Optimal Rate Allocation and Traffic Splits for Energy Efficient Routing in Ad Hoc Networks

Vikram Srinivasan, Carla F. Chiasserini, Pavan Nuggehalli, Ramesh R. Rao

Abstract— **In this paper, we address the problem of providing traffic quality of service and energy efficiency in ad hoc wireless networks. We consider a network that is shared by a set of sources, each one communicating with its corresponding destination using multiple routes. Each source is associated with a utility function which increases with the total traffic flowing over the available source-destination routes. The network lifetime is defined as the time until the first node in the network runs out of energy. We formulate the problem as one of maximizing the sum of the sources' utilities subject to the required constraint on network lifetime. We present a primal formulation of the problem, which uses penalty functions to take into account the system constraints, and we introduce a new methodology for solving the problem. The proposed approach leads to a flow control algorithm, which provides the optimal sources' rate and can be easily implemented in a distributed manner. When compared with the minimum transmission energy routing scheme, the proposed algorithm gives significantly higher sources' rates for same network lifetime guarantee.**

Keywords—**Mobile and wireless networks, Flow control, Quality of service.**

I. Introduction

The convergence of various technologies has made ubiquitous wireless access a reality and enabled wireless systems to support a large variety of applications, from Internet-based services to remote sensing.

We deal with ad hoc networks composed of battery-powered nodes, which communicate with each other using multihop wireless links. Each network node acts also as a router, forwarding data packets to other nodes. Since batteries can supply only a finite amount of energy, a major challenge in such networks is minimizing the nodes' energy consumption, which depends on the power spent by the nodes to transmit, receive, and process traffic. Clearly, a trade-off between energy consumption and traffic performance (e.g., throughput and delay) exists.

Several papers have addressed the issue of energy consumption in wireless ad hoc networks by proposing energy-aware routing algorithms [1], [2], [3], [4], [5], [6]. In particular, in [1] the so-called MTE (Minimum Transmission Energy) routing scheme is presented, which selects the route that uses the least amount of energy to transport a packet from the source to the destination. In [4], the concept of network lifetime is first defined as the period from the time instant when the network starts functioning to the time instant when the first node runs out of energy. The objective there is to maximize the network lifetime while guaranteeing the required traffic rate.

In this paper, we consider an ad hoc network composed of wireless nodes, each of which may have a different initial energy. The network is shared by a set of traffic sources and each source has a unique destination for all its data. Sources do not require a fixed bandwidth but can adjust their transmission rates to changes in network conditions (e.g., as in the case of Internet-based applications using TCP). Each source knows the set of routes that can be used to reach its destination; the possible routes can be discovered by applying a source routing algorithm, as in [7]. The advantage of using multiple paths is twofold: *(i)* It provides an even distribution of the traffic load, i.e., energy drain, over the network. *(ii)* In case of route disruption, the source is still able to send data to the destination by using the functioning routes.

Considering this scenario, we pose the following problem: given a required network lifetime, what is the most beneficial source rate allocation and flow control strategy?

To answer this question, we draw upon previous work on congestion pricing in wired networks [8], [9], [10], [11], [12], [13], [14]. Their approach consists in deriving the control schemes for the sources' traffic rate as solutions of an optimization problem. Each traffic source is associated with a utility function increasing in its transmission rate and subject to bandwidth constraints; the network objective is to maximize the sum of source utilities. The network problem is decomposed into several sub-problems each of them corresponding to a single traffic source. In [9], [10], it is shown that when a single path between a traffic source and its destination is considered and the objective function is strictly concave, solving the single source sub-problems is the same as solving the global network problem. In [13], [15], [16], the multipath case is ad-

C.F. Chiasserini is with Dipartimento di Elettronica, Politecnico di Torino, Torino, Italy. E-mail: chiasserini@polito.it . V. Srinivasan, P. Nuggehalli and R.R. Rao are with the Electrical and Computer Engineering Department, University of California, San Diego, La Jolla, CA. E-mail: {vikram,pavan,rao}@cwc.ucsd.edu .

dressed. Solving the optimization problem in the multipath case becomes more difficult because, even if the objective functions of the source sub-problems are strictly concave, the overall objective function may not be so. Hence, extensions of the approaches adopted for the single path case do not provide convergence to an optimal solution of the global network problem. Solutions to approximate versions of the problem are presented in [13], while an exact formulation is solved in [16].

In this paper, we use an optimization approach to address the problem of providing energy efficiency and traffic quality of service in ad hoc wireless networks. The network lifetime, as defined in [4], and the traffic rate over the available routes between each source-destination pair, are taken as measures of the network performance. Each source is associated with a utility function which increases with the traffic flowing over the available source-destination routes. We consider a primal formulation of the network optimization problem, where the objective is maximizing the sum of the sources' utilities for a required network lifetime guarantee. Then, in order to solve the problem in the multipath case, we present a new formulation, which makes use of penalty functions to take into account the system constraints [17]. We prove that the optimal solution of the proposed formulation converges to an optimal solution of the original problem and we show that the optimal solution can be obtained by applying a gradient descent method. By using the gradient descent technique, we devise a distributed flow control algorithm, named *ORSA (Optimal Rate Splitting and Allocation)*, that quickly converges to the optimal sources' rates.

The performance of the ORSA scheme is compared against the performance of the MTE algorithm [1]. Results show that, given the desired network lifetime, the ORSA algorithm allows for much higher sources' rates than the MTE scheme when *(i)* the source density in the network is less than 0.5 or *(ii)* the energy resources are unevenly distributed among the nodes. By increasing the number of available source-destination paths, higher sources' rates can be achieved. Results also suggest that an optimal number of source-destination routes can be found, that allows for high sources' rates while keeping the system complexity low.

The remainder of the paper is organized as follows. Section II describes the system model and a mathematical representation of the flow control problem. Section III introduces the methodology proposed for solving the optimization problem. Section IV provides numerical results; and, Section V reviews some related work. Finally, Section VI concludes the paper.

II. THE FLOW CONTROL PROBLEM

In this section, we first introduce the notation and assumptions that we use to model the system under study. Then, a mathematical representation of the network optimization problem is given, which takes into account both the sources' traffic rates and the network lifetime.

A. Notation and Assumptions

We model an ad hoc network with a set $\mathcal{N}$ of stationary wireless nodes, although the extension to a time-varying network topology is straightforward. We indicate the number of nodes by $|\mathcal{N}| = N$ and assume that the network is shared by a set $\mathcal{S}$ of sources. Let $\mathcal{D}$ be the set of destination nodes in the network; for the sake of simplicity, we assume that each source has a unique destination for all its traffic.

A path or a route, $r \subset \mathcal{N}$, is a subset of nodes. Let $\mathcal{R}$ be the set of routes. Let $\mathcal{R}(i), i \in \mathcal{N}$, be the set of routes that contain node i, $\mathcal{R}^{\mathcal{S}}(s)$, be the set of routes starting at node $s, s \in \mathcal{S}$, and $\mathcal{R}^{\mathcal{D}}(d)$, be the set of routes that end at node $d, d \in \mathcal{D}$. We define $\mathcal{N}^{\mathcal{S}}(s), s \in \mathcal{S}$, as the set of nodes belonging to any route in $\mathcal{R}^{\mathcal{S}}(s)$. For each source, we assume that the set of all possible routes toward the destination is known through a source routing algorithm such as the one proposed in [7].

Given a route r and a node $i \in r$, we let f_{ri} be the node immediately succeeding node i on route r. The energy required to transmit one unit flow from node i to the generic node j is denoted by $e_{ij}^{(tx)}$. We say that $e_{ij}^{(tx)} = \infty$ if no communication link exists between i and j. This parameter depends on the distance between nodes i and j, channel conditions, antenna gains, and receive/transmit powers.

Let x_s be the traffic rate that is associated with source $s, s \in \mathcal{S}$ and is split by s on its $|\mathcal{R}^{\mathcal{S}}(s)|$ routes. Let $y_r, r \in \mathcal{R}^{\mathcal{S}}(s)$ be the flow on route r, i.e., the fraction of traffic rate x_s routed through r; we have

$$x_s = \sum_{r \in \mathcal{R}^{\mathcal{S}}(s)} y_r \,. \tag{1}$$

Next, we assume that each node has a limited amount of available energy and denote by E_i the energy available at node $i, i \in \mathcal{N}$. We consider that energy costs are incurred in transmit and receive mode, while energy consumption due to traffic processing is neglected. The energy consumed per unit flow while receiving, denoted by $e^{(rx)}$, is assumed to be constant. Let γ_i be the power consumed by node $i, i \in \mathcal{N}$. Then,

$$\gamma_i = \sum_{r \in \mathcal{R}^{\mathcal{S}}(i)} y_r e_{i f_{ri}}^{(tx)} + \sum_{r \in \mathcal{R}^{\mathcal{D}}(i)} y_r e^{(rx)}$$

$$+ \sum_{\substack{r \in R(i) \\ r \notin \mathcal{R}^{\mathcal{S}}(i) \cup \mathcal{R}^{\mathcal{D}}(i)}} y_r \left(e^{(rx)} + e^{(tx)}_{i f_{ri}} \right) \quad (2)$$

where the first term on the right hand side is the power consumed to transmit the traffic generated by node i, the second term represents the power spent to receive the traffic of which i is the destination, and the third term is the transmission and reception cost due to the traffic that is relayed through i.

We define the network lifetime, L, as the time until the first node in the network runs out of energy, as first defined in [4]. By denoting by L_i the lifetime of node i, the network lifetime can be written as

$$L = \min_{i \in \mathcal{N}} L_i \,. \quad (3)$$

Let L_g be the required guarantee on the network lifetime. Then, the maximum energy consumption per unit time, or equivalently the maximum power consumption, allowed at node i is equal to

$$\Gamma_i = \frac{E_i}{L_g}. \quad (4)$$

By limiting the nodes' power consumption to Γ_i, we ensure that the network lifetime is at least equal to L_g. We define the 'congestion' of node i, denoted by Y_i, as

$$Y_i = \frac{\gamma_i}{\Gamma_i}\,. \quad (5)$$

When the power consumption of node i is equal to its maximum allowed value, Γ_i, we have $Y_i = 1$.

Observe that when a constraint on the level of the nodes' output power exists, Γ_i can be viewed as the maximum allowed value of transmission power.

B. Problem Statement

The optimization approach consists in deriving control mechanisms for the sources' traffic rates as solutions of an optimization problem. Different flow control algorithms can be obtained by varying the problem objective function or the solution approach. Below, we present the objective function to be maximized in our network problem, along with the constraints on the system variables that were introduced in the previous section.

For each source $s, s \in \mathcal{S}$, we define a utility function

$$V_s \left(\sum_{r \in \mathcal{R}^{\mathcal{S}}(s)} y_r \right) \quad (6)$$

where $V_s : R_+ \rightarrow R_+$ depends solely on the rate allocated to source s, with $\sum_{r \in \mathcal{R}^{\mathcal{S}}(s)} y_r = x_s$, and is assumed to be strictly concave, continuous, bounded and increasing in $x_s, x_s \in [0, \infty)$. Since the goal of the network is to maximize the utility of all sources while providing the desired lifetime, the centralized network problem can be written as

$$\max_{y_r, r \in \mathcal{R}} \sum_{s \in \mathcal{S}} V_s \left(\sum_{r \in \mathcal{R}^{\mathcal{S}}(s)} y_r \right) \quad (7)$$

$$\begin{array}{lll} \text{subject to} & y_r \geq 0 & \forall r \in \mathcal{R} \\ & \sum_{r \in \mathcal{R}^{\mathcal{S}}(s)} y_r \leq M_s & \forall s \in \mathcal{S} \\ & Y_i \leq 1 & \forall i \in \mathcal{N}\,. \end{array}$$

The first constraint emphasizes the non-negativity of the traffic rates. The second constraint says that the rate at each source $s, s \in \mathcal{S}$ must be less than a maximum value M_s. M_s depends on the characteristics of the system and/or the application requirements; a minimum rate requirement can be similarly specified. The third condition ensures that the network lifetime guarantee is met, i.e., the power consumption of any node in the network is always less than the maximum allowed consumption rate.

III. A PENALTY FUNCTION-BASED APPROACH

The objective function in (7) is strictly concave in x_s but is not strictly concave in $\{x_s, y_r\}$, thus a unique solution does not exist and the dual function is not differentiable. In this case, simple solution approaches based on the gradient descent method are not directly applicable [18].

Here, we propose a novel approach to solve (7), which uses exact penalty functions. A penalty function is said to be exact if a constrained nonlinear programming problem can be solved by a single minimization of an unconstrained problem [19]. We consider the following unconstrained optimization problem

$\mathbf{P_0}$:

$$\max_{y_r, r \in \mathcal{R}} \quad \sum_{s \in \mathcal{S}} V_s \left(\sum_{r \in \mathcal{R}^{\mathcal{S}}(s)} y_r \right) - \sum_{i \in \mathcal{N}} p(Y_i - 1) - \sum_{s \in \mathcal{S}} p(x_s - M_s) - \sum_{r \in \mathcal{R}} p(-y_r) \quad (8)$$

where Y_i is the congestion of node i and $p(t)$ is a scalar penalty function $p : R \rightarrow R$ given by

$$p(t) = \begin{cases} e^{(\alpha t)} - 1 & t \geq 0 \\ 0 & t < 0\,. \end{cases} \quad (9)$$

It is easy to verify that the function $p(t)$ defined in (9) satisfies the following assumptions:

$$\begin{array}{ll} 1. & p \text{ is convex} \\ 2. & p(t) = 0 \quad \forall t \leq 0 \\ 3. & p(t) > 0 \quad \forall t > 0\,. \end{array} \quad (10)$$

Then, we show that (9) is an exact penalty function, by using the result below [17, Prop. 1].

Theorem 1: Let $\tilde{y}$ be the solution of problem (8).

(a) A necessary condition for $\tilde{y}$ to be an optimal solution of problem (7) is that

$$\lim_{t\to 0^+} \frac{p(t)}{t} \geq \lambda_i \quad \forall \lambda_i \in \overline{\lambda} \tag{11}$$

for some Lagrange multiplier vector $\overline{\lambda} = \{\lambda_i\}$ of (7).

(b) A sufficient condition for problems (7) and (8) to have the same solution is

$$\lim_{t\to 0^+} \frac{p(t)}{t} > \lambda_i \quad \forall \lambda_i \in \overline{\lambda} \tag{12}$$

for some Lagrange multiplier vector $\overline{\lambda} = \{\lambda_i\}$.

We note that for sufficiently large values of α, part (b) of the theorem will be satisfied for any network. Thus, by solving the unconstrained problem given by (8), we also obtain a solution to (7).

A. Solving the Penalty Function-based Problem

The penalty function in (8) is not strictly convex, consequently neither is the objective function in $\mathbf{P_0}$. This implies that problem $\mathbf{P_0}$ does not have a unique solution and that the objective function is not differentiable; hence a gradient descent method can not be used to solve problem (8). An optimal solution, however, can be found by constructing a sequence of strictly concave and differentiable optimization problems whose solutions converge to an optimal solution of $\mathbf{P_0}$. Since these problems are strictly concave and differentiable, they possess a unique solution that can be obtained by applying the gradient descent method [18]. Other optimal solutions to $\mathbf{P_0}$ can be attained by selecting different penalty functions or different sequences of optimization problems.

Consider a function $p_n(t)$ defined as follows

$$p_n(t) \quad = \quad \begin{cases} e^{(\alpha t)} - 1 + \frac{1}{n} & t \geq 0 \\ e^{(\alpha t n)}/n & t < 0\,. \end{cases} \tag{13}$$

Note that $p_n(t)$ is strictly convex and differentiable in $t \in [-\infty, \infty)$. Now, consider the problem

$\mathbf{P_n}$:

$$\max_{y_r, r\in\mathcal{R}} \quad \sum_{s\in\mathcal{S}} V_s \left(\sum_{r\in\mathcal{R}^{\mathcal{S}}(s)} y_r \right) - \sum_{i\in\mathcal{N}} p_n(Y_i - 1) - \sum_{s\in\mathcal{R}^{\mathcal{S}}(s)} p_n(x_s - M_s) - \sum_{r\in\mathcal{R}} p_n(-y_r) \tag{14}$$

Notice that the contribution of each node to the sum in the second term of (14) is equal to (less than) 1 when the node's power consumption is equal to (less than) its maximum allowed value, i.e., when $Y_i = 1$ ($Y_i < 1$). Also, the greater the power consumption, the higher the value of the penalty function.

The objective function of $\mathbf{P_n}$ is strictly concave. In fact, since $p_n(-y_r)$ is strictly convex in y_r, the last term in (14) is strictly concave in $\{y_r\}$. As mentioned earlier, this implies that $\mathbf{P_n}$ has a unique solution and is differentiable; hence the solution can be obtained by the gradient descent method. All that we need to show is that the sequence of solutions $\{y_n\}$ converges to $\tilde{y}$, i.e., the solution to (8).

Let $q(n)$ be the optimal value of problem $\mathbf{P_n}$. Then, we have the following lemma.

Lemma 1: $q(n)$ is an increasing sequence.

Proof:

$$q(n) = \sum_{s\in\mathcal{S}} V_s \left(x_s^{(n)}\right) - \sum_{i\in\mathcal{N}} p_n \left(Y_i^{(n)} - 1\right) - \sum_{s\in\mathcal{S}} p_n \left(x_s^{(n)} - M_s\right) - \sum_{r\in\mathcal{R}} p_n \left(-y_r^{(n)}\right)\,. \tag{15}$$

Here $\left\{x_s^{(n)}\right\}$ and $\left\{y_r^{(n)}\right\}$ are the optimal rates for problem $\mathbf{P_n}$ and $Y_i^{(n)}$ is the corresponding congestion of node i. We have

$$p_n(t) \quad > \quad p_{n+1}(t)\,. \tag{16}$$

This implies that

$$\begin{aligned} q(n) &< \sum_{s\in\mathcal{S}} V_s \left(x_s^{(n)}\right) - \sum_{i\in\mathcal{N}} p_{n+1} \left(Y_i^{(n)} - 1\right) \\ &\quad - \sum_{s\in\mathcal{S}} p_{n+1} \left(x_s^{(n)} - M_s\right) - \sum_{r\in\mathcal{R}} p_{n+1} \left(-y_r^{(n)}\right) \\ &\leq q(n+1) \end{aligned} \tag{17}$$

where the first inequality follows from (16), while the second inequality follows from the definition of $q(n)$ in (15).

Note that $q(n)$ is bounded above by V^*, i.e., the value of (8). Since $\{q_n\}$ is a monotonically increasing bounded sequence, it has a limit q^* and $q^* \leq V^*$. It remains to show that $q^* = V^*$.

Fix any $\epsilon > 0$. For sufficiently large n we have

$$p_n(t) \quad < \quad p(t) + \frac{\epsilon}{3|\mathcal{R}|}\,. \tag{18}$$

This implies that

$$q_n \quad = \quad \max_{y_r r\in\mathcal{R}} \sum_{s\in\mathcal{S}} V_s(x_s) - \sum_{i\in\mathcal{N}} p_n(Y_i - 1) - \sum_{s\in\mathcal{R}^{\mathcal{S}}(s)} p_n(x_s - M_s) - \sum_{r\in\mathcal{R}} p_n(-y_r)$$

$$
\begin{aligned}
> & \sum_{s\in\mathcal{S}} V_s(\tilde{x}_s) - \sum_{i\in\mathcal{N}} p_n(\tilde{Y}_i - 1) \\
& - \sum_{s\in\mathcal{S}} p_n(\tilde{x}_s - M_s) - \sum_{r\in\mathcal{R}} p_n(-\tilde{y}_r)\,. \qquad (19)
\end{aligned}
$$

Here $\{\tilde{x}_s\}$ and $\{\tilde{y}_r\}$ are the optimal rates for problem (8), and $\tilde{Y}_i$ is the corresponding congestion of node i. Thus,

$$
\begin{aligned}
q_n \quad > & \sum_{s\in\mathcal{S}} V_s(\tilde{x}_s) - \sum_{i\in\mathcal{N}} p(\tilde{Y}_i - 1) \\
& - \sum_{s\in\mathcal{S}} p(\tilde{x}_s - M_s) - \sum_{r\in\mathcal{R}} p(-\tilde{y}_r) - \epsilon \\
\geq & \; V^* - \epsilon\,. \qquad (20)
\end{aligned}
$$

Since ϵ is arbitrary, we have $q^* = V^*$. ■

B. *The ORSA Algorithm*

As discussed in the previous section, the solution to problem (8) can be approached as closely as desired by choosing a sufficiently large value for n. Moreover, it was shown that the solution to problem $\mathbf{P_n}$ can be obtained by applying the gradient descent method. In the following, we present a distributed implementation of this algorithm, the so-called *ORSA (Optimal Rate Splitting and Allocation)* algorithm.

We consider the utility function for the generic source $s, s \in \mathcal{S}$, as [13]

$$
V_s(x_s) \;=\; \log(1 + x_s)\,. \qquad (21)
$$

Observe that V_s increases as the source rate increases; the log function is used in the expression of the source utility because it ensures proportional fairness.

Problem $\mathbf{P_n}$ can be therefore rewritten as

$\mathbf{P_n}$:

$$
\max_{y_r, r\in\mathcal{R}} \quad \sum_{s\in\mathcal{S}} U_s^{(n)}\left(x_s, \left\{Y_k, k \in \mathcal{N}^{\mathcal{S}}(s)\right\}\right) \qquad (22)
$$

with

$$
\begin{aligned}
U_s^{(n)}\left(x_s, \left\{Y_k, k \in \mathcal{N}^{\mathcal{S}}(s)\right\}\right) \;=& \\
& \log(1 + x_s) - \frac{1}{|\mathcal{S}|} \sum_{i\in\mathcal{N}} p_n(Y_i - 1) \\
& - p_n(x_s - M_s) - \sum_{r\in\mathcal{R}^{\mathcal{S}}(s)} p_n(-y_r)\,. \qquad (23)
\end{aligned}
$$

The optimal solution of (22) must satisfy the first order conditions [18]

$$
\begin{aligned}
\frac{\partial U_s^{(n)}}{\partial y_r} \;=& \; -\sum_{k\in\mathcal{N}^{\mathcal{S}}(s)} \frac{\partial U_s^{(n)}}{\partial Y_k}\frac{\partial Y_k}{\partial y_r} \\
& - \sum_{z\in\mathcal{S}\setminus\{s\}} \sum_{k\in\mathcal{N}^{\mathcal{S}}(z)} \frac{\partial U_z^{(n)}}{\partial Y_k}\frac{\partial Y_k}{\partial y_r} \\
=& \; w_r^s(n) + w_r(n) \\
& \forall r \in \mathcal{R}^{\mathcal{S}}(s) \quad \forall s \in \mathcal{S}\,. \qquad (24)
\end{aligned}
$$

In (24), the left hand side represents the marginal increase in utility for source s if s increases its rate on route r by a small amount. The first term on the right hand side represents the marginal decrease in source s's utility due to the increase in y_r; this term is denoted by $w_r^s(n)$. The second term on the right hand side represents the marginal decrease in utility for all other sources; we denote this term by $w_r(n)$. Hence, (24) says that each source node must increase the flow on each route, until the marginal increase in its utility is equal to the marginal decrease in the utility imposed on all nodes in the system.

In order to obtain the optimal solution in a decentralized fashion, we consider that each source $s, s \in \mathcal{S}$ solves the following problem [8],

$\mathbf{P_n^s}$:

$$
\begin{aligned}
\max_{y_r, r\in\mathcal{R}^{\mathcal{S}}(s)} \quad & U_s^{(n)}\left(x_s, \left\{Y_k, k \in \mathcal{N}^{\mathcal{S}}(s)\right\}\right) \\
& - \sum_{r\in\mathcal{R}^{\mathcal{S}}(s)} w_r(n) y_r\,. \qquad (25)
\end{aligned}
$$

Indeed, the above maximum is obtained when

$$
\begin{aligned}
\frac{\partial U_s^{(n)}}{\partial y_r} \;=& \; -\sum_{k\in\mathcal{R}^{\mathcal{S}}(s)} \frac{\partial U_s^{(n)}}{\partial Y_k}\frac{\partial Y_k}{\partial y_r} + w_r(n) \\
=& \; w_r^s(n) + w_r(n) \quad \forall r \in \mathcal{R}^{\mathcal{S}}(s) \qquad (26)
\end{aligned}
$$

which is the same condition as the one expressed in (24). This shows that, by solving sub-problem $\mathbf{P_n^s}$ for each $s \in \mathcal{S}$, we can attain the global optimum for $\mathbf{P_n}$ in a decentralized fashion.

Next, we introduce the distributed algorithm to be performed at each source s, in order to solve $\mathbf{P_n^s}$. By applying the gradient descent method, each source needs to compute the gradient of its utility function with respect to $y_r, r \in \mathcal{R}^{\mathcal{S}}(s)$. For each route $r \in \mathcal{R}^{\mathcal{S}}(s)$, the source algorithm is as follows.

Source Algorithm

1. Evaluate gradient (∇_r) on route r

   $\nabla_r = \frac{\partial U_s^{(n)}}{\partial y_r} - w_r^s(n) - w_r(n)$

   with $U_s^{(n)}$ as in (23)

2. $y_r(t) = y_r(t-1) + \nabla_r \epsilon$ /*update flow rate over r*/

where $\epsilon > 0$ is a scaling parameter that determines the step size for the gradient descent algorithm and t is the current update time. Notice that the source does not require $w_r^s(n)$ and $w_r(n)$ separately, but $w_r^s(n) + w_r(n)$.

Let route r be defined by $\{i_0, i_1, \ldots, i_l\}$, where i_0 is the source and i_l is the destination node and the others are intermediate nodes. The algorithm to compute $w_r^s(n) + w_r(n)$ for each route $r, r \in \mathcal{R}^{\mathcal{S}}(s)$, is given below.

Route Algorithm

1. $j = 0$, $w_r^s(n) = 0$, $w_r(n) = 0$
2. At node i_{l-j} /*update $w_r^s(n) + w_r(n)$*/

$$w_r^s(n) + w_r(n) = w_r^s(n) + w_r(n) + |\mathcal{R}(i_{l-j})| \frac{\partial p_n(Y_{i_{l-j}} - 1)}{\partial Y_{i_{l-j}}} \frac{\partial Y_{i_{l-j}}}{\partial y_r}$$

3. Relay $w_r^s(n) + w_r(n)$ to node i_{l-j-1}
   $j = j + 1$
   if ($l - j \neq 0$)
   Goto Step 2
   end if

Step 2 is derived from (24); each contribution is obtained by fixing node index k in (24) and summing over all the sources whose routes include k.

IV. Numerical Results

We consider the topology shown in Figure 1, which has been obtained by randomly distributing N nodes over a $Q \times Q$ region, with $N = 20$ and $Q = 1$. We assume that each node is characterized by a different maximum allowed power consumption, Γ_i with $i = 1, \ldots, N$. Recall from (4) that Γ_i's depend on the nodes' initial energy and the required network lifetime. We assume that Γ_i's are uniformly distributed random variables with mean equal to 0.75. The normalized value of energy consumed per unit flow in receive mode is assumed to be constant and equal to 0.01; while, the energy spent to transmit a unit flow depends on the distance between the transmitting and the receiving node. Using the DSR algorithm [7], we find multiple routes between each source-destination pair. The plots shown in the following are derived by averaging the results over 10 different runs, each of them corresponding to a different set of sources randomly chosen among the network nodes and different instances of the random variables Γ_i's.

Figure 2 presents the average source rate as a function of the number of sources in the network and compares the results obtained through the ORSA algorithm with the performance of the MTE scheme. For each source-destination pair, up to five available routes are considered; the MTE scheme always selects the route with the minimum energy cost among the available ones. Results are derived for two different distributions of the maximum allowed power consumption. Curves labeled in the plot by Var1 refer to the case where Γ_i's are uniformly distributed between 0.5 and 1, thus resulting in a variance in the nodes' maximum allowed power consumption roughly equal to 0.02. Curves labeled by Var2 refer to the case where Γ_i's are uniformly distributed with mean equal to 0.75 and variance of about 0.04. This corresponds to a variance of the nodes' initial energy in the Var2 case being twice the variance in the Var1 case.

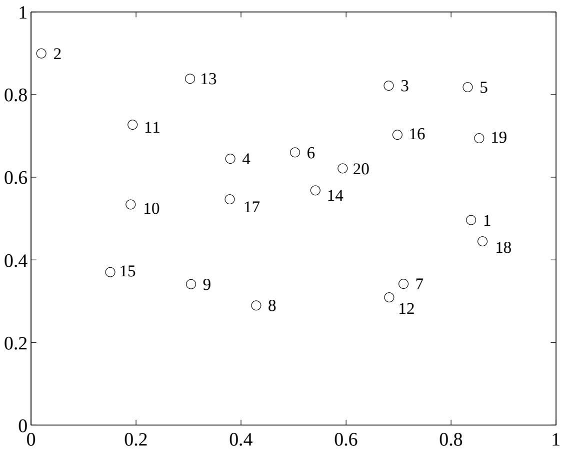

Fig. 1. Network topology with $N = 20$.

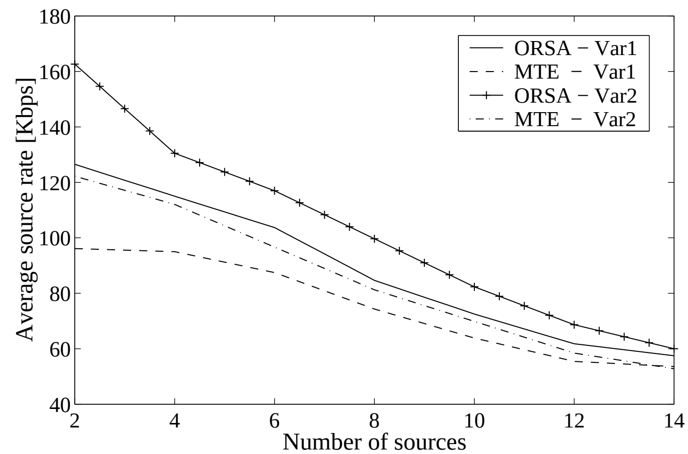

Fig. 2. Average source rate as a function of the number of sources for maximum number of available source-destination routes equal to 5. The performance of the ORSA algorithm is compared to the results obtained through the MTE algorithm for two different values of variance of the nodes' maximum allowed power consumption.

Figure 2 shows that the ORSA algorithm significantly

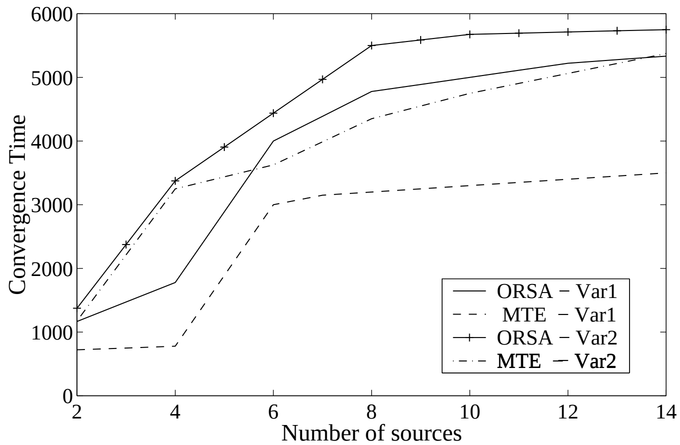

Fig. 3. Average convergence time of the rate allocation algorithm as a function of the number of sources in the network for maximum number of available source-destination routes equal to 5. The performance of the ORSA algorithm is compared to the results obtained through the MTE algorithm for two different values of variance of the nodes' maximum allowed power consumption.

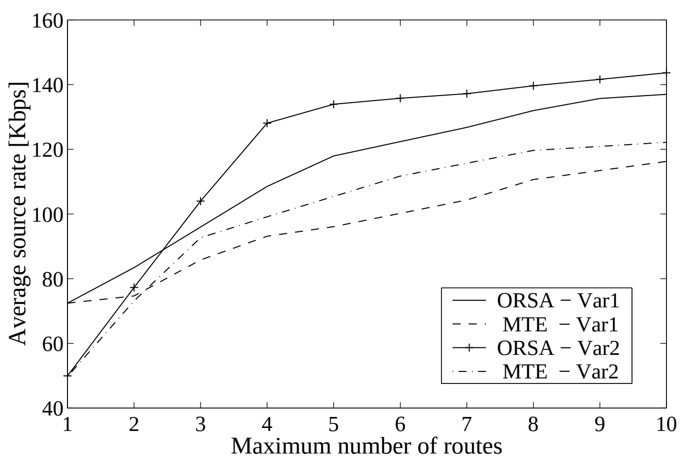

Fig. 4. Average source rate as a function of the maximum number of available source-destination routes for $S = 4$. The performance of the ORSA algorithm is compared to the results obtained through the MTE algorithm.

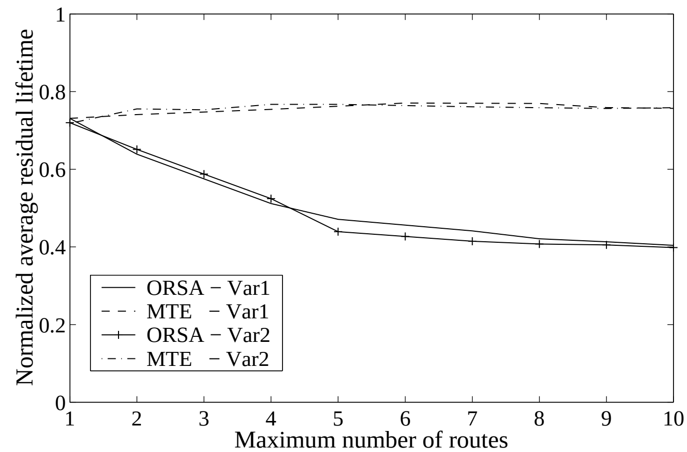

Fig. 5. Average of the normalized residual lifetime of the network nodes when the first node runs out of energy as a function of the maximum number of available source-destination routes. The performance of the ORSA algorithm is compared to the results obtained through the MTE algorithm for $S = 4$.

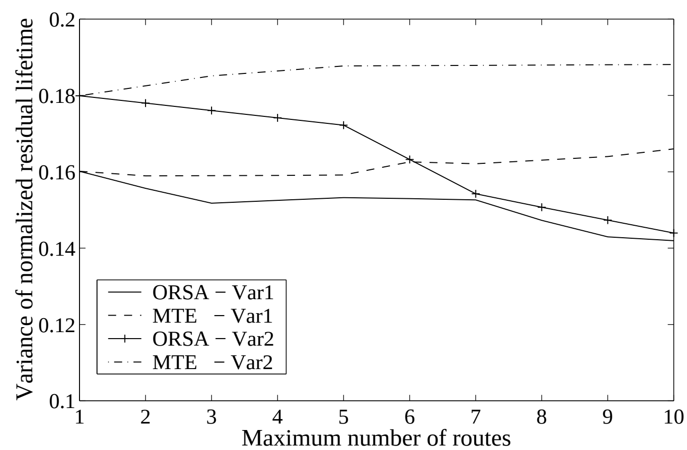

Fig. 6. Variance of the normalized residual lifetime of the network nodes when the first node runs out of energy as a function of the maximum number of available source-destination routes. The performance of the ORSA algorithm is compared to the results obtained through the MTE algorithm for $S = 4$.

outperforms the MTE scheme as long as the number of sources that are simultaneously active does not exceed half the total number of nodes in the network. For higher values of source density, the ORSA and MTE algorithms perform equally well. Interestingly, as the variance in the distribution of the nodes' initial energy increases, the gain of the ORSA scheme over the MTE algorithm increases. This suggests that a multipath scheme would be preferable when *(i)* the density of simultaneously active sources in the network is not very high or *(ii)* the energy resources are unevenly distributed among the network nodes. Otherwise, a single path routing scheme that has a lower complexity may be more desirable.

Under the same system assumptions, we derive the convergence time of the ORSA and MTE schemes, expressed as the number of algorithm iterations. Figure 3 shows that the MTE algorithm always outperforms the ORSA scheme; however, the gap in performance significantly decreases as the variance in the nodes' available energy becomes larger (curves labeled in the plot by Var2). Figure 3 also shows that the convergence time grows with the increase of the number of sources; in fact, greater the number of sources, the more likely it is that the traffic flow through the network nodes is changed at each iteration.

Figures 4–6 are derived by considering four active sources and varying the maximum number of available source-destination routes between 1 and 10.

Figure 4 compares the average source rate obtained through the ORSA and MTE schemes for the two cases, Var1 and Var2, as described above. The average source rate increases with the number of available routes. In fact, as the number of paths between each source-destination pair increases, a larger set of energy resources becomes available for optimization. When only one route is available, the ORSA and the MTE algorithms are the same. Instead, when the number of available routes is greater than one, 20–30% improvement can be obtained by the ORSA scheme over the MTE scheme. Also, Figure 4 shows that as the number of available routes increases beyond a certain threshold, the improvement in performance becomes marginal. This suggests that an optimal number of source-destination routes can be found, that allows for high sources' rates while keeping the system complexity low.

Next, in order to gain an insight into the dynamics of energy consumption, we study the residual lifetimes of the nodes at the end of the network's lifetime. Figures 5 and 6 present the mean and variance of the residual lifetime of the nodes, normalized to the desired network lifetime, L_g. With the MTE scheme, the mean and variance of the residual lifetimes are higher than that obtained through the ORSA scheme. This shows that the ORSA algorithm balances the load in the network, which results in a more equitable use of energy resources.

V. Related Work

Optimization based flow control schemes have been previously proposed in [8], [9], [10], [11], [12], [13], [14], [15], [16] in the context of wired networks.

Most of the work on flow control has been done under the assumption that a single path exists between each source-destination pair. In [9], [10], [13], each traffic source is associated with a utility function increasing in its transmission rate and subject to bandwidth constraints; the network objective is to maximize the sum of source utilities. The network problem is decomposed into several sub-problems each of them corresponding to a single traffic source. By introducing the notion of resource *price* and setting the price value according to the resource congestion level, source nodes can adjust their transmission rates so that the optimal trade-off between utility value and the price they have to pay for the network resources is achieved. While in [9], [13] sources select a willingness to pay and the network allocates the traffic rates, in [10] the sources determine their rates and pay the corresponding price. The two approaches correspond to a primal and a dual formulation of the optimization problem, respectively.

In [15], the same approach as in [10] is applied to the multipath case. The network fixes the price for each source-destination path and the source sends all its traffic on the route with the minimum price. The sources' rates, that are determined through this algorithm, are optimal only when the objective function of the dual problem is evaluated in the minimum price vector [20]. However, even in this case it is hard to obtain the optimal solution for the primal problem from the optimal solution for the dual problem due to the lack of strict concavity of the dual objective function. The optimal dual solution may yield multiple primal solutions, and some of them can be infeasible [16],[18, Chap. 6].

In [8], the authors apply the economic theory of pricing of congestible resources to a networking context. They detail the use of congestion pricing to efficiently allocate a single network resource among many users, and extend their work to different economic scenarios such as monopolistic and competitive environments. The focus there is on sharing a *single* network resource such as a ftp server or a router.

An algorithm based on a primal formulation of the rate control problem for the single-path case is presented in [14], and it is generalized to the multipath case in [16]. An exact solution is derived, which optimizes both flow control and routing in wired networks. In this case the congestion of the network resources is indicated through a binary variable ($1 =$"*congested*", $0 =$"*not congested*"). This approach is not suitable in a wireless context for the following reasons. In a wired capacitated network, each link is as good as any other, i.e., the design of a wired network ensures that the effort required to transmit a bit on any link is the same. On the contrary, in a wireless scenario, distance between the nodes determines the transmit effort. From an optimization viewpoint, in a wired context the Lagrange multipliers of the optimization problem are the price per unit flow. In a wireless context, these multipliers can be thought of as price per unit flow per unit distance. Thus, using binary indicators of congestion, as done in [14], [16], is not sufficient.

VI. Conclusions and Discussion

In this paper, we jointly addressed the problem of flow control and energy efficiency in wireless ad hoc networks. We considered a network scenario where multiple routes between each source-destination pair are known to the source. We associated to each source node a utility function which increases with the traffic flowing over the available source-destination routes, and we formulated the problem as one of maximizing the sum of the source utilities for a required network lifetime guarantee. We proposed a new methodology for solving the problem, that converges to the optimal solution. This technique enabled us to apply the gradient descent method and derive a simple and distributed flow control algorithm that provides the optimal sources' rates. In order to study the performance of the proposed algorithm, named ORSA, we compared it with the MTE scheme. We found that under the required constraint on network lifetime, the ORSA algorithm significantly outperforms the MTE algorithm when the source density in the network is less than 0.5. By increasing the number of available source-destination paths, higher sources' rates can be achieved. Results also suggest that an optimal number of source-destination routes can be found, that allows for high sources' rates while keeping the system complexity low.

The definition of lifetime that we considered allowed us to gain some insight in the interplay between energy efficiency and throughput. However, this definition is not completely satisfactory as it does not address some key issues of ad hoc networks. For instance, the death of a single node does not completely disrupt the communication between the various source-destination pairs. A more natural definition of network lifetime would be the time until no communication is feasible between some source-destination pair. With this definition, however, it is no longer straightforward to pose the problem in an analytically tractable manner. Also, a common lifetime guarantee for the whole network may be unfair. The fact that some nodes can be unduly burdened in relaying other nodes' traffic may be highly undesirable for some network applications and device characteristics. Therefore, we believe that devising appropriate performance metrics and designing algorithms that ensure a more equitable network behavior, are important aspects that still need to be addressed in future research.

REFERENCES

[1] T. Shepard, "Decentralized Channel Management in Scalable Multihop Spread Spectrum Packet Radio Networks," Tech. Rep. MIT/LCS/TR-670, Massachusetts Institute of Technology Laboratory for Computer Science, July 1995.

[2] S. Singh, M. Woo, C.S. Raghavendra, "Power-aware Routing in Mobile and Ad Hoc Networks," *Proc. of 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, Dallas, TX, Oct. 1998, pp. 181–190.

[3] V. Rodoplu, T.H. Meng, "Minimum Energy Mobile Wireless Networks," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 8, August 1999, pp. 1333–1344.

[4] J.-H. Chang, L. Tassiulas, "Energy Conserving Routing in Wireless ad hoc Networks," *Proc. of INFOCOM 2000,* Tel Aviv, Israel, March 2000.

[5] P. Manzoni, J.-C. Cano, "A Performance Comparison of Energy Consumption for Mobile Ad Hoc Network Routing Protocols," *Proc. of the 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems,* 2000, pp. 57–64.

[6] A. Michail, A. Ephremides, "Energy Efficient Routing for Connection-oriented Traffic in ad hoc Wireless Networks," *Proc. of the 11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC),* 2000, pp. 762–766.

[7] J. Broch, D.B. Johnson, D.A. Maltz, "The Dynamic Source Routing Protocol for Mobile Ad hoc Networks," draft-ietf-manet-dsr-03.txt, 22 Oct. 1999.

[8] J.K. Mackie-Mason, H.R. Varian, "Pricing Congestible Network Resources," *IEEE Journal on Selected Areas in Communications,* vol. 13, no. 7, Sep. 1995, pp. 1141–1149.

[9] F.P. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications,* vol. 8, no. 1, 1997, pp. 33–37.

[10] S.H. Low, D.E. Lapsley, "Optimization Flow Control - I: Basic Algorithm and Covergence," *IEEE/ACM Transactions on Networking,* vol. 7, no. 6, Dec. 1999.

[11] R. La, V. Ananthram, "Charge Sensitive TCP and Rate Control in the Internet," *Proc. of IEEE INFOCOM 2000*, Tel Aviv, Israel, Mar. 2000.

[12] S. Kunniyur, R. Srikant, "End-to-End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks," *Proc. of IEEE INFOCOM 2000*, Tel Aviv, Israel, Mar. 2000.

[13] F. Kelly, A. Maulloo, D. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of Operations Research Society,* vol. 49, no. 3, Mar. 1998, pp. 237–252.

[14] K. Kar, S. Sarkar, L. Tassiulas, "A Simple Rate Control Algorithm for Maximizing Total User Utility," *Proc. of IEEE Infocom 2001*, Anchorage, Alaska, Apr. 2001.

[15] S.H. Low, "Multipath Optimization Flow Control," *Proc. of IEEE International Conference on Networks (ICON)*, 2000, pp. 39–43.

[16] K. Kar, S. Sarkar, L. Tassiulas, "Optimization Based Rate Control for Multipath Sessions," *Technical Report ISR TR 2001-1,* Institute for Systems Research and University of Maryland, 2001.

[17] D.P. Bertsekas, "Necessary and Sufficient Conditions for a Penalty Method to Be Exact," *Mathematical Programming,* vol. 9, 1975, pp. 87–99.

[18] D.P. Bertsekas, *Nonlinear Programming,* Athena Scientific, 1995.

[19] G. Di Pillo, "Exact Penalty Methods," in *Algorithms for Continuous Optimization*, Kluwer Academy Publication, 1994.

[20] D.P. Bertsekas, J.N. Tsitsiklis, *Parallel and Distributed Computation,* Prentice-Hall International Editions, 1989.