2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)

Optimizing Queueing Cost and Energy Efficiency in a Wireless Network

Antonis Dimakis

Department of Informatics
Athens University of Economics and Business
Patission 76, Athens 10434, Greece
Email: dimakis@aueb.gr

Spyros Papafragkos

Department of Informatics
Athens University of Economics and Business
Patission 76, Athens 10434, Greece
Email: papafrag@aueb.gr

***Abstract*—We consider the joint minimization of queueing cost and power consumption in a wireless network, over all power control policies at the transmitters. Approximately optimal policies are derived through the use of a comparison to a single link system offering lower bounds for both performance metrics. The derived policies are of a simple form: the total power consumed at any instant is a monotonic function of a linear combination of the queue backlogs at that instant, while the portions of power allocated to each transmitter are such that the resulting link speeds are proportional to the queue backlogs. Simulations show that the lower bounds become tight -and so the policies become optimal- as the average power approaches the lowest value for which stability is maintained. Moreover, these policies are shown to perform better than gradient projection -based power adaptive algorithms such those arising in formulations using static optimization problems.**

I. Introduction

The efficient operation of a wireless network of nodes, whether in the cellular or the adhoc case presents a multitude of challenges to the network engineer. In contrast to the traditional wireline networks where network dimensioning is not considered a part of network protocol design, even if we disregard the intrinsic randomness of the wireless medium, the capacity of wireless networks constantly depends on decisions taken by multiple layers and by different users in real-time. It is for the network protocols to both realize and manage this capacity in an efficient and simple manner. Thus traditional quality of service metrics such as queueing delay as well as system-level design metrics such as power consumption are not only both considered important in the wireless setting, but they need to be addressed during operation.

In every system, except badly designed ones, a higher average power expenditure will result in improved delay performance. Conceptually, all possible combinations of average power and average queueing delay achievable under different selections of a protocols' parameters yield a delay vs. power tradeoff curve as in Fig.1. The particular point on the tradeoff curve which the system will operate on is chosen based on which type of cost (power or delay) the network operator prefers to minimize at the expense of the other. In this paper we seek power control algorithms that yield the best such tradeoff curves, i.e., we are interested in algorithms offering the minimum average queueing delay at any given level of average power expenditure. (Equivalently, any such algorithm minimizes the average power expenditure for any given level of average queueing delay.)

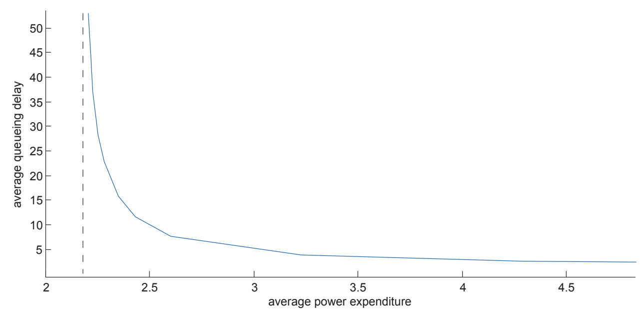

Fig. 1. The tradeoff curve between average power expenditure and the corresponding average queueing delay. The dashed line signifies the minimum possible power such that the input load is serviceable.

It is only recently that the problem of the joint minimization of delay and energy efficiency started to be addressed [1], [5], [3], [4], [6]. Most of these works start from a *static* optimization problem involving rates and/or power as the decision variables and then devise an iterative algorithm which converges at an optimal solution. The queueing delays are not taken into account explictly but it is through an identification of the variables of the dual problem with the queue backlogs that queueing delay is implicitly optimized as well. Moreover, due to the static nature of the associated optimization problems, the queues' evolution (i.e., system dynamics) is not taken into account in the optimization.

In this paper we account for the queues' evolution and both types of cost by using an optimal control problem

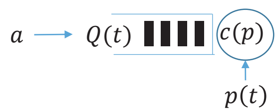

Fig. 2. The single queue problem: at any time t choose power level $p(t)$ such that the time-average cost (1) is minimized.

formulation. Although structural results of the optimal policies for standard queueing networks have long been known (see e.g.,[11], [10], [12]), to the best of our knowledge no similar results exist for interfering queues. A key element in our approach, motivated from work on the optimal control of queues at heavy traffic (e.g., see [14], [15], the book of Meyn [7] and references therein), is the consideration of *workload* instead of queueing delay as the queueing cost. At any time instant the workload is defined[1] as the minimum time required to serve all packets in the network queues if new packet arrivals were to be disregarded. Naturally, queueing delay and workload are related quantities so by optimizing the latter one implicitly optimizes the former as well. The benefit of using workload as the queueing cost is that then the problem of joint minimization of queueing cost *and* power can be addressed explicitly by treating the network as a one-dimensional single queue system where the structure of optimal controls is known. The optimal solution is expressed as a two-step decision. In the first "power control" step the total power to be consumed across all the transmitters at that instant is decided. In the second "scheduling" step, the total power is distributed to the transmitters so that the resulting speed at every transmitter is proportional to the backlog of it's queue.

The paper is organized as follows. In section III we introduce the basic notation and system model. In section IV we describe the MinTime scheduling algorithm and obtain a useful characterization through the auxiliary problem of emptying the network as fast as possible if no new arrivals were to be considered. Section V contains the basic properties of our algorithm obtained by considering the fluid model of our system. The power control step is studied in section VI. Certain issues are discussed in section VII.

II. The single-queue problem

Before moving to the general case let us consider the case of controlling the power of a single transmitter, when no interferers are present. Packets arrive at the transmit queue at an average rate a as depicted in Fig. 2, which at time t has length $Q(t)$. One wants to apply power $p(t)$ such that the average cost

$$\lim_T \frac{1}{T} \int_0^T \left(g(Q(t)) + p(t)\right) dt \tag{1}$$

is minimized, where $g(\cdot)$ is an increasing nonnegative function. Crabill first studied this problem in [11] and (not surprisingly) obtained that the optimal power $p(t)$ is given by a nondecreasing function of $Q(t)$, i.e., $p(t) = f(Q(t))$ for nondecreasing $f(\cdot)$. For most cases of interest, $f(\cdot)$ is not known in closed form but it can be obtained numerically by the results in [12]. This monotonicity is generalized to open and closed networks of queues in [10]. As explained in the introduction, in Section V we show that under a specific algorithm a network with many interfering queues essentially becomes a one-dimensional system for which we will take advatange of the monotonicity above to propose approximately optimal power controls.

III. System Model

Consider a wireless network comprised of n transmitters indexed by i, each associated to a separate receiver i. If p_i is the instantaneous transmit power of transmitter i and $p = (p_i, i = 1, \ldots, n)$ denotes the vector of transmit powers then the resulting rate for the i-th transmitter is

$$c_i(p) = b_i \log\left(1 + \frac{G_{ii}p_i}{\sum_{j\neq i} G_{ji}p_j + n_i}\right),$$

where $b_i > 0$ a constant, G_{ij} is the channel gain between the i-th transmitter and the j-th receiver, while n_i is the power of the background noise at the i-th receiver.

For any power level P, the *rate region* $R(P)$ is the set of all n-dimensional rate vectors $(c_i(p), i = 1, \ldots, n)$ attained as p ranges over all values $p = (p_i, i = 1, \ldots)$ which satisfy $\sum_i p_i \leq P$. Let $\partial_{\text{opt}} R(P)$ be the set of Pareto optimal, i.e., maximal, points of $R(P)$.

Lemma 1. *$\partial_{opt} R(P) = \{c(p) | p = (p_i, i = 1, \ldots, n)$ with $\sum_i p_i = P\}$.*

Proof: No rate vector $c(p) \in R(P)$ can be Pareto optimal if $\sum_i p_i < P$ since in that case $c((1+\epsilon)p) > c(p)$ and $c((1+\epsilon)p) \in R(P)$ for a sufficiently small $\epsilon > 0$.

Conversely, by Zander [9] there is no vector $c(p')$ with $c(p') \geq c(p)$ and $\sum_i p'_i = \sum_i p_i = P$ unless $c(p') = c(p)$. ■

Thus the rate region achievable for any power level P can be parameterized by the power vectors with total power P.

For any vector of queue lengths $q \in \mathbb{R}^n_+$, the *workload* $t_P(q)$ is given as the optimal value of the following problem:

$$t_P(q) = \min \quad t \tag{2}$$

$$\text{such that } \frac{q}{t} \leq \tilde{c}\,, \tag{3}$$

$$\text{over } \tilde{c} \in \text{Co}(R(P))$$

At each transmitter i, packets with unit average size are produced for transmission towards receiver i according to a Poisson process $A_i(t)$ of intensity a_i. Also, let $N_i, i = 1, \ldots n$

[1] A formal definition is given in (2).

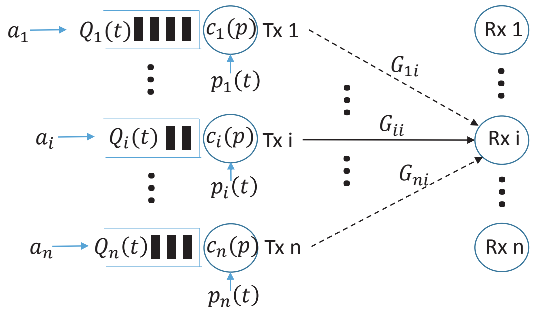

Fig. 3. The system model of n single-hop links.

be another set of Poisson processes of unit intensity and assume that $A_i, N_i, i = 1, ..., n$ are mutually independent.

Let $p(t) = (p_i(t), i = 1, \ldots, n)$ be any nonnegative set of powers actuated at the transmitters at time t. The number of backlogged packets $Q_i(t)$ at the i-th transmitter at time t, is given by the set of stochastic equations

$$Q_i(t) = Q_i(0) + A_i(t) - N_i\left(\int_0^t c_i(p(s))ds\right), i = 1, \ldots, n. \tag{4}$$

We say that the process of powers $p(t)$ is *feasible* if it is causal, i.e., it is adapted to the filtration generated by the queue sizes $Q(t) = (Q_i(t), i = 1, \ldots, n)$, and $p_i(t) = 0$ whenever $Q_i(t) = 0$. The latter ensures $Q(t) \geq 0$ for all t.

For every feasible process $p(t)$ we consider the resulting *average cost* given by

$$\limsup_T \frac{1}{T}\int_0^T \left(\epsilon t_{\sum_i p_i(t)}(Q(t)) + \sum_i p_i(t)\right) dt. \tag{5}$$

The cost rate is the sum of the total power consumption $P(t) = \sum_i p_i(t)$ at time t with the workload cost $\epsilon t_{P(t)}(Q(t))$ The constant $\epsilon > 0$ controls the tradeoff between low queueing cost and low average power consumption: lower values of ϵ give higher priority to lowering power consumption.

IV. The MinTime Scheduling Algorithm

In this section we introduce an algorithm for controlling the vector of powers $p(t)$ at every time t. For any nondecreasing nonnegative function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $f(0) = 0$, and a constant vector $\lambda \in \mathbb{R}_+^n$ consider the following algorithm which depends on the current queue lengths $Q(t)$:

Power control: The total power expenditure at time t is $P(t) = f(t_P(Q(t)))$, for some constant $P > 0$.

MinTime scheduling: Choose average service rate $\tilde{c} \in \mathrm{Co}(\partial_{\text{opt}} R(P))$ with $\tilde{c} \parallel Q(t)$ [2].

The power control step decides the instantaneous total power expenditure across transmitters, and is given as an increasing function of the workload $t_P(Q(t))$ with respect to some power level P (with $P \neq P(t)$ in general). In section VI we will consider specific choices for P and $f(\cdot)$.

In the scheduling step a service vector $\tilde{c}$ is chosen from $\mathrm{Co}(\partial_{\text{opt}}(R(P))$. By Lemma 1, $\mathrm{Co}(\partial_{\text{opt}}(R(P))$ is an n-dimensional convex set. Thus there is a unique choice of $\tilde{c} \in \mathrm{Co}(\partial_{\text{opt}} R(P))$ giving $\tilde{c} \parallel Q(t)$. Observe that if $\tilde{c}$ is not an extreme point of $\partial_{\text{opt}} R(P)$ then $\tilde{c}$ is not an attainable service rate since $\tilde{c} \notin R(P)$. Nonetheless through appropriate timesharing between extreme points of $\partial_{\text{opt}} R(P)$, the service rate $\tilde{c}$ is effected over a longer timescale. Whenever such timesharing is necessary, we will assume that the instantaneous service is a random choice of one of the extreme points of $\partial_{\text{opt}} R(P)$ with appropriate selection probabilities such that $\tilde{c}$ is effected on the average. In this case the average power consumption will be P by Lemma 1.

The choice of $\tilde{c}$ in the scheduling step can be interpreted as the service vector by which the network queues will empty in the *minimum time* if no further arrivals would be admitted and the total power expenditure is limited to P, i.e., it is the optimal solution of (2).

The dual problem of (2) is

$$\max_{\lambda \geq 0}\left[2\sqrt{\lambda \cdot q} - \max_{\tilde{c} \in \mathrm{Co}(\partial_{\text{opt}} R(P))} \lambda \cdot \tilde{c}\right]. \tag{6}$$

The first order conditions for optimality give $q = \tilde{c}\sqrt{\lambda \cdot q}$, and so we see that it agrees with the scheduling step of the algorithm. For any queue vector q and power level P let $\lambda^{q,P}$ denote the optimal dual variables. The service vector $\tilde{c}$ can be also interpreted as the optimal solution of a variant of the MaxWeight [2] algorithm

$$\max_{\tilde{c} \in \mathrm{Co}(\partial_{\text{opt}} R(P))} \lambda^{Q(t),P(t)} \cdot \tilde{c}. \tag{7}$$

Observe that since MinTime scheduling yields vectors $\tilde{c}$ parallel to q, in essence it solves the optimization problem

$$\max_{\tilde{c} \in \mathrm{Co}(\partial_{\text{opt}} R(P))} \frac{q \cdot \tilde{c}}{||\tilde{c}||},$$

as compared to MaxWeight scheduling which solves

$$\max_{\tilde{c} \in \mathrm{Co}(\partial_{\text{opt}} R(P))} q \cdot \tilde{c},$$

(see also [2], [3]).

V. Fluid Model Analysis

In this section we study the equilibrium properties of the MinTime scheduling algorithm introduced in the last section and establish a kind of optimality properties. We assume that the overall power expenditure $P(t)$ is fixed at a level P and focus only on the effect due to scheduling.

It turns out that the behaviour of a stochastic system in the long-term can be described by the so-called "fluid model" deterministic equations, which arise by disregarding all randomness in system dynamics. Thus, we treat arrivals as the continuous arrival of "fluid" at the same rate a as before, and look at the average effect $\tilde{c}$ of randomization by extreme points

[2]For two vectors x, y we let $x \parallel y$ denote that the two vectors are parallel, i.e., $x = \alpha y$ for some scalar α. Also Co(A) denotes the convex hull of A.

$c(p(t))$ of $R(P(t))$. For more on fluid models and their relation to the stochastic system, see [7]. The fluid model for (4) is

$$\dot{q}(t) = a - \tilde{c}\,, t \geq 0\,, \tag{8}$$

where $\tilde{c} \in \mathrm{Co}(\partial_{\mathrm{opt}} R(P))$ and is such that any solution satisfies $q(t) \geq 0$.

A basic question is whether for this power expenditure the network queues can ever empty. The answer is affirmative provided the condition $t_P(a) < 1$ holds, where t_P is defined in (2). Intuitively, this is true since the queues would not be stable if a unit time's worth of arrivals cannot be cleared at the same time - see [7]. We note here that if in the stochastic system (4) the power expenditure is kept fixed at P, then (stochastic) stability is characterized by the same condition [8]. We say P_* *is critical for the arrival rate vector* a if $t_{P_*}(a) = 1$. In this case the fluid model will never empty its' queues under any algorithm with $\sum_i p_i(t) \geq P$.

Now consider the case of MinTime scheduling with critical power P_*. The algorithm will give power vectors $p(t)$ which satisfy $\tilde{c}(p(t)) \parallel q(t)$. Since P_* is critical the queues will never empty but their sizes will vary and reach any of a set of possible nonzero equilibria:

Proposition 1. *Under the MinTime scheduling at critical power level P_* for a, for any initial queue vector $q(0)$ the queue sizes converge toward the set $I = \{q \geq 0 | q \parallel a\}$ of equilibria.*

Proof: The function $V(q) = ||q||^2||a||^2 - (q \cdot a)^2 \geq 0$ is a Lyapunov function for the system with respect to I. This is because

$$\begin{aligned}\dot{V}(q) &= -2\left(||a||^2 q \cdot \tilde{c} - (q \cdot a)(a \cdot \tilde{c})\right) \\ &\leq -2||a||^2\left(q \cdot \tilde{c} - ||q|| \cdot ||\tilde{c}||\right) = 0\end{aligned}$$

where the equality holds because $q \parallel \tilde{c}$ for MinTime scheduling. The inequality is strict except if $q \parallel a$ in which case $\tilde{c} \parallel a$ (by $q \parallel \tilde{c}$) and so $q \cdot a = ||q|| \cdot ||a||, a \cdot \tilde{c} = ||a|| \cdot ||\tilde{c}||$. ■

Now we provide a lower bound to any power control algorithm operating under a constant P, with respect to a linear combination of the queue sizes. Define the *localized workload* at the queue vector q as $w_P(q) = \lambda^{a,P} \cdot q$, where $\lambda^{a,P}$ is the optimal dual variables in (6) for parameters $q = a$ and P.

For any choice of service vector $\tilde{c}(t) \in \mathrm{Co}(\partial_{\mathrm{opt}} R(P)), t \geq 0$ we have

$$\begin{aligned}\dot{w}_P(q(t)) &= \lambda^{a,P} \cdot a - \lambda^{a,P} \cdot \tilde{c}(t) \\ &\geq \lambda^{a,P} \cdot a - \max_{\tilde{c} \in \mathrm{Co}(\partial_{\mathrm{opt}} R(P))} \lambda^{a,P} \cdot \tilde{c}\,,\end{aligned} \tag{9}$$

where the last expression is algorithm independent, and by (7) the $\tilde{c}$ where the maximum is attained satisfies $\tilde{c} \parallel a$. Thus if we consider the one dimensional system with state $w_P(t)$ evolving as

$$\begin{aligned}&\dot{w}_P(t) = w_P(0) + \lambda^{a,P} \cdot a - \max_{\tilde{c} \in \mathrm{Co}(\partial_{\mathrm{opt}} R(P))} \lambda^{a,P} \cdot \tilde{c}\,, \\ &\text{if } w_P(t) > 0\,,\end{aligned} \tag{10}$$

then $w_P(q(0)) \geq w_P(0)$ implies $w_P(q(t)) \geq w_P(t)$ for all $t \geq 0$ and any power selection algorithm with power expenditure P.

For the case of the MinTime algorithm if $q(0) \in I$ then $\tilde{c}(0) \parallel a$ as does the optimal $\tilde{c}$ in (9). Moreover this will be true for all times after t since by Proposition 1 I is an invariant set for $q(t)$. Thus $w_P(q(t)) = w_P(t)$ holds with equality if $q(0) \in I$ and this leads us to the following property which states that the MinTime algorithm attains lower values for $w_P(q(t))$ than any algorithm when both are started at states of equal value of workload.

Proposition 2. *Let $q_{MT}(t), q(t)$ be the respective evolutions of queues of systems under MinTime scheduling and any power selection algorithm, both having equal power expenditure P. Then $w_P(q_{MT}(0)) = w_P(q(0))$ implies $w_P(q_{MT}(t)) \leq w_P(q(t))$ for all $t \geq 0$.*

What happens if we compare algorithms for the same initial states, i.e., $q_{\mathrm{MT}}(0) = q(0)$? Clearly $w_P(q(t)) < w_P(q_{\mathrm{MT}}(t))$ is then possible for all t. Nevertheless, the lower bound $w_P(t)$ always satisfies $w_P(t) \leq w_P(q_{\mathrm{MT}}(t))$, if $w_P(0) = w_P(q_{\mathrm{MT}}(t))$, and hence $w_P(t) \leq w_P(q_{\mathrm{MT}}(t))$ holds for all t. In what follows we argue that actually this holds with (approximate) equality, i.e., $w_P(t) \approx w_P(q_{\mathrm{MT}}(t))$ will hold for all t *after some long initial period*, if $P = P_*$. This is due to the randomness in the arrival and service times which the fluid model analysis ignored: the fluid model provides an accurate description of the system dynamics over not very long time periods.

To see what happens over longer timescales we argue informally as follows. If the initial queues $Q_{\mathrm{MT}}(0)$ at the actual (i.e., nonfluid) system have sizes proportional to some large scalar parameter n indicating the size of the system, then over the time interval $[0, nt]$ the queue sizes will change by amounts proportional to n since the number of arrivals and departures over that period are themselves proportional to n. Thus, $(Q_{\mathrm{MT}}(nt) - Q_{\mathrm{MT}}(0))/n = t(a - \tilde{c}(p)) + O(\sqrt{n})/n$ provided p does not change significantly during $[0, nt]$. The $O(\sqrt{n})$ term is due to the random deviations about the mean values in the arrival and service times. Thus at or near the equilibrium set of states I, $Q_{\mathrm{MT}}(nt) - Q_{\mathrm{MT}}(0) = O(\sqrt{n})$ holds. Over a longer period $[0, n^2 t]$ we will have $(Q_{\mathrm{MT}}(n^2 t) - Q_{\mathrm{MT}}(0))/n = O(1)$, i.e., the accumulated random deviations will have a $O(1)$ effect after time $O(n^2)$, and thus should not be ignored if one is interested in equilibrium properties. Notice that $Q_{\mathrm{MT}}(n^2 t)$ will never be more that $O(n)$ away of I, since because of the nonzero drift towards I, corrections of the size $O(n^2)$ over the time period $[0, n^2]$ will be into effect.

Now $w_{P_*}(Q_{\mathrm{MT}}(n^2 t)/n$ will also change by an $O(1)$ amount because $w(\cdot)$ is a linear function. The accumulated noise that drive the excursions around means is zero mean[3], so the system effectively performs a random walk over the set I, viz. the direction a. At some finite time this random walk will hit 0, and so it will couple with the lower bound $w(t)$

[3]It is a martingale.

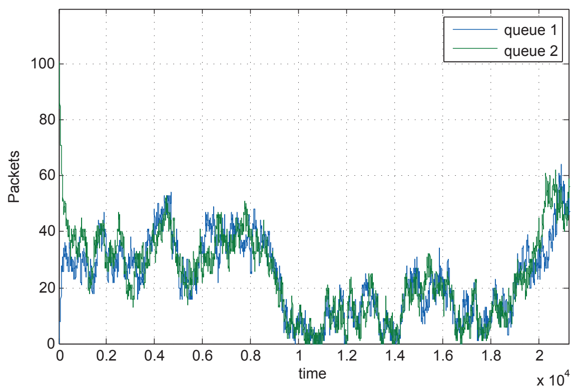

Fig. 4. The queue evolutions in a network with two symmetric links under the MinTime scheduling algorithm. The queue lengths equalize almost immediately.

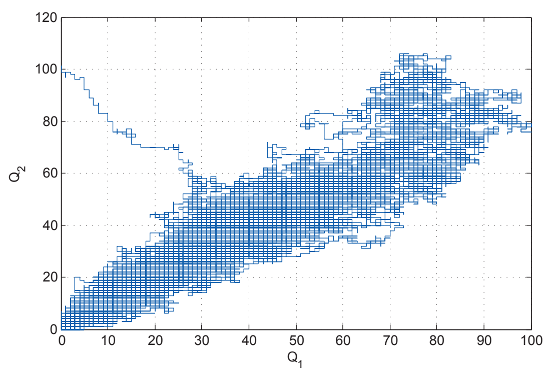

Fig. 5. The phase-space diagram of the queue evolution in Fig. 4

process or any other localized workload process started at lower workload levels. Since the fluid analysis guarantees $w_{P_*}(t) \leq w_{P_*}(Q(t))/n$ (modulo $o(n)$ terms) for the actual queue sizes under any algorithm, we have $w_{P_*}(Q_{\rm MT}(n^2t)) \leq w_{P_*}(Q(n^2t))$ (again modulo $o(n)$ terms) for all sufficiently long t.

In Figs. 4- 5 we simulate the trajectories of the queue sizes in a network of two symmetric links where $Q_1(0) = 0, Q_2(0) = 100$ for a specific power control (see the "sqrt" rule at the end of the next section). The arrival rates at the two links are the same, so one expects that the queue sizes will equalize. The initial phase prior to the system hitting the invariant set I is short.

VI. Optimizing the Total Power Expenditure

In the previous section we argued that the MinTime algorithm attains the least possible value of workload (modulo some small deviations due to randomness) in a pathwise sense, compared to any other algorithm using the same fixed power expenditure P. Here we allow $P(t)$ to vary in order to minimize the average cost criterion (5).

This power can never fall below the critical level P_* required for queue stability. On the other hand, as ϵ in (5) tends to zero, the optimal behavior essentially drives average power to the lowest possible level since it becomes increasingly more expensive than the queueing cost. Thus for small ϵ we expect the total power selection by the optimal algorithm to give values $P(t) \approx P_*$, and so $w_{P_*}(q_{\rm MT}(t))$ is an approximate lower bound for the localized workload of any other algorithm started at the same state, as explained in the last section.

After an initial period, when the queues under MinTime scheduling converge to I, the value of $w_{P_*}(q_{\rm MT}(t))$ will change according to $\dot{w}_{P_*}(q_{\rm MT}(t)) = \lambda^{a,P_*} \cdot \alpha - C(P(t))$, where

$$C(P) = \max_{\tilde{c} \in {\rm Co}(\partial_{\rm opt} R(P))} \lambda^{a,P_*} \cdot \tilde{c}\,. \tag{11}$$

Thus the workload $w_{P_*}(t)$ evolves similarly to the contents of a single controlled queue with unit rate arrivals and a service rate $C(P)$ controlled through the total average power expenditure P, i.e., $\dot{w}_{P_*}(t) = 1 - C(P)$, where P is the control input. The optimal policy of choosing $P(t)$ with respect to the average cost criterion

$$\limsup_{T} \frac{1}{T} \int_0^T (\epsilon w_{P_*}(t) + P(t))\, dt\,,$$

would produce the minimum average cost also for (5). if $w_{P_*}(t) \approx t_{P(t)}(Q_{\rm MT}(t))$ held exactly. As explained in Section II, we know that the optimal policy for the average cost criterion (5) is such that the instantaneous service rate is a nondecreasing function of the current queue size. Moreover, the service rate is always at least the arrival rate. Motivated by this, we heuristically consider policies which select the total power level $P(t)$ based solely on the current workload $t_{P(t)}(Q(t)) \approx t_{P_*}(Q(t))$, i.e., $P(t) = f(t_{P_*}(Q(t)))$ for increasing functions f with $f(x) \geq P_*$ if $x \neq 0$[4].

In Fig. 6 we obtain the tradeoff curves by simulating a network of two symmetric links with $G_{11} = G_{22} = 0.8, G_{12} = G_{21} = 0.3$. For power control we use $f(x) = P_* + \epsilon\sqrt{x}$ ("sqrt") and $f(x) = P_* + \epsilon \log(x)$ ("log"), and compare with two algorithms based on the following problems of power minimization: The "Primal" algorithm solves

$$\min \sum_i p_i + \epsilon D(p)$$

iteratively for the penalty function $D(P) = \sum_i (c_i(p) - a_i)^{-1}$, motivated from the average queueing delay in an M/M/1 queue. The "Primal" algorithm is just gradient projection, and the different points of the tradeoff curve are obtained by varying ϵ. The "Primal-dual" algorithm is again gradient projection on the Lagrangian over both primal and dual variables of the following problem:

$$\min_{p \geq 0} \epsilon \sum_i p_i \text{ such that } c(p) \geq a.$$

The tradeoff curve is obtained by varying ϵ.

Notice that both sqrt and log with MinTime yield uniformly better operating points than the others. The primal-dual algorithm has the worse performance and this may be due to the

[4] $f(0) = 0$ of course.

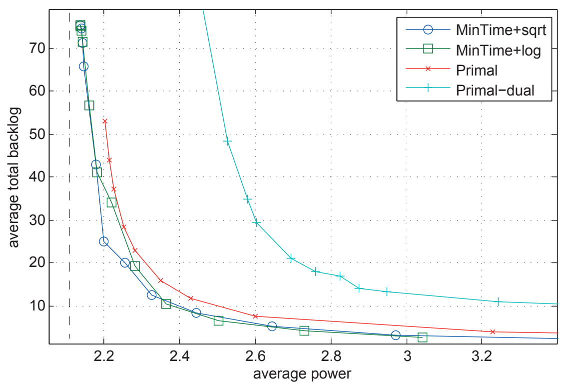

Fig. 6. The tradeoff curves between average power expenditure and the corresponding average total backlog for MinTime scheduling under two different power control schemes. "Primal" and "Primal-dual" are based on two different static formulations of power minimization.

fact that it short queues are very noisy signals -and hence not very informative- of the value of dual variables. That is, the correspondence between dual variables and queues shown to exist in deterministic models, it breaks down in the presence of large noise. The results in Fig. 6 are generally robust in the choice of f. Thus, simple stationary rules may be adequate for balancing delay and power.

VII. Discussion

The problem of solving the optimization problem (7) involved in MinTime scheduling is not touched in this paper. In [16] the authors show that (2) is NP complete and it is even hard to approximate in polynomial time. Thus we cannot hope for a computationally efficient algorithm for solving (2) in every case. Nevertheless, there are some initial thoughts of how to do this in specific cases such as in the "low SNR" regime, i.e., when no timesharing is necessary. For example, observe that the Foschini-Miljanic distributed power control algorithm [13] computes the minimum power expenditure under which the constraint (3) is satisfiable for any pair of queue length vector w and value for t. If this expenditure matches P, then this means that t is indeed the value of workload at q; if the expenditure exceeds P then the network can empty within time t only at a higher than P power consumption. Thus, t in fact underestimates the true workload $t_P(q)$ and so one could run the Foschini-Miljanic algorithm again for a higher t value this time, and continue until the power expenditure to empty by t converges to P. We plan to pursue this idea further and provide an extension of the Foschini-Miljanic algorithm which also computes the workload for any given vector of queue sizes.

We believe that the conceptual approach in this paper is valuable, in that it shows how the power control part of the problem can be decoupled into a scheduling and total power expenditure selection step and solved easily because it collapses to a single dimension where structural results for optimal policies are available.

Acknowledgment

This work was funded by the research program "CROWN", through the Operational Program "Education and Lifelong Learning 2007-2013" of NSRF, which has been co-financed by EU and Greek national funds.

References

[1] M. Chiang, "Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control", IEEE J. Sel. Areas Comm., vol. 23, no. 1, pp. 104-116, 2005.

[2] L. Tassiulas and A. Ephremides, "Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks", IEEE Trans. on Automatic Control, vol. 37, no. 12, pp. 1936- 1949, 1992.

[3] A. Giannoulis, K. P. Tsoukatos and L. Tassiulas, "Cross-layer Power Control in Wireless Networks", SIGMETRICS Performance Evaluation Review, vol. 34, no. 2, pp. 29-31, 2006.

[4] A. Giannoulis, K. P. Tsoukatos and L. Tassiulas, "Maximum Throughput Power Control in CDMA Wireless Networks", ICC 2006, pp. 4457-4462, 2006.

[5] A. Stolyar, "Greedy Primal-Dual Algorithm for Dynamic Resource Allocation in Complex Networks", Queueing Systems, vol. 54, pp. 203-220, 2006.

[6] L. Georgiadis, M. J. Neely, and L. Tassiulas, "Resource Allocation and Cross-Layer Control in Wireless Networks", Foundations and Trends in Networking, vol. 1, no. 1, pp. 1-144, 2006.

[7] S. Meyn. *Control techniques for complex networks*. Cambridge University, 2007.

[8] J. Dai, "On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models," Ann. Appl. Probab., vol. 5, pp. 49V77, 2005.

[9] J. Zander, "Performance of Optimum Transmitter Power Control in Cellular Radio Systems", IEEE Transactions on Vehicular Technology, VT-41, no. 1, pp. 57-62, February 1992.

[10] Weber, R.R., and Stidham, S., "Optimal Control of Service Rates in Networks of Queues," Advances in Applied Probability, vol. 19, 202-218, 1987.

[11] Crabill, T. B., "Optimal Control of a Service Facility with Variable Exponential Service Times and Constant Arrival Rate," Management Science, pp. 560-566, 1972.

[12] J. George and M. Harrison,"Dynamic Control of a Queue with Adjustable Service Rate", Operations Research, vol. 49, no. 5, pp. 720-731, 2001.

[13] G. J. Foschini and Z. Miljanic, "A Simple Distributed Autonomous Power Control Algorithm and its Convergence", IEEE Transactions on Vehicular Technology, VT-42, no. 4, pp. 641-646, April 1993.

[14] Kelly, F. P. and Laws, C. N. "Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling," Queueing Systems Theory Appl., vol. 13, pp. 47V86, 1993.

[15] M. Harrison, "Correction: Brownian Models of Open Processing Networks: Canonical Representation of Workload", The Annals of Applied Probability, vol. 16, no. 3, pp. 1703-1732, 2006.

[16] D. Shah, D. N. C. Tse, and J. N. Tsitsiklis, "Hardness of Low Delay Network Scheduling", IEEE Transactions on Information Theory, vol. 57, no. 12, pp. 7810-7817, 2011.