi^{(l-1)})/\partial r_k(j)$ is reduced on iteration $l-1$ for some k such that $\partial A^\phi(\phi^{(l-1)})/\partial r_k(j) < \partial A^\phi(\phi^{(l)})/\partial r_i(j)$. This reduction at node k however implies a reduction at some node k' of smaller differential cost at iteration $l-2$ and so forth. Since this sequence of differential costs is decreasing with decreasing l and since (from (D40)) the differential cost at a given node is nondecreasing with decreasing l, each node in the sequence must be distinct. Since there are $n-1$ nodes other than the given destination available for such a sequence, the initial l in such a sequence satisfies $l \leq n-2$. On the other hand, if $\partial A^\phi(\phi^{(l)})/\partial r_i(j)$ is unchanged for all i, j, we see from (D40) that $\phi^{(l)}$ satisfies the sufficient conditions to minimize A^ϕ and then ϕ also minimizes A^ϕ contrary to our hypothesis; thus we must have $m \leq n$.

Now observe that the middle expression in (D40), for $l = 0$, is a continuous function of ϕ and consequently $\partial A^\phi(\phi^{(1)})/\partial r_i(j)$ is a continuous function of ϕ for all i, j. It follows by induction that $\partial A^\phi(\phi^{(l)})/\partial r_i(j)$ is a continuous function of ϕ for all i, j and for $l \leq m-1$. Finally $\phi^{(m-1)} \notin \Phi_3$, so it must satisfy the conditions of case 1 or 2; it will be observed that the analysis there apply equally to $\phi^{(m-1)}$ because of the continuity of $\partial A^{\phi}(\phi^{(m-1)})/\partial r_i(j)$ as a function of ϕ. This completes the proof. ■

Our last lemma will be stated in greater generality than required since it is a global convergence theorem for algorithms that avoids the usual continuity constraint on the algorithm. (See Luenberger [13]) for a good discussion of global convergence).

Lemma 12.7: Let Φ be a compact region of Euclidean N space. Let Γ be a mapping from Φ into Φ and let A^{ϕ} be a continuous real valued function in Φ. Assume that $A^{\phi}(\Gamma(\phi)) \leq A^{\phi}(\phi)$ for all $\phi \in \Phi$. Let A^{ϕ}_{min} be the minimum of A^{ϕ} over Φ and let Φ_{min} be the set of $\phi \in \Phi$ such that $A^{\phi}(\phi) = A^{\phi}_{min}$. Assume that for every $\phi \in \Phi - \Phi_{min}$, there is an $\epsilon > 0$ and an integer $m \geq 1$ such that for all $\phi^* \in \Phi$ satisfying $|\phi - \phi^*| < \epsilon$, we have $A^{\phi}(\Gamma^m(\phi^*)) < A^{\phi}(\phi)$. Then for all $\phi \in \Phi$,

$$\lim_{m\to\infty} A^{\phi}(\Gamma^m(\phi)) = A^{\phi}_{min}. \tag{D41}$$

Proof: See [1]. ■

Proof of Theorem 6.3: Let Φ be the set of loop-free routing variable ϕ such that $A^{\phi}(\phi) \leq A_0$. We have verified that Γ maps loop-free routing variables into loop-free routing variables, and from Lemma 12.5, $A^{\phi}(\Gamma(\phi)) \leq A^{\phi}(\phi)$ for $\phi \in \Phi$. Thus Γ is mapping from Φ into Φ. It is obvious that Φ is bounded and easy to verify that any limit of loop-free variables with $A^{\phi}(\phi) \leq A_0$ is also loop-free with $A^{\phi}(\phi) \leq A_0$. Thus ϕ is compact. The final assumption of Lemma 12.7 is established by Lemma 12.6. Thus Lemma 12.7 asserts the conclusion of Theorem 6.3. ■