

Low-Complexity and Distributed Energy Minimization in Multi-hop Wireless Networks

Longbi Lin, Xiaojun Lin, and Ness B. Shroff

Center for Wireless Systems and Applications (CWSA)
School of Electrical and Computer Engineering, Purdue University
West Lafayette, IN 47907, USA
{llin, linx, shroff}@ecn.purdue.edu

Abstract—In this work, we study the problem of minimizing the total power consumption in a multi-hop wireless network subject to a given offered load. It is well-known that the total power consumption of multi-hop wireless networks can be substantially reduced by jointly optimizing power control, link scheduling, and routing. However, the known optimal cross-layer solution to this problem is centralized, and with high computational complexity. In this paper, we develop a low-complexity and distributed algorithm that is provably power-efficient. In particular, under the node exclusive interference model, we can show that the total power consumption of our algorithm is at most twice as large as the power consumption of the optimal (but centralized and complex) algorithm. Our algorithm is not only the first such distributed solution with provable performance bound, but its power-efficiency ratio is also tighter than that of another sub-optimal centralized algorithm in the literature.

Index Terms—Energy Aware Routing, Duality, Mathematical Programming/Optimization, Cross-Layer Optimization, Simulations

I. INTRODUCTION

There has been significant recent interest in developing control protocols for multi-hop wireless networks. Many applications can benefit from the deployment of these networks. For instance, sensors can form multi-hop wireless networks for such applications as habitat monitoring [2], and the management of sewer overflow events [20]. Vehicles can form multi-hop wireless networks to exchange safety messages and traffic information [17]. Wireless LAN devices can form multi-hop mesh networks to provide wireless broadband access [1].

A key issue in developing control protocols for multi-hop wireless networks is to reduce the energy or power consumption. This is obviously an important issue for battery-powered networks since the power consumption often limits the lifetime of the network. Even for networks with access to power sources, the transmission power of the communication links may still need to be properly controlled, e.g., due to health or regulatory concerns.

In this work, we are interested in the problem of minimizing the total power consumption of a multi-hop wireless network, subject to a given offered load. It is well-known that the total power consumption of multi-hop wireless networks can be substantially reduced by jointly optimizing power control, link

scheduling, and routing. However, known optimal solutions require centralized computation and high computational complexity. In this paper, we propose a new low-complexity and distributed solution to this problem under a widely-used interference model, called the node-exclusive interference model. Using this model, the work in [4] developed a centralized solution that yielded a 3-approximation ratio (the resultant power consumption is within a factor of 3 from the optimal power consumption). In contrast, in this paper, we will obtain a $(2 + \varepsilon)$ -approximation algorithm that is fully distributed, where $\varepsilon > 0$ is an arbitrarily small constant. To the best of our knowledge, our proposed algorithm is the first *distributed* solution in the literature with a provable performance bound.

Our solution approach is inspired by the recent progress in using imperfect scheduling algorithms to develop distributed cross-layer congestion control and scheduling algorithms in multi-hop wireless networks. We first formulate the energy minimization problem into a special form that naturally leads to a distributed solution. We then map the solution to corresponding components of the cross-layer control protocols, and rigorously establish the power-efficiency of the resulting algorithm.

Our work is also related to the study of energy-aware routing protocols for minimizing energy consumption and extending network lifetime [10], [11], [13], [22], [25]. These works assume that the system capacity is battery-limited instead of interference-limited and therefore do not consider scheduling constraints. In contrast, our work explicitly considers scheduling, jointly with power control and routing.

The intellectual contribution of this work is summarized as follows:

- We develop a low-complexity and distributed joint routing, power control, and scheduling algorithm for multi-hop wireless networks with provable power-efficiency ratio. Further, our algorithm can guarantee a better power-efficiency level than some existing centralized algorithms.
- Our solution cannot be obtained by extending the known optimal solution in the literature [8], [18]. Instead, we develop an optimization approach to the energy minimization problem that naturally leads to distributed solutions. We also develop rigorous techniques to identify the

convergence properties and quantify the power-efficiency of the resulted control protocol.

The rest of this paper is organized as follows: in Section II, we present the system model and formulate the energy minimization problem. In Section III, we present our algorithm and discuss its implications. In Section IV, we discuss our main result on the convergence properties and the power-efficiency of the dual algorithm. Numerical results are provided in Section V. Concluding remarks are presented in Section VI.

II. PROBLEM FORMULATION

We model a wireless multi-hop network by a directed graph $G(V, E)$, where V is the set of vertices representing the nodes, and E is the set of edges representing the communication links. We use $N_o(v)$ and $N_i(v)$ to denote the sets of outgoing and incoming links of node v , respectively. Their union $N(v)$ forms the set of all links incident on node v .

The system is time-slotted. We adopt the following node-exclusive interference model that is used to characterize FH-CDMA and UWB system with perfect orthogonal spreading codes and low power-spectrum density. Under this model, a node can only receive from or transmit to at most one node at any time-slot m . Further, each link is power-controlled. That is, if node-exclusiveness is satisfied, we assume that the possible data rate R_e of link e is a result of its power assignment, p_e . We use $h(R_e)$ to denote the power consumption of supporting data rate of R_e . It is assumed that $h(\cdot)$ is a non-decreasing, convex function satisfying $h(0) = 0$. An example of $h(\cdot)$ is the power-rate relationship in an Additive White Gaussian Noise (AWGN) channel. Other interference models can be incorporated in this formulation, and we will discuss this in Section VI.

Each packet may take multiple hops to be delivered from source to destination. Let T_{vd} denote the long-term average data rate of the flow that needs to be supported from source node v to destination node d . We use D to denote the set of destinations.

The joint energy minimization problem is now formulated as follows:

$$\begin{aligned}
 (*) \quad & \min_{\vec{f}, \vec{R}} \quad \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \sum_{e \in E} h(R_e(m)), \quad (1) \\
 & \text{subject to} \quad \vec{R}(m) \text{ satisfies node-exclusiveness, } \forall m, \\
 & \quad \sum_d f_e^d = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M R_e(m), \forall e, \quad (2) \\
 & \quad \sum_{e \in N_o(v)} f_e^d - \sum_{e \in N_i(v)} f_e^d - T_{vd} \geq 0, \\
 & \quad \forall d, \text{ and } \forall v \neq d, \quad (3)
 \end{aligned}$$

where the quantity f_e^d can be interpreted as the average amount of data rate on link e allocated for destination d . The constraints in (2) require that the long-term average data rate, determined by the power allocation, should be able to support the total average data rate ($\sum_d f_e^d$) on each link.

The constraints in (3) require that the total outgoing flow of a node should be able to support the total incoming and locally generated flow, for all destinations. Note that we use $\forall d$ as a shorthand notation for all nodes d belonging to D , the set of destinations. We will refer to the above problem as Problem (*).

III. SOLUTION METHODOLOGY

A. Approximating the Energy Minimization Problem

The optimal solutions developed in [8], [18] attempt to solve Problem (*). However, their solutions contain a scheduling component with high computational complexity. In order to compute at which power and at what time each link e should be activated, these solutions need to solve a complex global optimization problem in each time-slot.

In this paper, in order to obtain a low-complexity and distributed solution, we take a different approach. We first approximate (*) by another optimization problem that is easier to solve. The following Lemma [4] provides the first step in this direction.

Lemma 1: For some time-slot m when link e is activated, the instantaneous data rate $R_e(m)$ is independent of m in any power-optimal scheme.

It is worth noting that $R_e(m) = R_e$ holds when link e is activated. As a result, $\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M R_e(m)$ is equal to the product of R_e and the frequency that link e is activated. Therefore, by (2), the objective function of Problem (*) can now be written as

$$\sum_{e \in E} \frac{\sum_d f_e^d}{R_e} h(R_e),$$

where $\frac{\sum_d f_e^d}{R_e}$ is the fraction of time-slots that link e is activated.

Further, using the results from low-complexity scheduling [16], we have

- *Fact 1:* In the optimal solution to (*), we must have $\sum_{e \in N(v)} \frac{\sum_d f_e^d}{R_e} \leq 1, \forall v \in V$.
- *Fact 2:* Under the node-exclusive interference model, if $\sum_{e \in N(v)} \frac{\sum_d f_e^d}{R_e} \leq \frac{1}{2} - \eta, \forall v \in V$, where $\eta > 0$ is an arbitrarily small number, then a maximal schedule [19] can be computed such that each link is activated for $\frac{\sum_d f_e^d}{R_e}$ fraction of time-slots. We will discuss more about the role of maximal scheduling in our solution in Section III-D.

Based on these two facts, in the rest of the paper, we will replace the scheduling constraints by

$$\sum_{e \in N(v)} \frac{\sum_d f_e^d}{R_e} \leq \beta, \forall v \in V. \quad (4)$$

Problem (*) can then be reformulated as the following

problem:

$$(A) \quad \min_{\vec{f}, \vec{R}} \quad \sum_{e \in E} \frac{\sum_d f_e^d}{R_e} h(R_e), \quad (5)$$

subject to (3) and (4)
 $(\vec{f}, \vec{R}) \in X,$

where $X = \{(\vec{f}, \vec{R}) : R_e \geq 0, f_e^d \geq 0, \forall e, d\}$. The formulation in (A) is not only easier to solve, but it also produces natural bounds for proving the power efficiency ratio of our solution. Indeed, solving (A) with $\beta = 1$ provides a lower bound on the minimum power of (*), while $\beta = \frac{1}{2} - \eta$ provides an upper bound.

Remark: This problem appears in the formulation of [4]. However, our solution is different from this point on. As we mentioned earlier, their solution is a centralized one with an approximation ratio of 3, while our solution is a distributed one with a better approximation ratio.

B. Handling the Non-Convexity

In Problem (A), the objective function and constraints (4) are non-convex. Problems of this type are considered to be difficult in general. To overcome this difficulty, the following change of variable is performed:

$$t_e = \frac{\sum_d f_e^d}{R_e}, \quad \forall e \in E.$$

The physical meaning of t_e is the *fraction of time-slots* that link e is activated. It can also be interpreted as the *load* on link e . The latter interpretation seems more appropriate when we deal with $t_e(m)$ for each time-slot in the dual solution. Note that $t_e = 0$ implies that $f_e^d = 0$ for any destination d and any link e : if a link is not activated (or the load is zero), it cannot support any non-zero data rate. The long-term average power consumption Θ from link e can therefore be defined as

$$\Theta(\vec{f}_e, t_e) = \begin{cases} 0, & t_e = 0, \\ t_e h\left(\frac{\sum_d f_e^d}{t_e}\right), & t_e > 0. \end{cases} \quad (6)$$

Using the above notation, Problem (A) can be transformed into

$$(B) \quad \min_{\vec{f}, \vec{t}} \quad \sum_{e \in E} \Theta(\vec{f}_e, t_e), \quad (7)$$

subject to $\sum_{e \in N(v)} t_e \leq \beta, \forall v \in V,$ (8)

$$\begin{aligned} \sum_{e \in N_o(v)} f_e^d - \sum_{e \in N_i(v)} f_e^d - T_{vd} &\geq 0, \\ \forall d, \text{ and } \forall v \neq d, & \\ (\vec{f}, \vec{t}) &\in Y, \end{aligned} \quad (9)$$

where $Y = \{(\vec{f}, \vec{t}) : f_e^d \geq 0, \forall e, d; 0 \leq t_e \leq 1, \sum_d f_e^d \leq at_e, \forall e, \text{ for some } a \geq 0\}$.

Remark: The condition $\sum_d f_e^d \leq at_e$, for some $a \geq 0$, in the definition of Y is required to guarantee that the optimal solution from Problem (B) can be transformed to the optimal

solution of Problem (A). Without this condition, Problem (B) can be trivially solved by setting $\vec{t} = 0$ and \vec{f} large enough for the flow constraint. Such a solution cannot be transformed to an optimal solution of Problem (A), since the instantaneous data rate \vec{R} must be finite.

To show that there is no duality gap in solving Problem (B), we need the following Lemma:

Lemma 2: Let

$$\theta(f, t) = \begin{cases} 0, & t = 0, \\ th\left(\frac{f}{t}\right), & t > 0. \end{cases} \quad (10)$$

If $h(\cdot)$ is a convex function on \mathbb{R}^+ , then $\theta(f, t)$ is also convex on $C = \bigcup_{a \in [0, +\infty)} \{(f, t) : 0 \leq f \leq at, t \geq 0\}$.

We have included a proof of Lemma 2 in the Appendix.

Remark: A function of the form $g(f, t) = th\left(\frac{f}{t}\right)$, where $t > 0$ is known as the *perspective* of function $h(f)$. As shown in [5], the perspective is one of the transformations that preserve convexity. The transformation in the above Lemma includes the $t = 0$ case, and therefore can be viewed as a slightly generalized version of the perspective. Our proof in the Appendix takes a different approach than [5].

Note that θ is non-convex over the entire $\mathbb{R}^+ \times \mathbb{R}^+$. However, the convexity of θ over C is enough for our purpose, since $(\vec{f}, \vec{t}) \in Y$ in Problem (B). Due to Lemma 2, the objective function $\Theta(\vec{f}, \vec{t})$ of (B) is convex, and the entire problem is a convex program. We can then use the duality approach to solve the problem.

C. Distributed Algorithm Based on Lagrange Duality

To use the duality approach to solve the above problem, we first form the Lagrangian:

$$\begin{aligned} L(\vec{f}, \vec{t}, \vec{\mu}, \vec{q}) &= \sum_{e \in E} \Theta(\vec{f}_e, t_e) + \sum_{v \in V} \mu_v \left(\sum_{e \in N(v)} t_e - \beta \right) \\ &\quad - \sum_{v, d} q_v^d \left(\sum_{e \in N_o(v)} f_e^d - \sum_{e \in N_i(v)} f_e^d - T_{vd} \right), \end{aligned}$$

where $\vec{\mu} \geq 0$ and $\vec{q} \geq 0$ are the Lagrange multipliers, and $\vec{f}_e = \{f_e^d\}_{d \in D}$. For ease of notation, we define $q_d^d = 0$, for all d . By rearranging the order of summation, the above equation can be transformed into the following:

$$\begin{aligned} L(\vec{f}, \vec{t}, \vec{\mu}, \vec{q}) &= \sum_{e \in E} c_e(\vec{f}_e, t_e) - \beta \sum_v \mu_v + \sum_{v, d} (q_v^d T_{vd}), \end{aligned}$$

where

$$\begin{aligned} c_e(\vec{f}_e, t_e) &= \Theta(\vec{f}_e, t_e) + (\mu_{x(e)} + \mu_{r(e)})t_e \\ &\quad - \sum_d (q_{x(e)}^d - q_{r(e)}^d) f_e^d, \end{aligned} \quad (11)$$

and $x(e)$ and $r(e)$ are the transmission node and reception node of link e , respectively.

The dual objective function is

$$\begin{aligned} D(\vec{\mu}, \vec{q}) &= \min_{(\vec{f}, \vec{t}) \in Y} L(\vec{f}, \vec{t}, \vec{\mu}, \vec{q}), \\ &= \sum_{e \in E} \left[\min_{(\vec{f}_e, t_e) \in Y_e} c_e(\vec{f}_e, t_e) \right] - \beta \sum_v \mu_v + \sum_{v,d} (q_v^d T_{vd}), \end{aligned} \quad (12)$$

where Y_e denotes the constraint set on link e , namely

$$Y_e = \{(\vec{f}_e, t_e) : 0 \leq f_e^d \leq H, \forall d; 0 \leq t_e \leq 1; \sum_d f_e^d \leq at_e, \text{ for some } a \geq 0\}. \quad (13)$$

To ensure convergence, we have added $f_e^d \leq H$, where H is a large constant to be determined later. See the proof of Theorem 4 (in Appendix) for details. The minimization of the Lagrangian can now be decomposed into a minimization on each link. Note that all the information needed in minimizing $c_e(\vec{f}_e, t_e)$ is local to link e .

The dual optimization problem is

$$(C) \quad \max_{\vec{\mu} \geq 0, \vec{q} \geq 0} D(\vec{\mu}, \vec{q}). \quad (14)$$

As mentioned before, the convexity of $\Theta(\vec{f}, \vec{t})$ can be inferred from Lemma 2. Based on this, the following Theorem [12] establishes the relationship between the primal problem (B) and the dual problem (C) above.

Theorem 3: (Strong Duality)

Problem (B) is feasible and its optimal value Θ^* is finite. Further, there is no duality gap: Θ^* can be found by solving its dual optimization problem.

Proof: We can always choose a large enough f_e^d to satisfy the flow constraint (9), and small enough t_e to satisfy the scheduling constraint (8). Therefore, it is evident that problem (B) is feasible and its optimal value Θ^* is finite.

In fact, we can choose \vec{f}, \vec{t} in such a way that all constraints in (8) and (9) strictly satisfy:

$$\begin{aligned} \sum_{e \in N(v)} t_e &< \beta, \quad \forall v \in V; \\ \sum_{e \in N_o(v)} f_e^d - \sum_{e \in N_i(v)} f_e^d - T_{vd} &> 0, \quad \forall d, \text{ and } \forall v \neq d. \end{aligned}$$

Therefore problem (B) satisfies the Slater constraint qualification [21]. Since this condition is satisfied, and the constraints (8) and (9) are polyhedral in nature, to show that there is no duality gap, it suffices to show the objective function $\Theta(\vec{f}, \vec{t})$ is convex over Y . Note that $\Theta(\vec{f}, \vec{t})$ can be derived by applying multiple vector compositions on $\theta(f, t)$. The convexity of $\Theta(\vec{f}, \vec{t})$ can therefore be inferred from Lemma 2. ■

The next step is to solve the dual problem in a distributed fashion. We can show that $D(\vec{\mu}, \vec{q})$ is convex and its subgradient is given by

$$\begin{aligned} \frac{\partial D}{\partial \mu_v} &= \sum_{e \in N(v)} t_e - \beta, \\ \frac{\partial D}{\partial q_v^d} &= - \left[\sum_{e \in N_o(v)} f_e^d - \sum_{e \in N_i(v)} f_e^d - T_{vd} \right]. \end{aligned}$$

We can then use the following subgradient method to solve the dual problem.

Distributed Energy Minimization Algorithm

At each iteration m ,

- 1) At link e , the data rate \vec{f}_e and the link assignment t_e are determined by:

$$\begin{aligned} &(\vec{f}_e(m), t_e(m)) \\ &= \operatorname{argmin}_{(\vec{f}_e, t_e) \in Y_e} c_e(\vec{f}_e, t_e, m) \\ &= \operatorname{argmin}_{(\vec{f}_e, t_e) \in Y_e} \left[\Theta(\vec{f}_e, t_e) + (\mu_{x(e)}(m) + \mu_{r(e)}(m)) t_e \right. \\ &\quad \left. - \sum_d (q_{x(e)}^d(m) - q_{r(e)}^d(m)) f_e^d \right]. \end{aligned} \quad (15)$$

- 2) At node v , the dual variables are updated by:

$$\begin{aligned} \mu_v(m+1) &= \left\{ \mu_v(m) + \alpha_m \left[\sum_{e \in N(v)} t_e(m) - \beta \right] \right\}^+; \\ q_v^d(m+1) &= \left\{ q_v^d(m) - \alpha_m \left[\sum_{e \in N_o(v)} f_e^d(m) \right. \right. \\ &\quad \left. \left. - \sum_{e \in N_i(v)} f_e^d(m) - T_{vd} \right] \right\}^+. \end{aligned} \quad (16)$$

Remark: At each iteration m , the data rate and schedule are chosen to minimize the total cost c_e , given the current implicit costs $(\vec{\mu}, \vec{q})$. To maximize the dual objective function, the dual variables are updated according to the subgradient ascent procedure in (16) and (17), where $\{\alpha_m\}$ are the stepsizes. In the above algorithm, we use the same stepsize for $\vec{\mu}$ and \vec{f} at each iteration. This is simply for ease of notation: the stepsizes can be different for each dual variable. Since the primal cost function is not strictly convex, the dual function may not be differentiable. The stepsizes need to be chosen carefully to guarantee the convergence of the dual variables. We will discuss this issue in the Section IV.

The above exercise of using Lagrange duality is standard. Nonetheless, there are some questions that are not answered by the above algorithm alone. One of them is how the algorithm maps to different protocol components. We will now study this problem.

D. Mapping to Network Protocol Components

To decide the routing and scheduling, each link solves (15) by minimizing $c_e(\vec{f}_e, t_e, m)$. Note that there are three components in $c_e(\vec{f}_e, t_e, m)$:

- The term $\Theta(\vec{f}_e, t_e)$ is the power cost, i.e., the power consumption of supporting an average data rate of \vec{f}_e while link e being activated for t_e fraction of time.
- The term $(\mu_{x(e)}(m) + \mu_{r(e)}(m))t_e$ is the scheduling cost. We will show that $t_e(m)$ will either be 0 or 1 in order for link e to minimize $c_e(\vec{f}_e, t_e)$. If $t_e(m)$ is chosen to be 1, then $\mu_{x(e)}(m+1) \geq \mu_{x(e)}(m)$ from (16): scheduling the transmission on link e gives rise to increased (or at least the same) scheduling costs on the transmission/reception nodes. If $t_e(m)$ is chosen to be 0, then $\mu_{x(e)}(m+1) \leq \mu_{x(e)}(m)$: an idle link gives the system more flexibility in scheduling other links, which leads to a potential decrease in the scheduling costs on the transmission/reception nodes.
- The term $\sum_d (q_{x(e)}^d(m) - q_{r(e)}^d(m))f_e^d$ is the *utility* of supporting the total data rate on link e . The term q_v^d can be interpreted as an approximation of the scaled version of the queue length at node v for destination d [15] if constant stepsizes are used. As in (17), if $f_e^d(m) > 0$, it helps reduce the workload at the transmission node, but add to the workload at the reception node.

The following transformation leads to a better understanding of the minimization of $c_e(\vec{f}_e, t_e)$:

$$f_e^d(m) = R_e^d(m)t_e(m), \quad (18)$$

where $R_e^d(m)$ is the instantaneous data rate allocated on link e for destination d , and $f_e^d(m)$ is the resultant average data rate when transmitting for $t_e(m)$ fraction of the time. Substitute the above equation into $c_e(\vec{f}_e, t_e, m)$, we have (all time index dropped for ease of notation):

$$c_e(\vec{f}_e, t_e) = t_e l_e(\vec{R}_e), \quad (19)$$

where $l_e(\vec{R}_e)$ is defined as

$$l_e(\vec{R}_e) = h\left(\sum_d R_e^d\right) + (\mu_{x(e)} + \mu_{r(e)}) - \sum_d \left[(q_{x(e)}^d - q_{r(e)}^d) R_e^d \right]. \quad (20)$$

Since $t_e \geq 0$, to minimize (19), we should first minimize $l_e(\vec{R}_e)$ as a function of \vec{R}_e . Note that function $h(\cdot)$ takes as input parameter the sum of the data rates allocated for all the destinations on this link. In other words, from the viewpoint of power consumption, it is indifferent which destination the data rate is allocated for, as long as the total data rate is the same. As a result, the minimum of $l_e(\vec{R}_e)$ is attained when *all the data rates are allocated to the destination with the maximum positive backlog difference*. Let

$$\hat{d} = \operatorname{argmax}_d (q_{x(e)}^d - q_{r(e)}^d), \quad (21)$$

then $R_e^{\hat{d}} > 0$ if $(q_{x(e)}^{\hat{d}} - q_{r(e)}^{\hat{d}}) > 0$, and $R_e^d = 0$ for all $d \neq \hat{d}$. With this observation, the minimization of $l_e(\vec{R}_e)$ is relatively simple. For example, if $h(x) = e^x - 1$, and $\max_d (q_{x(e)}^d - q_{r(e)}^d) > 0$, then $R_e^{\hat{d}} = [\log(q_{x(e)}^{\hat{d}} - q_{r(e)}^{\hat{d}})]^+$, and $R_e^d = 0$ for all other destination d .

Now that \vec{R}_e has been chosen to minimize $l_e(\vec{R}_e)$ in (20), the next step is to determine the value of t_e over the interval $[0, 1]$ to minimize $c_e(\vec{f}_e, t_e) = t_e l_e(\vec{R}_e)$. Clearly, the optimal t_e value is

$$\hat{t}_e = \begin{cases} 1, & \text{if } \min_{\vec{R}_e} l_e(\vec{R}_e) \leq 0, \\ 0, & \text{if } \min_{\vec{R}_e} l_e(\vec{R}_e) > 0. \end{cases} \quad (22)$$

Remark: There are two possible scenarios where $\hat{t}_e = 0$ (no load is assigned in this time-slot):

- All backlog differences $(q_{x(e)}^d - q_{r(e)}^d)$ are negative, which means the utility does not increase by transporting data to next hop on this link for any destination.
- Although some backlog differences $(q_{x(e)}^d - q_{r(e)}^d)$ are positive, as in (20), the *utility of transporting data to next hop* is not large enough to outweigh the *importance of saving more energy* (the $h(\sum_d R_e^d)$ term) and/or the *need to gain more flexibility in scheduling* (the $(\mu_{x(e)} + \mu_{r(e)})$ term). As a result, optimal $l_e(\vec{R}_e)$ is non-negative, and no load is assigned to this link.

To summarize, the minimization of $c_e(\vec{f}_e, t_e)$ on each link naturally translates into the following protocol components:

- 1) *Routing:* Choose only the flow \hat{d} with maximum positive backlog difference (cf. (21)). This is the flow that should receive service.
- 2) *Power control:* Choose \vec{R}_e to minimize $l_e(\vec{R}_e)$. This is the power link e should use.
- 3) *Link assignment:* Choose t_e to minimize $t_e l_e(\vec{R}_e)$ in such a way that t_e takes its maximum value 1 if the optimal $l_e(\vec{R}_e)$ is less than or equal to 0; and 0 otherwise. This determines the amount of time link e should be on.
- 4) *Maximal Scheduling:* As mentioned before, given the fraction of up-time, or load, on each link, a scheduling component is needed to determine the exact time-slots link e should be on. In other words, the link assignment from previous step may have scheduling conflicts, and therefore cannot be used as a transmission schedule as it is. A scheduling component can resolve such conflicts by delaying the transmission by a certain number of time-slots. Clearly, given the link assignment $\vec{t}(m)$, the delay can be different for different links. With the node-exclusive interference model, by choosing $\beta = \frac{1}{2} - \eta$, it can be shown [7], [24] that a scheduling policy based on maximal matching can stabilize the system (the delay of transmission mentioned above is bounded). Maximal schedules such as this can be implemented in a distributed fashion. We refer readers to [16] for more details on the distributed implementation of maximal schedules.

As we have seen thus far, the duality approach exploits the problem structure and decomposes the primal problem into sub-problems on each link that are much simpler. In addition, some of the quantities produced/monitored by this algorithm actually help network engineering. For instance, the Lagrange multiplier q_v^d can also be interpreted as the *shadow price* of the corresponding constraint. If a small change occurs in the amount of supported traffic from node v to node d , q_v^d measures the sensitivity of the optimal power consumption with respect to this perturbation. In a different networking setting where the network has some control over the traffic matrix \vec{T} , information such as \vec{q} can be used as guidelines to optimize power consumption.

IV. PERFORMANCE ANALYSIS

In this section, first we are interested in the following question: *Under what condition do the dual variables in the Distributed Energy Minimization algorithm converge?*

Furthermore, as we have seen in Section III, there is no guarantee that $(\vec{f}(m), \vec{t}(m))$ will converge, even if the dual variables converge. For example, $t_e(m)$ is either 0 or 1 from the distributed dual algorithm. A natural question to ask then is the following: *In what sense is the primal solution optimal?* The two Theorems below answer both of the above questions from two different perspectives.

A. Convergence Result with Constant Stepsizes

The following theorem establishes the stability and the optimality of the proposed dual algorithm.

Theorem 4: (Stability and Optimality)

Let the stepsizes in the Distributed Energy Minimization algorithm be constant, i.e., $\alpha_m = h$, for all m . Let Φ be the set of $(\vec{\mu}, \vec{q})$ that maximizes $D(\vec{\mu}, \vec{q})$, and $d((\vec{\mu}, \vec{q}), \Phi) = \min_{(\vec{\mu}^*, \vec{q}^*) \in \Phi} \|(\vec{\mu}, \vec{q}) - (\vec{\mu}^*, \vec{q}^*)\|$. Given any $\varepsilon > 0$, there exists some $h_0 > 0$ such that, for any $h \leq h_0$ and any initial implicit costs $(\vec{\mu}_0, \vec{q}_0)$, there exists a time M_0 such that for all $m > M_0$,

$$d((\vec{\mu}(m), \vec{q}(m)), \Phi) < \varepsilon, \quad (23)$$

and

$$\limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{\tau=1}^m \sum_{e \in E} \Theta(\vec{f}_e(\tau), t_e(\tau)) < \Theta^* + \varepsilon. \quad (24)$$

The proof of Theorem 4 is included in the Appendix.

The above Theorem shows that, when stepsizes are small, the dual variables eventually converge to within a small neighborhood of the optimal dual solution. Further, the power consumption from the distributed algorithm is asymptotically optimal: although the primal variables $(\vec{f}(m), \vec{t}(m))$ may not converge, by using $(\vec{f}(m), \vec{t}(m))$ for each time-slot m , the long-term average of the resultant power consumption is arbitrarily close to the optimal power consumption.

B. Optimality of Primal Variables

As shown in Theorem 4, with constant stepsizes, the dual variables eventually converge to within a small neighborhood of the optimal dual solution. When appropriate diminishing stepsizes are used in the dual algorithm, the following Theorem shows that the entire sequence of the dual variables converges to one point. In this case, we also obtain a different interpretation of the primal optimality.

Theorem 5: (Primal Optimality with Diminishing Stepsizes)

(a) Let the stepsizes $\{\alpha_m\}$ in the Distributed Energy Minimization algorithm satisfy the following conditions:

$$\sum_{m=1}^{\infty} \alpha_m^2 < +\infty, \quad \sum_{m=1}^{\infty} \alpha_m = +\infty. \quad (25)$$

Then for any nonnegative starting point $(\vec{\mu}_0, \vec{q}_0)$, the dynamics of the Distributed Energy Minimization algorithm converges to the optimal value of Problem (B):

$$\lim_{m \rightarrow \infty} (\vec{\mu}(m), \vec{q}(m)) = (\vec{\mu}^*, \vec{q}^*), \quad (26)$$

$$\lim_{m \rightarrow \infty} D(\vec{\mu}(m), \vec{q}(m)) = \Theta^*, \quad (27)$$

where $(\vec{\mu}^*, \vec{q}^*)$ is a maximum point of $D(\vec{\mu}, \vec{q})$.

(b) Let the stepsizes $\{\alpha_m\}$ be chosen as

$$\alpha_m = \frac{\kappa}{m + \rho}, \quad (28)$$

where κ and ρ are some positive scalars (note that the above definition of $\{\alpha_m\}$ satisfies the conditions (25)). If the long-term average of the vector $(\vec{f}(m), \vec{t}(m))$ converges, i.e.,

$$\lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \vec{f}(m) = \vec{f}^*, \quad \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \vec{t}(m) = \vec{t}^*, \quad (29)$$

then the time-average version of $(\vec{f}(m), \vec{t}(m))$ is the optimal solution to Problem (B): $\sum_{e \in E} \Theta(\vec{f}_e^*, \vec{t}_e^*) = \Theta^*$.

The proof of Theorem 5 is included in the Appendix.

Remark: In the Distributed Energy Minimization algorithm, $t_e(m)$, for example, can only be 0 or 1. From this perspective, our solution carries the same flavor as some of the related work [8], [14], namely, in the optimal solution, *the resource, be it power or fraction of up-time, is used to the maximum extent, if the link is activated for the current time-slot.* However, one would expect the optimal solution of t_e to Problem (B) to be a number anywhere from 0 to β for most of the links in a typical setting (cf. (8)). Theorem 5 reconciles the difference between these two viewpoints. Although $(\vec{f}(m), \vec{t}(m))$ is not a continuous mapping from the underlying converging implicit costs (dual variables $(\vec{\mu}, \vec{q})$), as long as the long-term average of primal variables $(\vec{f}(m), \vec{t}(m))$ converges, the limit is an optimal solution to the primal problem.

C. Power-Efficiency Ratio

Setting $\beta = 1$ or $\beta = \frac{1}{2} - \eta$ in (4) gives necessary or sufficient conditions for schedulability, in terms of stabilizing the overall queueing system [9]. From the necessary condition of $\beta = 1$, it is evident that the maximum loss in throughput under the node exclusive interference model is $\frac{1}{2} - \eta$. This is the throughput loss ratio in approximating (*) using Problem (B) with $\beta = \frac{1}{2} - \eta$.

To derive the approximation ratio of our algorithm, the throughput loss needs to be translated into power loss. We apply the same first order approximation of the rate-power function as in [4]. More specifically, in an AWGN channel, the total power consumption can be approximated as $\sum_e \frac{B_e}{t_e}$, where B_e is a flow-related constant independent of t_e . Let (\vec{f}, \vec{t}) be the optimal solution to Problem (B) with $\beta = 1$. It is evident that $(\vec{f}, \frac{\vec{t}}{2+\varepsilon})$, where ε is a small positive constant, is a feasible solution to Problem (B) with $\beta = \frac{1}{2} - \eta$. This feasible solution results in a power consumption that is at most $(2+\varepsilon)$ that of the optimal value with $\beta = 1$. Since the optimal value of Problem (B) with $\beta = 1$ is a lower bound on the minimum power from (*), we conclude that the power-efficiency ratio of our algorithm is upper-bounded by $(2+\varepsilon)$.

V. NUMERICAL RESULTS

In this illustrative example, we consider a 7-node network, whose topology is depicted in Figure 1. The power-rate function is of the following form:

$$R_e = W \log_2 \left[1 + \frac{\sigma_e p_e}{N_0 W} \right],$$

where $W = 1.0$ MHz is the available bandwidth, $\sigma_e = 1.6 \times 10^{-13}$ is the channel gain of link e , $N_0 = 1.6 \times 10^{-18}$ mW/Hz is the noise spectral density, p_e is the transmission power, and R_e is the resultant instantaneous data rate of link e . This network supports two flows, as shown in Table I.

TABLE I
THE TWO FLOWS SUPPORTED BY THE NETWORK

	source	destination	data rate	paths
flow 1	1	7	250 kbps	1-7, 1-2-7
flow 2	3	6	500 kbps	3-2-6, 3-4-5-6

The node-exclusive interference model is considered, and $\beta = (0.5 - 10^{-4})$ in Problem (A). The length of each time-slot is 1 second. The results reported in this section are the average over a moving time window of length 120 seconds.

To show that the joint energy minimization algorithm can adapt to variations in the input parameters, we apply the following changes in the system setting. At time $t = 4000$ s, the channel gain $\sigma_{(1,7)}$ of the direct link between node 1 and node 7 is decreased from 1.6×10^{-13} to 0.4×10^{-13} . At time $t = 8000$ s, the data rate of flow 2 (from node 3 to node 6) is reduced from 500 kbps to 250 kbps.

For each setting, offline computation is carried out to find the optimal value of Problem (A), which is given by the dashed line in Figure 2. The power consumption from the

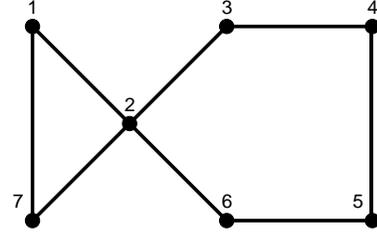


Fig. 1. Network topology

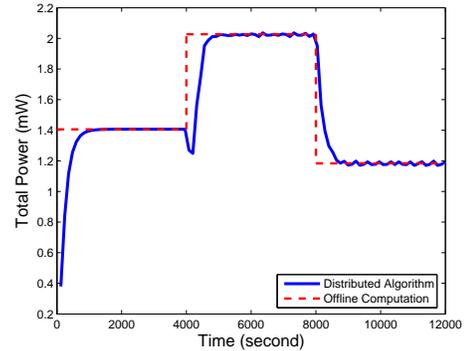


Fig. 2. Power consumption from distributed algorithm and offline computation

proposed algorithm is shown as the solid line in the same figure. This simulation result shows that the Distributed Energy Minimization algorithm is capable of computing the optimal solution in a distributed manner, and automatically tracking the optimal operating point once the system parameters change.

Figure 3 shows the average data rates f_e^d for different flows on four links. We now take a closer look at the routing of the flows.

- In the initial state, flow 1 concentrates on the minimum energy path, namely, link (1, 7).
- At $t = 4000$ s, the channel gain $\sigma_{(1,7)}$ reduces by 75%, and part of flow 1 is shifted to path 1 – 2 – 7. Since the scheduling capacity of node 2 is saturated, a larger percentage of flow 2 is then routed through path 3 – 4 – 5 – 6 in the optimal solution.
- At $t = 8000$ s, the traffic that the network has to support between node 3 and node 6 reduces (flow 2 is reduced to 250 kbps). As a consequence, part of the scheduling capacity of node 2 is freed, and more of flow 1 takes path 1 – 2 – 7 to reduce the overall power consumption.

The above example shows that the interaction between routing, scheduling, and power control is relatively complex even in a wireless network of small size. The correct way to deal with such interaction is difficult to summarize into

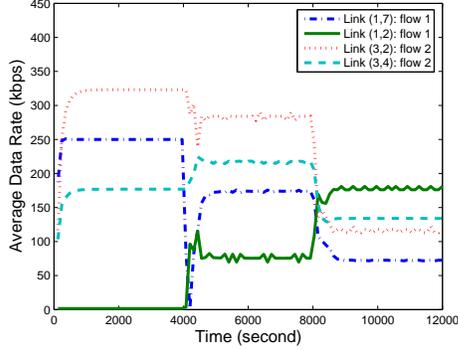


Fig. 3. Average data rates for different flows on four links

general approaches such as minimum energy approach or load balancing approach. In our Distributed Energy Minimization algorithm, as shown in Figure 2 and Figure 3, low-complexity and distributed operation on each link accomplishes this joint optimization even in networks with non-stationarity.

VI. CONCLUSION

In this paper, we propose a joint power control, link scheduling, and routing algorithm to minimize the power consumption in multi-hop wireless networks. The known cross-layer solution to this problem is centralized, and with high computational complexity. In contrast, our algorithm is distributed, and with low computational complexity. We establish the power efficiency ratio of our solution, and show this performance bound, achieved in a distributed manner, is provably tighter than a centralized solution in the literature. We map our solution to corresponding components of the cross-layer control protocols, and discuss the implication of our result on network protocol design. The stability and optimality of our solution is studied, and the simulation results verify that our distributed solution computes and tracks the optimal operating point when the system parameters change.

In this work, we focus on the well-studied node exclusive interference model. In a more general interference model, we can define Γ_e to be the set of links that interfere with link e . (The node-exclusive model can be viewed as a special case with $\Gamma_e = N(x(e)) \cup N(r(e))$.) Let ω_e denote the maximum number of links that can be scheduled simultaneously in Γ_e , and $\omega_{\max} = \max_e \omega_e$. It can be shown that a distributed maximal schedule can stabilize the system while the throughput is reduced by at most a factor of $1/\omega_{\max}$. In this case, the left hand side of the scheduling constraint in our problem formulation need to be modified to be a summation over all links incident with one or more nodes in Γ_e [6], [7], [23], [24]. As a result, the relationship of dual variables $\bar{\mu}$ and the set of nodes V can be a many-to-many mapping, instead of a one-to-one mapping in the node exclusive model. Nonetheless, since all the information needed to solve the decomposed problem is still local to each link, the general methodology presented in this paper can be carried through to this type of interference

model.

APPENDIX

A. Proof of Lemma 2:

To begin with, we make the following two observations:

First, if f/t equals to some fixed $\tau > 0$, then $\theta(f, t)$ is proportional to t .

Secondly, given any $t > 0$, it is evident that $\theta(f, t)$ is convex in f .

To show that $\theta(f, t)$ is convex in C , we want to show

$$\begin{aligned} & \alpha\theta(f_1, t_1) + (1 - \alpha)\theta(f_2, t_2) \\ & \geq \theta(\alpha f_1 + (1 - \alpha)f_2, \alpha t_1 + (1 - \alpha)t_2), \end{aligned} \quad (30)$$

for any $\alpha \in [0, 1]$, and any (f_1, t_1) and (f_2, t_2) in C .

If t_1 or t_2 equals zero, then Inequality (30) holds. To see that, without loss of generality, let us assume $t_1 = 0$. To avoid trivial case, let us also assume that $t_2 > 0$ and $\alpha \neq 1$. Then $\theta(f_1, t_1) = 0$ by definition. It follows that

$$\begin{aligned} & \alpha\theta(f_1, t_1) + (1 - \alpha)\theta(f_2, t_2) \\ & = (1 - \alpha)\theta(f_2, t_2) \\ & = (1 - \alpha)t_2 h\left(\frac{f_2}{t_2}\right) \\ & = (1 - \alpha)t_2 h\left(\frac{(1 - \alpha)f_2}{(1 - \alpha)t_2}\right) \\ & = \theta((1 - \alpha)f_2, (1 - \alpha)t_2) \\ & = \theta(\alpha f_1 + (1 - \alpha)f_2, \alpha t_1 + (1 - \alpha)t_2). \end{aligned}$$

The last equality holds since $t_1 = 0$ and $(f_1, t_1) \in C$ imply $f_1 = 0$. Therefore, Inequality (30) holds in this case.

From now on, it is assumed that both t_1 and t_2 are positive. Consider the point $(t_1 f_2 / t_2, t_1)$. This point is chosen in such a way that it has the same t -coordinate as (f_1, t_1) and is on the same direction as vector (f_2, t_2) . From the convexity of $h(\cdot)$, we have

$$\begin{aligned} & \lambda\theta(f_1, t_1) + (1 - \lambda)\theta\left(\frac{t_1 f_2}{t_2}, t_1\right) \\ & = t_1 \left[\lambda h(f_1, t_1) + (1 - \lambda)h\left(\frac{t_1 f_2}{t_2}, t_1\right) \right] \\ & \geq t_1 h(\bar{f}, t_1) \\ & = \theta(\bar{f}, t_1), \end{aligned} \quad (31)$$

where $\bar{f} = \lambda f_1 + (1 - \lambda)\frac{t_1 f_2}{t_2}$, for any $\lambda \in [0, 1]$.

The next step is to select the correct value of λ to show the convexity of $\theta(\cdot)$ using the convexity of $h(\cdot)$. Given any $\alpha \in [0, 1]$, λ is chosen to be

$$\lambda = \frac{\alpha t_1}{\alpha t_1 + (1 - \alpha)t_2}.$$

The value of λ is chosen this way such that (\bar{f}, t_1) lies in the direction of vector (\hat{f}, \hat{t}) , where (\hat{f}, \hat{t}) is the convex combination of (f_1, t_1) and (f_2, t_2) with weight parameter α ,

as in the RHS of (30). By substituting the value of λ into (31), and simplifying the expression, we have

$$\begin{aligned} & \alpha t_1 \theta(f_1, t_1) + (1 - \alpha) t_2 t_1 h \left(\frac{f_2}{t_2} \right) \\ \geq & (\alpha t_1 + (1 - \alpha) t_2) t_1 h \left(\frac{\alpha f_1 + (1 - \alpha) f_2}{\alpha t_1 + (1 - \alpha) t_2} \right). \end{aligned}$$

Inequality (30) then follows. \blacksquare

B. Proof of Theorem 4:

The proof technique here is similar to [18].

We first show the boundedness of the subgradient of the dual objective function. In (13), we choose $H \geq |V|^2 \max_{v \neq d} T_{vd}$, where $|V|$ is the number of nodes in the network. Let \bar{Y} denote the Cartesian product of all Y_e in (13). It is clear that $\bar{Y} \subset Y$, where Y is defined in Problem (B). Note that the introduction of H does not change the solution. The optimal solution (\vec{f}^*, \vec{t}^*) to Problem (B) is in \bar{Y} for the following reason: Θ is a non-decreasing function of f_e^d , and therefore any (\vec{f}, \vec{t}) with $f_e^d > H$ for some d and e is clearly sub-optimal even if all the traffic is routed through link e . So the solution generated from the Distributed Energy Minimization algorithm coincides with the optimal solution to Problem (B). The benefit from enforcing such an upper bound on f_e^d is that the subgradient of the dual objective function is then bounded.

Given the subgradient of the dual objective function is bounded, it can be shown [3] that, given any $\varepsilon > 0$, there exists some $h_1 > 0$ such that, for any $h \leq h_1$ and any initial implicit costs $(\vec{\mu}_0, \vec{q}_0)$, there exists a time M_0 such that (23) holds for all $m > M_0$. Therefore, as time index m increases, $(\vec{\mu}(m), \vec{q}(m))$ converges to within a small neighborhood of the set of the maximizer of $D(\vec{\mu}, \vec{q})$ if the constant stepsize is chosen to be small enough. This implies that the sequence $\{(\vec{\mu}(m), \vec{q}(m))\}_m$ is bounded.

To show (24), we consider the following Lyapunov function:

$$V(\vec{\mu}(m), \vec{q}(m)) = \frac{1}{2} \sum_{v,d} [q_v^d(m)]^2 + \frac{1}{2} \sum_v \mu_v^2(m).$$

The subgradient of $D(\vec{\mu}, \vec{q})$ at time-slot m can be written as

$$\begin{aligned} \Delta \mu_v(m) &= \sum_{e \in N(v)} t_e(m) - \beta, \\ \Delta q_v^d(m) &= - \left[\sum_{e \in N_o(v)} f_e^d(m) - \sum_{e \in N_i(v)} f_e^d(m) - T_{vd} \right]. \end{aligned}$$

Using the above notation, the one-step drift of the Lyapunov

function can be calculated as follows:

$$\begin{aligned} & V(\vec{\mu}(m+1), \vec{q}(m+1)) - V(\vec{\mu}(m), \vec{q}(m)) \\ &= \frac{1}{2} \sum_v \left\{ [(\mu_v(m) + h \Delta \mu_v(m))^+]^2 - \mu_v^2(m) \right\} \\ & \quad + \frac{1}{2} \sum_{v,d} \left\{ [(q_v^d(m) + h \Delta q_v^d(m))^+]^2 - [q_v^d(m)]^2 \right\} \\ & \quad \text{(from (16) and (17))} \\ & \leq \frac{1}{2} \sum_v \left\{ [\mu_v(m) + h \Delta \mu_v(m)]^2 - \mu_v^2(m) \right\} \\ & \quad + \frac{1}{2} \sum_{v,d} \left\{ [q_v^d(m) + h \Delta q_v^d(m)]^2 - [q_v^d(m)]^2 \right\} \\ & \leq h \sum_v \mu_v(m) \Delta \mu_v(m) + h \sum_{v,d} q_v^d(m) \Delta q_v^d(m) + h^2 W, \end{aligned}$$

where W is a constant large enough to guarantee the last inequality. It is possible to choose such a constant since the subgradient of $D(\vec{\mu}, \vec{q})$ is bounded. Adding $h \sum_{e \in E} \Theta(\vec{f}_e(m), t_e(m))$ to both sides of the above formula, we have

$$\begin{aligned} & V(\vec{\mu}(m+1), \vec{q}(m+1)) - V(\vec{\mu}(m), \vec{q}(m)) \\ & \quad + h \sum_{e \in E} \Theta(\vec{f}_e(m), t_e(m)) \\ & \leq h \left[\sum_{e \in E} \Theta(\vec{f}_e(m), t_e(m)) + \sum_v \mu_v(m) \Delta \mu_v(m) \right. \\ & \quad \left. + \sum_{v,d} q_v^d(m) \Delta q_v^d(m) \right] + h^2 W \\ & = h D(\vec{\mu}(m), \vec{q}(m)) + h^2 W \\ & \quad \text{(from (12) and (15))} \\ & \leq h D^* + h^2 W \\ & = h \Theta^* + h^2 W, \end{aligned}$$

where D^* is the maximum dual value, which is also equal to the minimum power Θ^* from the primal problem.

Summing the above inequality over $m = 1, 2, \dots, M$, and dividing both sides by M , we have

$$\begin{aligned} & \frac{V(\vec{\mu}(M+1), \vec{q}(M+1)) - V(\vec{\mu}(1), \vec{q}(1))}{M} \\ & \quad + \frac{h}{M} \sum_{m=1}^M \sum_{e \in E} \Theta(\vec{f}_e(m), t_e(m)) \\ & \leq h \Theta^* + h^2 W. \end{aligned} \tag{32}$$

Note that all $V(\cdot)$ is bounded since the sequence $\{(\vec{\mu}(m), \vec{q}(m))\}_m$ is bounded. For any given $\varepsilon > 0$, the first term on the LHS of (32) can be bounded by $\varepsilon/2$ for M large enough:

$$\frac{1}{M} \sum_{m=1}^M \sum_{e \in E} \Theta(\vec{f}_e(m), t_e(m)) \leq \Theta^* + hW + \frac{\varepsilon}{2}.$$

For any given $\varepsilon > 0$, there exist some $h_2 > 0$ such that $hW < \varepsilon/2$ for all $h < h_2$. Under this condition, the following is evident:

$$\limsup_{M \rightarrow \infty} \frac{1}{M} \sum_{m=1}^M \sum_{e \in E} \Theta(\vec{f}_e(m), t_e(m)) < \Theta^* + \varepsilon.$$

Let $h_0 = \min(h_1, h_2)$, and Theorem 4 follows. \blacksquare

C. Proof of Theorem 5:

(a) As in the beginning of the proof of Theorem 4, we can show the boundedness of the subgradient. We then invoke Proposition 8.2.6 from [3], and by the diminishing stepsize rule (25), we have (26) and

$$\lim_{m \rightarrow \infty} D(\mu(\vec{m}), \vec{q}(m)) = D^* = \Theta^*, \quad (33)$$

where the last equality follows from Theorem 3.

(b) Define the average fraction of up-time of link e over the first m time slots as

$$t_e^{\text{avg}}(m) = \frac{1}{m} \sum_{\tau=1}^m t_e(\tau).$$

Similarly, define the average data rate of link e , dedicated for destination d , over the first m time slots as

$$f_e^{d,\text{avg}}(m) = \frac{1}{m} \sum_{\tau=1}^m f_e^d(\tau).$$

Let $\varepsilon > 0$ be a given arbitrarily small number. Since $\lim_{m \rightarrow \infty} (\vec{\mu}_m, \vec{q}_m) = (\vec{\mu}^*, \vec{q}^*)$, and $\alpha_m = \frac{\kappa \rho}{m+\rho} > \frac{\kappa}{2m}$ for large m (from (25)), we can choose M_0 large enough such that for all $m \geq M_0$, the following holds:

$$|q_v(m) - q_v^*| \leq \frac{\kappa \varepsilon}{12}, \quad (34)$$

$$\alpha_m > \frac{\kappa}{2m}, \quad (35)$$

for any node v .

Consider link e , whose transmission node $x(e)$ or reception node $r(e)$ satisfies $q_v^{d,*} > 0$ for some d .

For large M , clearly $q_v^d(m) > 0$ for all $m \geq M$. Therefore, the $(\dots)^+$ from (17) has no effect when m is large:

$$-\alpha_m \left(\sum_{e \in N_o(v)} f_e^d(m) - \sum_{e \in N_i(v)} f_e^d(m) - T_{vd} \right) = q_v^d(m+1) - q_v^d(m).$$

Sum the above equation for m ranging from M_0 to some $M_1 > M_0$:

$$\begin{aligned} & \sum_{m=M_0}^{M_1} \left[-\alpha_m \left(\sum_{e \in N_o(v)} f_e^d(m) - \sum_{e \in N_i(v)} f_e^d(m) - T_{vd} \right) \right] \\ &= q_v^d(M_1 + 1) - q_v^d(M_0). \end{aligned}$$

Since $\alpha_m > \frac{\kappa}{2m} \geq \frac{\kappa}{2M_1}$ and (34), we have

$$\begin{aligned} & \frac{\kappa}{2M_1} \left| \sum_{m=M_0}^{M_1} \left(\sum_{e \in N_o(v)} f_e^d(m) - \sum_{e \in N_i(v)} f_e^d(m) - T_{vd} \right) \right| \\ & \leq \left| \sum_{m=M_0}^{M_1} [\alpha_m \left(\sum_{e \in N_o(v)} f_e^d(m) - \sum_{e \in N_i(v)} f_e^d(m) - T_{vd} \right)] \right| \\ & \leq |q_v^d(M_1 + 1) - q_v^d(M_0)| \\ & \leq \frac{\kappa \varepsilon}{6}. \end{aligned} \quad (36)$$

Exchange the order of summation on the LHS of (36), we arrive at

$$\left| \tilde{\Delta}_{vd}(M_0, M_1) - \tilde{T}_{vd}(M_0, M_1) \right| \leq \frac{\varepsilon}{3}, \quad (37)$$

where

$$\tilde{\Delta}_{vd}(M_0, M_1) = \sum_{e \in N_o(v)} \tilde{f}_e^d(M_0, M_1) - \sum_{e \in N_i(v)} \tilde{f}_e^d(M_0, M_1),$$

$$\tilde{f}_e^d(M_0, M_1) = \frac{1}{M_1} \sum_{m=M_0}^{M_1} f_e^d(m),$$

$$\tilde{T}_{vd}(M_0, M_1) = T_{vd} \frac{M_1 - M_0 + 1}{M_1}.$$

Given M_0 ,

$$\lim_{M_1 \rightarrow \infty} \tilde{\Delta}_{vd}(M_0, M_1) = \sum_{e \in N_o(v)} f_e^{d,*} - \sum_{e \in N_i(v)} f_e^{d,*}, \quad (38)$$

$$\lim_{M_1 \rightarrow \infty} \tilde{T}_{vd}(M_0, M_1) = T_{vd}. \quad (39)$$

From (39) and (38), we can choose M_1 large enough such that

$$\begin{aligned} & \left| \tilde{\Delta}_{vd}(M_0, M_1) - \left[\sum_{e \in N_o(v)} f_e^{d,*} - \sum_{e \in N_i(v)} f_e^{d,*} \right] \right| \leq \frac{\varepsilon}{3}, \\ & \left| \tilde{T}_{vd}(M_0, M_1) - T_{vd} \right| \leq \frac{\varepsilon}{3}. \end{aligned}$$

From (37), we have

$$\left| \sum_{e \in N_o(v)} f_e^{d,*} - \sum_{e \in N_i(v)} f_e^{d,*} - T_{vd} \right| \leq \varepsilon.$$

Since the above is true for any given $\varepsilon > 0$, we conclude that

$$\sum_{e \in N_o(v)} f_e^{d,*} - \sum_{e \in N_i(v)} f_e^{d,*} = T_{vd}, \quad (40)$$

for all e satisfying $q_{x(e)}^{d,*} > 0$ or $q_{r(e)}^{d,*} > 0$ for some d . In other words, all the constraints on \vec{f} are active for $q_v^d > 0$. This is known as the *complimentary slackness* condition. Following the same approach, the complimentary slackness can be shown for constraints on \vec{t} :

$$\sum_{e \in N(v)} t_e^* = \beta, \quad (41)$$

for all $e \in N(v)$, $\mu_v^* > 0$.

Furthermore, since the vector $(\vec{f}(m), \vec{t}(m))$ minimize $L(\vec{f}, \vec{t}, \vec{\mu}(m), \vec{q}(m))$ for each time-slot m in the Dual Energy Minimization algorithm, it is easy to show

$$(\vec{f}^*, \vec{t}^*) = \underset{(\vec{f}, \vec{t})}{\operatorname{argmin}} L(\vec{f}, \vec{t}, \vec{\mu}^*, \vec{q}^*). \quad (42)$$

Therefore,

$$\begin{aligned} & \sum_{e \in E} \Theta(\vec{f}_e^*, t_e^*) \\ &= \sum_{e \in E} \Theta(\vec{f}_e^*, t_e^*) + \sum_{v \in V} \mu_v^* \left(\sum_{e \in N(v)} t_e^* - \beta \right) \\ & \quad - \sum_{v,d} q_v^{d,*} \left(\sum_{e \in N_o(v)} f_e^{d,*} - \sum_{e \in N_i(v)} f_e^{d,*} - T_{vd} \right), \\ & \quad (\text{from complimentary slackness (40) and (41)}) \\ &= L(\vec{f}^*, \vec{t}^*, \vec{\mu}^*, \vec{q}^*) \\ &= \min_{(\vec{f}, \vec{t})} L(\vec{f}, \vec{t}, \vec{\mu}^*, \vec{q}^*) \\ & \quad (\text{from (42)}) \\ &= D(\vec{\mu}^*, \vec{q}^*) \\ &= \Theta^*. \end{aligned}$$

The last equality is true from the strong duality in Theorem 3.

Note that $\{(f_e^*, t_e^*)\}$ is also prime feasible, otherwise $(\vec{\mu}^*, \vec{f}^*)$ will diverge in view of (16), (17), and the fact that $\sum_{m=1}^{\infty} \alpha_m = +\infty$. ■

REFERENCES

- [1] I. F. Akyildiz, X. Wang, and W. Wang. Wireless mesh networks: a survey. *Comput. Netw. ISDN Syst.*, 47(4):445–487, 2005.
- [2] Intel Research Laboratory at Berkeley. Habitat monitoring on great duck island, 2002. Available on <http://www.greatduckisland.net/>.
- [3] D. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Nashua, NH, 2003.
- [4] R. Bhatia and M. Kodialam. On power efficient communication over multi-hop wireless networks: Joint routing, scheduling and power control. In *IEEE INFOCOM'04*, Hong Kong, March 2004.
- [5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- [6] P. Chaporkar, K. Kar, and S. Sarkar. Throughput Guarantees in Maximal Scheduling in Wireless Networks. In *Proceedings of 43d Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, September 2005.
- [7] P. Chaporkar, K. Kar, and S. Sarkar. Achieving Queue Length Stability Through Maximal Scheduling in Wireless Networks. In *Proceedings of Information Theory and Applications Inaugural Workshop*, University of California, San Diego, February 2006.
- [8] R. L. Cruz and A. V. Santhanam. Optimal Routing, Link Scheduling and Power Control in Multi-hop Wireless Networks. In *Proceedings of IEEE INFOCOM*, San Francisco, April 2003.
- [9] B. Hajek and G. Sasaki. Link Scheduling in Polynomial Time. *IEEE Transactions on Information Theory*, 34(5):910–917, September 1988.
- [10] K. Kar, M. Kodialam, T. V. Lakshman, and L. Tassiulas. Routing for network capacity maximization in energy-constrained ad-hoc networks. *IEEE INFOCOM'03*, 2003.
- [11] Q. Li, J. A. Aslam, and D. Rus. Online power-aware routing in wireless ad-hoc networks. In *Proc. the Seventh Annual International Conference on Mobile Computing and Networking (ACM Mobicom'01)*, pages 97–107, 2001.
- [12] L. Lin, X. Lin, and N. B. Shroff. Low-complexity and distributed energy minimization in multi-hop wireless networks. Technical report, Purdue University, 2006. Available on <http://web.ics.purdue.edu/~llin/paper/tech06.pdf>.
- [13] L. Lin, N. B. Shroff, and R. Srikant. Energy-aware routing in sensor networks: A large systems approach. In *the Third Annual Conference on Wireless On demand Network Systems and Services (WONS)*, Les Menuires, France, January 2006.
- [14] X. Lin and N. B. Shroff. The Impact of Imperfect Scheduling on Cross-Layer Rate Control in Multihop Wireless Networks. In *Proceedings of IEEE INFOCOM*, Miami, FL, March 2005.
- [15] X. Lin and N. B. Shroff. Simplification of network dynamics in large systems. *IEEE/ACM Transactions on Networking*, October 2005.
- [16] X. Lin, N. B. Shroff, and R. Srikant. A Tutorial on Cross-Layer Optimization in Wireless Networks. *IEEE Journal on Selected Areas in Communications*, 2006, to appear.
- [17] J. Luo and J.-P. Hubaux. A survey of inter-vehicle communication. Technical report, EPFL, Lausanne, Switzerland, 2004.
- [18] M. J. Neely. Energy Optimal Control for Time Varying Wireless Networks. *IEEE Transactions on Information Theory*, 52(2):2915–2934, July 2006.
- [19] S. Rasool and X. Lin. Distributed Opportunistic Scheduling in Multi-Carrier Wireless Networks. *to be submitted to IEEE INFOCOM*, 2006.
- [20] T. P. Ruggaber and J. W. Talley. Detection and control of combined sewer overflow events using embedded sensor network technology. In *World Water Congress*, Anchorage, Alaska, May 2005.
- [21] M. Slater. Lagrange multipliers revisited. Cowles foundation discussion papers, mathematics 403, Cowles Foundation, Yale University, 1950.
- [22] C.-K. Toh. Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks. *IEEE communications Magazine*, 39(6):138–147, June 2001.
- [23] X. Wu and R. Srikant. Bounds on the Capacity Region of Multihop Wireless Networks Under Distributed Greedy Scheduling. In *Proceedings of IEEE INFOCOM*, Barcelona, Spain, April 2006.
- [24] X. Wu, R. Srikant, and J. R. Perkins. Queue-Length Stability of Maximal Greedy Schedules in Wireless Network. In *Proceedings of Information Theory and Applications Inaugural Workshop*, University of California, San Diego, February 2006.
- [25] Y. Xue, Y. Cui, and K. Nahrstedt. A utility-based distributed maximum lifetime routing algorithm for wireless networks. In *QSHINE*, Los Alamitos, CA, USA, 2005.