University of Pennsylvania
ScholarlyCommons

Departmental Papers (ESE) | Department of Electrical & Systems Engineering

4-19-2009

# Layers and layer interfaces in wireless networks

Alejandro Ribeiro
*University of Pennsylvania*, aribeiro@seas.upenn.edu

Copyright 2009 IEEE. Reprinted from:
Ribeiro, A., "Layers and layer interfaces in wireless networks," Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on , vol., no., pp.2557-2560, 19-24 April 2009
URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4960144&isnumber=4959496

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/ese_papers/512
For more information, please contact repository@pobox.upenn.edu.

# Layers and layer interfaces in wireless networks

**Abstract**
This paper proposes an optimal architecture for wireless networks based on layers and layer interfaces. In the presence of fading the architecture is shown to be optimal. The result follows from a subgradient descent algorithm on the dual function of a generic wireless networking optimization problem. The fact that these non-convex optimization problems have nonetheless zero duality gap is exploited.

**Keywords**
concave programming, duality (mathematics), gradient methods, network interfaces, radio networks, layers-and-layer interface, nonconvex optimization problem, subgradient descent algorithm, wireless network optimization problem, zero duality gap, Cross-layer design, Optimization, Wireless networks

**Comments**
Copyright 2009 IEEE. Reprinted from:

Ribeiro, A., "Layers and layer interfaces in wireless networks," Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on , vol., no., pp.2557-2560, 19-24 April 2009

URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4960144&isnumber=4959496

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Pennsylvania's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by writing to pubs-permissions@ieee.org. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

This conference paper is available at ScholarlyCommons: http://repository.upenn.edu/ese_papers/512

# LAYERS AND LAYER INTERFACES IN WIRELESS NETWORKS

*Alejandro Ribeiro*

Dept. of Electrical and Systems Engineering – University of Pennsylvania

**ABSTRACT**

This paper proposes an optimal architecture for wireless networks based on layers and layer interfaces. In the presence of fading the architecture is shown to be optimal. The result follows from a subgradient descent algorithm on the dual function of a generic wireless networking optimization problem. The fact that these non-convex optimization problems have nonetheless zero duality gap is exploited.

**Keywords:** Wireless networks. Optimization. Cross-layer design.

## 1. INTRODUCTION

Optimal design is one of the most promising alternatives for future wireless networks [1, 2]. Desired operating points are defined as solutions of optimization problems. Fundamental properties of wireless networks and protocols are obtained from characteristics of these optimization problems and algorithms that solve them. An interesting observation is that even if the might be suboptimal, layered architectures follow from the decomposition of Lagrangian dual problems [3]. Contrary to established wisdom, it has been shown recently that in the presence of fading layered architectures are in fact optimal [5].

The goal of this paper is to introduce an optimal architecture based on layers and layer interfaces. Layers maintain variables of interest to the network, while interfaces maintain auxiliary variables. Layers exchange variables only with adjacent interfaces and interfaces only with adjacent layers. Over time the network finds an optimal operating point that maximizes a given utility. Although the architecture presented here is novel, similar architectures have been reported elsewhere, see e.g., [3]. The main contribution of this paper is to show its optimality in the presence of fading.

## 2. OPTIMAL WIRELESS NETWORK

Consider an ad-hoc wireless network composed of $J$ user terminals $\{T_i\}_{i=1}^J$. Terminal $T_i$ wants to deliver packets for different application level flows generically denoted by $k$, with the flow $k$ intended for destination $T^k$. Network connectivity is modeled with a graph $\boldsymbol{\mathcal{G}}(v, e)$ with vertices $v := [1, J]$ and edges $e \in \mathcal{E}$ connecting pairs of vertices $(i, j)$ when and only when $T_i$ and $T_j$ can communicate with each other; see Fig. 1. The adjacency of $i$ is denoted $n(i) := \{j : (i, j) \in \mathcal{E}\}$. Each terminal $\{T_j\}_{j\in n(i)}$ that can communicate with $T_i$ will be referred to as a neighbor and the set of all neighbors as $T_i$'s neighborhood. Network nodes communicate using a set of frequency tones $f \in \boldsymbol{\mathcal{F}}$. The channel from $T_i$ to $T_j$ is denoted as $h_{ij}^f$ and modeled as a random variable. Channel gains of all network links are collected in the vector $\boldsymbol{h}$.

Terminals $T_i$ select various variables that determine the flow of information through the network. For given channel realizations $\boldsymbol{h}$, terminal $T_i$ determines a power profile $p_{ij}^f(\boldsymbol{h})$ used for sending packets to $T_j$ on the tone $f$ when the channel vector realization is $\boldsymbol{h}$. Power profiles determine $T_i$'s power consumption $p_i$ and the capacity $c_{ij}$ of the $T_i \to T_j$ link. For every flow $k$, $T_i$ sends packets to neighboring terminals $\{T_j\}_{j\in n(i)}$ at an average rate $r_{ij}^k$. Likewise it receives packets from neighbors at a rate $r_{ji}^k$. Finally, variables $a_i^k$ determine the rate at which $T_i$ accepts packets of the flow $k$ from applications. These variables are not independent of each other. They must satisfy constraints that will be explained shortly [cf. (2)-(4)].

Network design calls for selecting powers $p_i$, link capacities $c_{ij}$, rates $r_{ij}^k$, arrivals $a_i^k$ and power profiles $p_{ij}^f(\boldsymbol{h})$ that are optimal in some sense. Thus, concave $U_i^k(a_i^k)$ and convex $V_i(p_i)$ functions, respectively, representing the value of rate $a_i^k$ and the cost of power $p_i$ are introduced. The optimal operating point is then defined as the set of variables solving the optimization problem

$$P = \max \sum_{i,k} U_i^k(a_i^k) - \sum_i V_i(p_i) \tag{1}$$

$$c_{ij} \le \mathrm{E}_{\boldsymbol{h}}\left[\sum_{f\in\boldsymbol{\mathcal{F}}} C_{ij}\left(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})\right)\right] \tag{2}$$

$$p_i \ge \mathrm{E}_{\boldsymbol{h}}\left[\sum_{j\in n(i)}\sum_{f\in\boldsymbol{\mathcal{F}}} p_{ij}^f(\boldsymbol{h})\right] \tag{3}$$

$$a_i^k \le \sum_{j\in n(i)}\left(r_{ij}^k - r_{ji}^k\right), \qquad \sum_k r_{ij}^k \le c_{ij}. \tag{4}$$

Of the two constraints in (4) the first one requires the rate $a_i^k$ at which packets are accepted from applications to be smaller than the difference between the aggregate departure rates (to neighbors) $\sum_{j\in n(i)} r_{ij}^k$ and arrival rates (from neighbors) $\sum_{j\in n(i)} r_{ji}^k$. The second constraint requires the total rate $\sum_k r_{ij}^k$ sent from $T_i$ to $T_j$ for all flows to be smaller than the link's capacity $c_{ij}$. The constraint in (3) states that the average power consumption $p_i$ is obtained by summing over all links $j \in n(i)$ and tones $f \in \boldsymbol{\mathcal{F}}$ and taking expected value over channel realizations $\boldsymbol{h}$. The capacity constraint in (2) is a similar average over fading states and tones. The function $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$ maps channels and powers into link capacities so that the capacity $c_{ij}^f(\boldsymbol{h})$ of the link $T_i \to T_j$ on the tone $f$ is $C_{ij}(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h}))$. The function $C(\cdot)$ is determined by terminal's capabilities and operating conditions. If, e.g., terminals perform single user detection, link capacity is determined by the signal to noise plus interference ratio (SINR). Please refer to [5] for a more detailed account of the model in (1)-(4).

All problem variables have to be non-negative, but this is left implicit in (1)-(4). Also implicit in (1)-(4) are power constraints $p_i \le p_{\max i}$ and $p_{ij}^f(\boldsymbol{h}) \le p_{\max}$, arrival rate requirements $a_{\min i}^k \le a_i^k \le a_{\max i}^k$ and upper bound constraints $c_{ij} \le c_{\max}$ and $r_{ij}^k \le r_{\max}$ on link capacities and link flow rates. These constraint define a box $B$ of feasible variables. They will be implicit in general and make explicit when demanded by clarity. For future reference define the vector valued power distribution $\boldsymbol{p}(\boldsymbol{h})$ with components $p_{ij}^f(\boldsymbol{h})$ and $\boldsymbol{X}$ the set of primal variables $c_{ij}$, $p_i$, $r_{ij}^k$ and $a_i^k$ for all possible subindexes – i.e., all $i$ and all $j \in n(i)$ for $c_{ij}$, all $i$ for $p_i$ and so on. Further define $f(\boldsymbol{X})$ as the utility function in (1) and $\boldsymbol{h}\left[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})\right] \ge \boldsymbol{0}$ the constraints (2)-(4) so that (1)-(4) can be written in generic form as

$$P = \max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} f(\boldsymbol{X}); \qquad \text{st } \boldsymbol{h}\left[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})\right] \ge \boldsymbol{0} \tag{5}$$

where $B := \left\{(\boldsymbol{X}, \boldsymbol{p}(h)) : 0 \le p_{ij}^f(\boldsymbol{h}) \le p_{\max}, 0 \le p_i \le p_{\max i}, a_{\min i}^k \le a_i^k \le a_{\max i}^k, 0 \le c_{ij} \le c_{\max},\ 0 \le r_{ij}^k \le r_{\max}\right\}$ is the box outlined above.

The function $C(\cdot)$ in (2), and as a consequence $\boldsymbol{h}\left[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})\right]$ in (5), is not concave in general. Therefore, (1) is a difficult optimization problem. This difficulty notwithstanding, properties of wireless networks can be derived from properties of (1). For this purpose introduce multipliers $\boldsymbol{\Lambda}$ and the Lagrangian

$$\boldsymbol{\mathcal{L}}[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h}), \boldsymbol{\Lambda}] = f(\boldsymbol{X}) + \boldsymbol{\Lambda}^T \boldsymbol{h}\left[\boldsymbol{X}, \boldsymbol{p}(\boldsymbol{h})\right]. \tag{6}$$

The dual function is obtained by maximizing the Lagrangian over the



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 13, 2009 at 10:48 from IEEE Xplore. Restrictions apply.

primal variables

$$g[\mathbf{\Lambda}] = \max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} \boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\mathbf{\Lambda}] \tag{7}$$

where we emphasize the maximization is performed over the set of primal variables satisfying the box constraints. The dual problem is finally defined as

$$D = \min_{\mathbf{\Lambda}\geq 0} g[\mathbf{\Lambda}]. \tag{8}$$

Because (1)-(4) is non-convex it may be the case that $D > P$, implying that working with the dual problem entails loss of optimality. However, it has been proved that as long as the cumulative distribution function of the fading channels is continuous the duality gap is null [5]:

**Theorem 1 (Ribeiro-Giannakis 2008)** *Let $P$ denote the solution of the primal problem* (1) *and $D$ the solution of its dual in* (8)*. If the channel cumulative distribution function (cdf) is continuous, then*

$$P = D. \tag{9}$$

Because of Theorem 1 the dual problem in (8) can be solved in lieu of the primal problem (5). THis fact can be exploited to show that if optimal Lagrange multipliers are known, layering can be rendered optimal in wireless networking [5]. In this paper we study a subgradient descent algorithm and show that it induces an optimal separation in layers and layer interfaces.

## 3. COMPUTATION OF OPTIMAL LAGRANGE MULTIPLIERS

Solving the optimal wireless networking problem in (1) can be reduced to finding the optimal dual variables $\mathbf{\Lambda}^*$ of (8). Because the dual function $g(\mathbf{\Lambda})$ is convex, descent algorithms can be used to find $\mathbf{\Lambda}^*$. However, $g(\mathbf{\Lambda})$ need not be differentiable, and it certainly will not be in some cases. The challenge is therefore to find such descent direction. This prompts the definition of subgradient that we introduce next.

**Definition 1 Subgradient** *We say that $\check{\boldsymbol{g}}(\mathbf{\Lambda}_0)$ is a subgradient of the convex dual function $g(\mathbf{\Lambda})$ at $\mathbf{\Lambda} = \mathbf{\Lambda}_0$ if for every $\mathbf{\Lambda} \geq 0$ we have*

$$g(\mathbf{\Lambda}) \geq g(\mathbf{\Lambda}_0) + \check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda} - \mathbf{\Lambda}_0). \tag{10}$$

The hyperplane $g(\mathbf{\Lambda}_0) + \check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda} - \mathbf{\Lambda}_0) = 0$ defined by the subgradient direction $\check{\boldsymbol{g}}(\mathbf{\Lambda}_0)$ and the point $(\mathbf{\Lambda}_0, g(\mathbf{\Lambda}_0))$ supports $g(\mathbf{\Lambda})$ in the sense that it touches $g(\mathbf{\Lambda})$ at $\mathbf{\Lambda} = \mathbf{\Lambda}_0$ and is below $g(\mathbf{\Lambda})$ at any other point. The fundamental property of a subgradient is that it always points toward the optimal argument. Formally, let $\mathbf{\Lambda} = \mathbf{\Lambda}^*$ in (10), and reorder terms to obtain

$$\check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda}_0 - \mathbf{\Lambda}^*) \geq g[\mathbf{\Lambda}_0] - g(\mathbf{\Lambda}^*) = g[\mathbf{\Lambda}_0] - D \geq 0 \tag{11}$$

where we replaced $g(\mathbf{\Lambda}^*) = D$ and use the fact that $D$ is the minimum value of $g(\mathbf{\Lambda})$. Given that the inner product of $\check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)(\mathbf{\Lambda}_0 - \mathbf{\Lambda}^*)$ is positive, (11) proves that the angle between $\check{\boldsymbol{g}}^T(\mathbf{\Lambda}_0)$ and $\mathbf{\Lambda}_0 - \mathbf{\Lambda}^*$ is less than $\pi/2$. Therefore, the negative of the subgradient points "towards", i.e., with an angle of less than $\pi/2$ radians, the optimal argument.

A subgradient of the dual function can be obtained from the arguments that maximize the Lagrangian for given $\mathbf{\Lambda}$ multipliers as detailed by the following theorem. This as well as subsequent results in Theorems 3 - 5 are known for finite-dimensional optimization problems, [4]. We present them here for the (infinite-dimensional) variational problem (5). The proofs here are patterned after those in [4]; see [5].

**Theorem 2** *With $\mathbf{\Lambda}_0 \geq 0$ an arbitrary dual variable and $\boldsymbol{X}^\dagger(\mathbf{\Lambda}_0)$ primal variables that maximize the Lagrangian function in* (6) *for $\mathbf{\Lambda} = \mathbf{\Lambda}_0$*

$$\left(\boldsymbol{X}^\dagger(\mathbf{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h},\mathbf{\Lambda}_0)\right) \in \arg\max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} \boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\mathbf{\Lambda}_0]. \tag{12}$$

*Then a subgradient of the dual function at $\mathbf{\Lambda} = \mathbf{\Lambda}_0$ is given by*

$$\check{\boldsymbol{g}}(\mathbf{\Lambda}_0) = \boldsymbol{h}\left[\boldsymbol{X}^\dagger(\mathbf{\Lambda}_0), \boldsymbol{p}^\dagger(\boldsymbol{h},\mathbf{\Lambda}_0)\right]. \tag{13}$$

In general, there is more than one argument maximizing (12). Therefore the $\arg\max$ operator does not specify a value but a set, as signified by the $\in$ symbol in (12). We interpret $\boldsymbol{X}^\dagger(\mathbf{\Lambda})$ as any element of this set.

### 3.1. Subgradient descent algorithm

A descent algorithm to compute optimal multipliers $\mathbf{\Lambda}^*$ and minimum dual value $D = P$ is obtained using the subgradient of the dual function described in Theorem 2. With iterations indexed on $t$, start with given dual variables $\mathbf{\Lambda}(t)$ and compute arguments $[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h},t)]$ that maximize the Lagrangian in (6),

$$\begin{aligned}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h},t)] &\in \arg\max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} \boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\mathbf{\Lambda}(t)] \\ &= \arg\max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} f(\boldsymbol{X}) + \mathbf{\Lambda}^T(t)\boldsymbol{h}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h})] \end{aligned} \tag{14}$$

Using (13) we have that a subgradient of the dual function at $\mathbf{\Lambda} = \mathbf{\Lambda}(t)$ is given by $\check{\boldsymbol{g}}(t) := \check{\boldsymbol{g}}[\mathbf{\Lambda}(t)] = \boldsymbol{h}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h},t)]$. Therefore, the dual variable is updated as

$$\begin{aligned}\mathbf{\Lambda}(t+1) &= \Big[\mathbf{\Lambda}(t) - \epsilon_t \check{\boldsymbol{g}}(t)\Big]^+ \\ &= \Big[\mathbf{\Lambda}(t) - \epsilon_t \boldsymbol{h}[\boldsymbol{X}(t), \boldsymbol{p}(\boldsymbol{h},t)]\Big]^+. \end{aligned} \tag{15}$$

where $[\cdot]^+$ denotes the componentwise maximum of 0 and the value between parenthesis and $\epsilon_t$ is a properly selected step-size; see Theorems 4 and 5. Because the negative of the subgradient $-\check{\boldsymbol{g}}(t)$ points towards $\mathbf{\Lambda}^*$ it is expected that iterates of (15) are progressively closer to $\mathbf{\Lambda}^*$. As the following standard result shows, this is indeed true in some sense.

**Theorem 3** *Consider the subgradient descent iteration in* (15) *and define the dual value at iteration $t$ as $g(t) := g[\mathbf{\Lambda}(t)]$. Let $G := \max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} \left\|\boldsymbol{h}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h})]\right\|$ be a bound on the norm of the subgradient of the dual function. The 2-norm distances $\left\|\mathbf{\Lambda}(t) - \mathbf{\Lambda}^*\right\|$ of iterates $\mathbf{\Lambda}(t)$ to the optimal argument $\mathbf{\Lambda}^*$ at times $t$ and $t+1$ satisfy the relation*

$$\left\|\mathbf{\Lambda}(t+1) - \mathbf{\Lambda}^*\right\|^2 \leq \left\|\mathbf{\Lambda}(t) - \mathbf{\Lambda}^*\right\|^2 + \epsilon_t^2 G^2 - 2\epsilon_t\big[g(t) - D\big]. \tag{16}$$

Because all primal variables are constrained to the bounded region $B$, the bound $G$ on the subgradient norm is finite. Given that $D$ denotes the minimum of $g(t)$ it is clearly true that $g(t) - D \geq 0$. Thus, at each iteration the distance between the current dual iterate $\mathbf{\Lambda}(t)$ and the optimal dual variable $\mathbf{\Lambda}^*$ is reduced by (at least) $2\epsilon_t\big[g(t) - D\big]$ and increased by (at most) $\epsilon_t^2 G^2$. For small $\epsilon_t$ we expect the reduction $2\epsilon_t\big[g(t) - D\big]$ to dominate the increase $\epsilon_t^2 G^2$ and consequently for $\mathbf{\Lambda}(t)$ to approach $\mathbf{\Lambda}^*$.

For fixed step size $\epsilon_t = \epsilon$ for all $t$, however, there is a limit on how close $\mathbf{\Lambda}(t)$ can come to $\mathbf{\Lambda}^*$. For any given $\epsilon$, $\epsilon^2 G^2$ will eventually become larger than $2\epsilon\big[g(t) - D\big]$ preventing the optimality gap $\big[g(t) - D\big]$ to go to zero. This is not a limitation of the analysis but a consequence of the fact that for non-differentiable functions the norm of the subgradient $\|\check{\boldsymbol{g}}(\mathbf{\Lambda})\|$ does not necessarily vanish as $\mathbf{\Lambda}$ approaches $\mathbf{\Lambda}^*$. Therefore, the iteration in (3.1) is not convergent. Rather, the iterates $\mathbf{\Lambda}(t)$ approach $\mathbf{\Lambda}^*$ until $\epsilon^2 G^2$ starts dominating $2\epsilon\big[g(t) - D\big]$.

This motivates the use of vanishing step-size sequences, i.e., $\lim_{t\to\infty} \epsilon_t = 0$, so that as the duality gap $\big[g(t) - D\big]$ approaches zero, so does $\epsilon_t$. This allows for $2\epsilon_t\big[g(t) - D\big]$ to always dominate $\epsilon_t^2 G^2$ leading to the following classical convergence result.

**Theorem 4** *Consider the subgradient descent iteration in* (15) *with vanishing step sizes $\epsilon_t$. Require the sum of step sizes to be divergent, i.e.,*

$$\sum_{t=1}^{\infty} \epsilon_t = \infty, \quad \lim_{t\to\infty} \epsilon_t = 0. \tag{17}$$

*Then, the limit of the sequence of iterates $\mathbf{\Lambda}(t)$ exists and*

$$\lim_{t\to\infty} \mathbf{\Lambda}(t) = \mathbf{\Lambda}^*. \tag{18}$$

The conditions (17) on the step-size sequence are certainly minimal. E.g., sequences of the form $\epsilon_t = \epsilon_1/(t+\epsilon_2)^\alpha$ with $\alpha > 0$ for arbitrary positive constants $\epsilon_1$ and $\epsilon_2$ satisfy (17). Nonetheless, constant step sizes $\epsilon_t = \epsilon$ for all $t$, are still desirable in some cases. In this case it can be proven that as $t \to \infty$, $\mathbf{\Lambda}(t)$ "stays close" to $\mathbf{\Lambda}^*$.



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 13, 2009 at 10:48 from IEEE Xplore. Restrictions apply.

**Theorem 5** *Consider the subgradient descent iteration in* (15) *with constant step sizes* $\epsilon_t = \epsilon$ *for all* $t$*. With* $G := \max_{(\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}))\in B} \left\|\boldsymbol{h}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h})]\right\|$ *the subgradient norm bound of Theorem 3, it holds:*

**(i)** *The best dual value at time* $t$*,* $g_{\text{best}}(t) := \min_{s\in[0,t]} g(t)$*, converges to a value within* $\epsilon G^2/2$ *of the optimum* $D$*, i.e.,*

$$\lim_{t\to\infty} g_{\text{best}}(t) - D \leq \epsilon G^2/2. \tag{19}$$

**(ii)** *The average of the dual iterates* $\bar{\boldsymbol{\Lambda}}(t) := (1/t)\sum_{s=1}^{t} \boldsymbol{\Lambda}(s)$*, converges to a point whose optimality gap is less than* $\epsilon G^2/2$*, i.e.,*

$$g\left[\lim_{t\to\infty} \bar{\boldsymbol{\Lambda}}(t)\right] - D \leq \epsilon G^2/2. \tag{20}$$

As commented after Theorem 3, the subgradient descent algorithm (14) - (15) does not necessarily converge for fixed step sizes. Nonetheless, a reasonable approximation to $\boldsymbol{\Lambda}^*$ is achieved by $\boldsymbol{\Lambda}_{\text{best}}(t)$ defined as the argument for which $g[\boldsymbol{\Lambda}_{\text{best}}(t)] = g_{\text{best}}(t)$. The quality of this approximation is measured in the optimality gap $g[\boldsymbol{\Lambda}_{\text{best}}(t)] - D$ that can be made arbitrarily small with adequately selected step size $\epsilon$.

By its own definition $\boldsymbol{\Lambda}_{\text{best}}(t)$ is the best approximation to $\boldsymbol{\Lambda}^*$ that can be obtained by (14) - (15) with fixed step sizes. Finding $\boldsymbol{\Lambda}_{\text{best}}(t)$, though, requires access to the dual values $g(t)$, which might not be available; see e.g., Section 3.2. In such circumstances a similarly good approximation to $\boldsymbol{\Lambda}^*$ is the average $\bar{\boldsymbol{\Lambda}}(t)$ of iterates $\boldsymbol{\Lambda}(t)$.

### 3.2. Layers and layer interfaces

Implementing the subgradient descent iteration (14)-(15) uncovers details in the interaction between layers. Define components of the Lagrange multipliers $\boldsymbol{\Lambda}$ so that $\lambda_{ij}$ is associated with the capacity constraints in (2), $\mu_i$ with the power constraint in (3), and $\nu_i^k$ and $\xi_{ij}$ with the flow and rate constraints in (4). Using this explicit notation the Lagrangian $\boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda}(t)]$ used for the primal iteration in (14) becomes

$$\begin{aligned}\boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda}(t)] &= \\ &= \sum_{i,k}\Big(U_i^k(a_i^k) - \nu_j^k(t)a_j^k\Big) + \sum_i \Big(\mu_i(t)p_i - V_i(p_i)\Big) \\ &+ \sum_{i,j}\Big(\xi_{ij}(t) - \lambda_{ij}(t)\Big)c_{ij} + \sum_{i,j,k}\Big(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\Big)r_{ij}^k \\ &+ \mathrm{E}_{\boldsymbol{h}}\left[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}\Big(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h})\Big) - \mu_i(t)p_{ij}^f(\boldsymbol{h})\right].\end{aligned} \tag{21}$$

Except for the last term $\mathrm{E}_{\boldsymbol{h}}[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}\Big(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h})\Big) - \mu_i(t)p_{ij}^f(\boldsymbol{h})]$, the Lagrangian $\boldsymbol{\mathcal{L}}[\boldsymbol{X},\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{\Lambda}(t)]$ is a sum of terms that depend on one primal variable only. The first term is a weighted sum of $a_i^k$ variables for all $i,k$, the second sums different powers $p_i$, the third one sums all links capacities $c_{ij}$ and the last one does the same for all $r_{ij}^k$. Therefore, the maximization required for the primal iteration (14) can be separated in specific subproblems associated with each of these variables. The elements of $\boldsymbol{X}(t)$ in (14) are thus

$$a_i^k(t) = \max_{a_{\min i}^k \leq a_i^k \leq a_{\max i}^k} \left[U_i^k(a_i^k) - \nu_i^k(t)a_i^k\right] \tag{22}$$

$$r_{ij}^k(t) = \max_{0\leq r_{ij}^k \leq r_{\max}} \left[\Big(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\Big)r_{ij}^k\right] \tag{23}$$

$$c_{ij}(t) = \max_{0\leq c_{ij}\leq c_{\max}} \left[\left(\xi_{ij}(t) - \lambda_{ij}(t)\right)c_{ij}\right] \tag{24}$$

$$p_i(t) = \max_{0\leq p_i \leq p_{\max i}} \left[\mu_i(t)p_i - V_i(p_i)\right] \tag{25}$$

Also, in the last term of (21), the maximization can be brought into the expected value operator. The elements of the power distribution $\boldsymbol{p}(\boldsymbol{h};t)$ in (14) can thus be computed separately for each fading state $\boldsymbol{h}$, i.e.,

$$\boldsymbol{p}(\boldsymbol{h};t) = \max_{0\leq p_{ij}^f(\boldsymbol{h})\leq p_{\max}} \left[\sum_{i,j,f}\lambda_{ij}(t)C_{ij}\Big(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h})\Big) - \mu_i(t)p_{ij}^f(\boldsymbol{h})\right]. \tag{26}$$

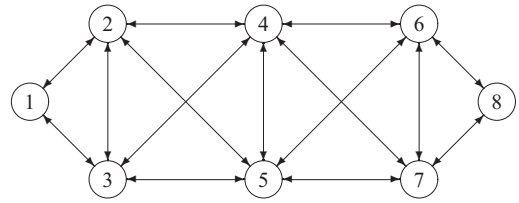


**Fig. 1**. Connectivity graph of an example wireless network.

The subgradient $\check{\boldsymbol{g}}(t) = \boldsymbol{h}[\boldsymbol{X}(t),\boldsymbol{p}(\boldsymbol{h};t)]$ can be likewise separated. The components of the vector function $\boldsymbol{h}[\boldsymbol{X}(t),\boldsymbol{p}(\boldsymbol{h};t)]$ are as specified in (2)-(4). Therefore, the dual iteration (15) can be written explicitly as

$$\nu_i^k(t+1) = \nu_i^k(t) + \epsilon_t\left[\sum_{j\in n(i)}\Big(r_{ij}^k(t) - r_{ji}^k(t)\Big) - a_i^k(t)\right] \tag{27}$$

$$\xi_{ij}(t+1) = \xi_{ij}(t) + \epsilon_t\left[c_{ij}(t) - \sum_k r_{ij}^k(t)\right] \tag{28}$$

$$\lambda_{ij}(t+1) = \lambda_{ij}(t) + \epsilon_t\left[\mathrm{E}_{\boldsymbol{h}}\left[\sum_{f\in\boldsymbol{\mathcal{F}}} C_{ij}\Big(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h};t)\Big)\right] - c_{ij}(t)\right] \tag{29}$$

$$\mu_i(t+1) = \mu_i(t) + \epsilon_t\left[p_i - \mathrm{E}_{\boldsymbol{h}}\left[\sum_{j\in n(i)}\sum_{f\in\boldsymbol{\mathcal{F}}} p_{ij}^f(\boldsymbol{h};t)\right]\right] \tag{30}$$

The argument to be optimized in (22) is solely parameterized by $\nu_i^k(t)$. Thus, given the multiplier $\nu_i^k(t)$ associated with the flow conservation constraint, $a_i^k(t)$ is determined. Likewise, $r_{ij}^k(t)$ is determined by flow conservation multipliers $\nu_i^k(t)$ and $\nu_j^k(t)$ and link capacity constraints multipliers $\xi_{ij}(t)$. In general, all the primal iterations (22)-(26) depend on multipliers associated with no more than two types of constraints. The dual iterations (29)-(30) have a similar property. The update of $\mu_i(t)$ in (30) for instance, depends on the total power $p_i(t)$ and the power distribution $\boldsymbol{p}(\boldsymbol{h};t)$. In general, the multipliers' updates depend on no more than two different types of primal variables.

The fact that primal and dual variable updates depend on only two types of variables prompts a interpretation of (22)-(30) in terms of layers and layer interfaces. The flow control problem (22) is associated with the transport layer, the link rate problem (23) with the routing layer, link capacity (24) and power control (25) problems are solved at the link layer and power distribution (26) pertains to the physical layer. Because the dual variables in (22)-(26) are not optimal, it becomes necessary to communicate variables across layer interfaces. These interfaces are defined by the dual variable updates (27)-(30). Thus, the update of multipliers $\nu_i^k(t)$ in (27) defines the interface between the network and transport layer and (28) the link to network layer interface. Because there are two problems being solved at the link layer, (29) defines the interface between the physical layer and the link capacity subproblem and (30) between physical layer and power control subproblem.

Fig. 2 shows a schematic representation of the layers and their interfaces. At the bottom of the stack the physical layer solves (26) to find the power distribution $\boldsymbol{p}(\boldsymbol{h};t)$. Due to coupling that in general is introduced by the function $C_{ij}\left(\boldsymbol{h}^f,\boldsymbol{p}^f(\boldsymbol{h})\right)$ the physical layer optimization cannot be separated in per-terminal optimization problems and is therefore represented as a common substrate supporting per-terminal stacks. To compute $\boldsymbol{p}(\boldsymbol{h};t)$ the physical layer receives multipliers $\lambda_{ij}(t)$ and $\mu_i(t)$ from the physical-link interface.

At the link layer each terminal maintains variables representing the average link capacities $c_{ij}(t)$ to neighbors $T_j$, $j \in n(i)$ and the average transmitted power $p_i(t)$. These are computed by solving (24) and (25). In turn, this requires dual variables $\lambda_{ij}(t)$ and $\mu_i(t)$ communicated from the physical-link interface and $\xi_{ij}(t)$ communicated from the link-network interface.



Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 13, 2009 at 10:48 from IEEE Xplore. Restrictions apply.

Layers and interfaces at terminal $T_i$

| | |
|---|---|
| $a_i^k(t) = \max_{a^k_{\min i} \le a_i^k \le a^k_{\max i}} \left[U_i^k(a_i^k) - \nu_i^k(t) a_i^k\right]$ | $a_i^k$ for all $k$ |
| $\nu_i^k(t+1) = \nu_i^k(t) + \epsilon_t \left[\sum_{j\in n(i)} \left(r_{ij}^k(t) - r_{ji}^k(t)\right) - a_i^k(t)\right]$ | $\nu_i^k$ for all $k$ |
| $r_{ij}^k(t) = \max_{0\le r_{ij}^k \le r_{\max}} \left[\left(\nu_i^k(t) - \nu_j^k(t) - \xi_{ij}(t)\right) r_{ij}^k\right]$ | $r_{ij}^k$ for $j \in n(i)$ and all $k$ |
| $\xi_{ij}(t+1) = \xi_{ij}(t) + \epsilon_t \left[c_{ij}(t) - \sum_k r_{ij}^k(t)\right]$ | $\xi_{ij}$ for $j \in n(i)$ |
| $c_{ij}(t) = \max_{0\le c_{ij}\le c_{\max}} \left[\left(\xi_{ij}(t) - \lambda_{ij}(t)\right) c_{ij}\right]$ | $c_{ij}$ for $j \in n(i)$ |
| $p_i(t) = \max_{0\le p_i \le p_{\max i}} \left[\mu_i(t) p_i - V_i(p_i)\right]$ | $p_i$ |
| $\lambda_{ij}(t+1) = \lambda_{ij}(t) + \epsilon_t \left[\mathbb{E}_{\boldsymbol{h}}\left[\sum_{f\in\boldsymbol{\mathcal{F}}} C_{ij}\left(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h};t)\right)\right] - c_{ij}(t)\right]$ | $\lambda_{ij}$ for $j \in n(i)$ |
| $\mu_i(t+1) = \mu_i(t) + \epsilon_t \left[p_i - \mathbb{E}_{\boldsymbol{h}}\left[\sum_{j\in n(i)} \sum_{f\in\boldsymbol{\mathcal{F}}} p_{ij}^f(\boldsymbol{h};t)\right]\right]$ | $\mu_i$ |

$$\boldsymbol{p}(\boldsymbol{h},t) = \max_{0\le p_{ij}^f(\boldsymbol{h}) \le p_{\max}} \left[\sum_{i,j,f} \lambda_{ij}(t) C_{ij}\left(\boldsymbol{h}^f, \boldsymbol{p}^f(\boldsymbol{h})\right) - \mu_i(t) p_{ij}^f(\boldsymbol{h})\right]$$

**Fig. 2**. The subgradient descent iteration (22)-(30) can be interpreted in terms of layers and layer interfaces. Layers keep variables of interest to the network, e.g., link transmission rates $r_{ij}^k$ at the network layer, that they update according to primal iterations (22)-(26). Layer interfaces maintain (auxiliary) dual variables updated as per the dual iterations (27)-(30). Communication of variables across layers and interfaces is restricted to adjacent entities; i.e., layers receive variables from, and transmit to, adjacent interfaces. Interfaces exchange variables with adjacent layers. Note that in general the physical layer optimization problem cannot be separated in per-terminal problems.

As is true for physical and link, all layers compute network variables of interest based on dual variables received from adjacent interfaces. That way, the network layer maintains variables $r_{ij}^k$ for neighbors $j \in n(i)$ and flows $k$ that determine local routing decisions. These are updated as per (23) using multipliers $\xi_{ij}(t)$ received from the link-network interface and $\nu_i^k$ and $\nu_j^k$, $j \in n(i)$ from the network-transport interface. The transport layer, finally, keeps variables $a_i^k$ determining the rate at which packets pertaining to the $k$-th flow are accepted into the network by terminal $T_i$. These are updated as per (22) using multipliers $\nu_i^k$ received from the network-transport interface.

Interfaces in turn, update dual variables using information received from adjacent layers. The physical-link interface computes dual variables $\lambda_{ij}(t)$ for $j \in n(i)$ and $\mu_i(t)$. This is fitting because the multipliers $\lambda_{ij}(t)$ and $\mu_i(t)$ are respectively associated with the link capacity (2) and power (3) constraints that relate physical-level variables $\boldsymbol{p}(\boldsymbol{h})$ and link-level quantities $c_{ij}$ and $p_i$. The updates (29) and (30) carried at the physical-link interface require variables $\boldsymbol{p}(\boldsymbol{h};t)$ communicated from the physical layer and variables $c_{ij}(t)$ and $p_i(t)$ from the link layer.

Likewise, the link-network interface keeps one multiplier $\xi_{ij}(t)$ per neighbor $T_j$, $j \in n(i)$. These are associated with the rate constraints in (4) that couple link variables $c_{ij}$ and network variables $r_{ij}^k$. Updates of $\xi_{ij}(t)$ are specified in (28), being determined by variables $c_{ij}(t)$ and $r_{ij}^k(t)$ respectively communicated from the link and network layers. The network-transport interface, finally, maintains dual variables $\nu_i^k(t)$ associated with the flow conservation constraints in (4) that couple network $r_{ij}^k$ and transport $a_i^k$ variables. These $\nu_i^k(t)$ variables are updated as per (27) using $r_{ij}^k(t)$ and $a_i^k(t)$ received from the network and transport layer respectively.

As time progresses, interfaces' variables $\lambda_{ij}(t)$, $\mu_i(t)$, $\xi_{ij}(t)$ and $\nu_i^k(t)$ converge to optimal multipliers $\lambda_{ij}^*$, $\mu_i^*$, $\xi_{ij}^*$ and $\nu_i^{k*}$ [cf. Theorem 4] – or a point close to them if the step size $\epsilon_t$ is fixed [cf. (5)] – enabling computation of optimal network variables $\boldsymbol{p}^*(\boldsymbol{h})$, $p_i^*$, $c_{ij}^*$ ${r_{ij}^k}^*$ and $a_i^{k*}$.

## 4. CONCLUDING REMARKS

This paper has described the separability of wireless networking problems into layers and layer interfaces. This was shown as a consequence of the implementation of a subgradient descent algorithm for the dual function. Similar architectures have been reported elsewhere. The main contribution of this paper is to show that this architecture is optimal.

The algorithmic complexity incurred by the layered architecture in Fig. 2 is determined by the complexity of the optimal power allocation problem (26). The design of optimal wireless networks requires algorithms to efficiently solve this problem in a distributed manner.

## 5. REFERENCES


[1] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1452–1463, August 2006.

[2] 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.

[3] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle, "Layering as optimization decomposition," *Proc IEEE*, vol. 95, no. 1, pp. 255–312, January 2007.

[4] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, Heilderberg, Germany, 1985.

[5] A. Ribeiro and G. Giannakis, "Separation theorems of wireless networking," *IEEE Trans. Inf. Theory*, vol. submitted, January 2008, Available at http://www.seas.upenn.edu/ aribeiro/.




Authorized licensed use limited to: University of Pennsylvania. Downloaded on October 13, 2009 at 10:48 from IEEE Xplore. Restrictions apply.