OPTIMAL WIRELESS NETWORKS BASED ON LOCAL CHANNEL STATE INFORMATION

Yichuan Hu, Alejandro Ribeiro

Department of Electrical and Systems Engineering, University of Pennsylvania

ABSTRACT

We consider distributed algorithms to optimize random access multihop wireless networks in the presence of fading. Since the associated optimization problem is neither convex nor amenable to distributed implementation, we introduce a problem approximation. This approximation is still not convex, but it has zero duality gap and can be solved and decomposed into local subproblems in the dual domain. The solution method is through a stochastic subgradient descent algorithm that operates without knowledge of the fading's distribution and leads to an architecture composed of layers and layer interfaces. With limited amount of message passing among terminals and small computational cost, the proposed algorithm converges almost surely in an ergodic sense.

Index Terms— Wireless networking, cross-layer design, random access.

1. INTRODUCTION

Optimal design is emerging as the future paradigm for wireless networking. The fundamental idea is to select operating points as solutions of optimization problems, which, inasmuch as optimization criteria are properly chosen, yield the best possible network. Results in this field include architectural insights, e.g. [1], and protocol design, e.g., [2], but a drawback shared by most of these works is that they rely on global channel state information (CSI); i.e., the optimal variables of a terminal depend on the channels between all pairs of terminals in the network. While availability of global CSI is plausible in certain situations, it is unlikely to hold if time varying fading channels are taken into account.

We consider optimal design of wireless networks when, due to presence of random fading, only local CSI is available. As a consequence, operating variables of each terminal are selected as functions of the channels linking the terminal with neighboring nodes. A further consequence is that random access appears as the natural medium access choice. If transmission decisions depend on local channels only and these channels are random and independent for different terminals, transmission decisions can be viewed as random and resultant link capacities as limited by collisions. Thus, we can restate our goal as the development of algorithms to find optimal operating points of wireless random access networks in the presence of fading. We remark that even though channel-adaptive random access has a long history in wireless communications, e.g., [3, 4], the idea has not been migrated to wireless networks. Indeed, while works on optimal wireless random access networks do exist, e.g., [5, 6], fading is not considered as part of the optimization problem.

The paper begins by introducing an optimization problem that defines the optimal random access network (Section 2). Since this problem is not convex we proceed to a suboptimal approximation through a problem that while still not convex has zero duality gap (Section 2.1). We further observe that solution is simpler in the dual domain – and equivalent because of the lack of duality gap – and proceed to develop stochastic dual descent algorithms that converge to the optimal operating point (Section 3). The resultant algorithm decomposes in a layered architecture and can be implemented in a distributed manner (Section 3.2). Recent results on ergodic stochastic optimization algorithms [7] are finally leveraged to show that the proposed algorithm yields operating points that are almost surely close to optimal (Section 3.1). Remarkable properties of the proposed algorithm are that it does not necessitate access to the channel's probability distribution and that it can handle non-covex functions mapping transmitted power to channel capacity.

Work in this paper is supported by the Army Research Office grant P-57920-NS and the National Science Foundation CAREER award CCF-0952867.

2. PROBLEM FORMULATION

Consider an ad-hoc wireless network consisting of J terminals $\{T_i\}_{i=1}^J$. Network connectivity is modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $v \in \mathcal{V} := \{1, \ldots, J\}$ representing J terminals and edges $e = (i, j) \in \mathcal{E}$ connecting pairs of terminals that can communicate with each other. Denote the neighborhood of T_i as $\mathcal{N}(i) := \{j \mid (i, j) \in \mathcal{E}\}$ and define the interference neighborhood of the link (i, j) as the set of nodes $\mathcal{M}_i(j) := \mathcal{N}(j) \cup \{j\} \backslash \{i\}$ whose transmission can interfere with a transmission from i to j. The network supports a set $\mathcal{K} := \{1, \ldots, K\}$ of end-to-end flows through multihop transmissions. The average rate at which k-flow packets are generated at T_i is denoted by a_i^k. Terminal T_i transmits these packets to neighboring terminals at average rates r_{ij}^k and, consequently, receives k-flow packets from neighbors at average rates r_{ji}^k. Accounting of k-flow packets implies that exogenous rates a_i^k and endogenous rates r_{ij}^k at T_i satisfy

$$a_i^k = \sum_{j \in \mathcal{N}(i)} \left(r_{ij}^k - r_{ji}^k \right), \quad \text{for all } i \in \mathcal{V}, \text{ and } k \in \mathcal{K}. \tag{1}$$

Further denote the capacity of the link from T_i to T_j as c_{ij}. Since packets of different flows k are transmitted from T_i to T_j at rates r_{ij}^k it must be

$$\sum_{k \in \mathcal{K}} r_{ij}^k \leq c_{ij}, \quad \text{for all } (i, j) \in \mathcal{E}. \tag{2}$$

Unlike wireline networks where c_{ij} are fixed, link capacities in wireless networks are dynamic since channel conditions and transmission policies change over time. Let time be divided into slots indexed by t and denote the time-varying block fading channel between T_i and T_j at time t as $h_{ij}(t)$. Channel gains at different times and/or at different links are assumed independent. Channel gains $h_{ij}(t)$ of link (i, j) are further assumed identically distributed with probability distribution function (pdf) $m_{h_{ij}}(\cdot)$. We assume no channel realization has nonzero probability, something that is true for models used in practice. For reference, define the vector of terminal T_i outgoing channels $\mathbf{h}_i(t) := \{h_{ij}(t) | j \in \mathcal{N}(i)\}$ and the vector of all channels $\mathbf{h}(t) := \{h_{ij}(t) | (i, j) \in \mathcal{E}\}$. Denote their pdfs as $m_{\mathbf{h}_i}(\cdot)$ and $m_{\mathbf{h}}(\cdot)$, respectively.

Based on the channel state $\mathbf{h}_i(t)$ of its outgoing links, T_i decides whether to transmit or not on link (i, j) in time slot t by determining the value of a scheduling function $q_{ij}(t) := Q_{ij}(\mathbf{h}_i(t)) \in \{0, 1\}$. If $q_{ij}(t) = 1$, T_i transmits on link (i, j) and remains silent otherwise. We further restrict T_i to communicate with, at most, one neighbor per time slot. Upon defining $q_i(t) := Q_i(\mathbf{h}_i(t)) := \sum_{j \in \mathcal{N}(i)} Q_{ij}(\mathbf{h}_i(t))$ to indicate a transmission from T_i to some neighbor we must have $q_i(t) \in \{0, 1\}$. We emphasize that $q_{ij}(t) := Q_{ij}(\mathbf{h}_i(t))$ depends on local outgoing channels only and not on global CSI. Further note that terminals have access to local CSI $\mathbf{h}_i(t)$, but underlying pdfs $m_{\mathbf{h}_i}(\cdot)$ are unknown.

Besides channel access decisions, terminals also adapt transmission power to local CSI through a power control function $p_{ij}(t) := P_{ij}(\mathbf{h}_i(t))$ taking values in $[0, p_{ij}^{\max}]$. Here, $p_{ij}^{\max}$ represents the maximum allowable instantaneous power on link (i, j). The average power consumed by T_i is then given as the expected value over channel realizations of the sum of $P_{ij}(\mathbf{h}_i)$ over all $j \in \mathcal{N}(i)$, i.e.,

$$p_i = \mathbb{E}_{\mathbf{h}_i} \left[\sum_{j \in \mathcal{N}(i)} P_{ij}(\mathbf{h}_i) Q_{ij}(\mathbf{h}_i) \right]. \tag{3}$$

If T_i transmits to T_j in time slot t, $p_{ij}(t)$ and $h_{ij}(t)$ determine the transmission rate through a function $C_{ij}\big(h_{ij}(t) p_{ij}(t)\big)$, whose form depends on modulation and coding. To keep analysis general we do not restrict $C_{ij}\big(h_{ij}(t)p_{ij}(t)\big)$ to a specific form. We just assume that it is a nonnegative increasing function of the product $h_{ij}(t)p_{ij}(t)$ taking finite values for finite arguments. This restriction is lax enough to allow for discontinuous rate functions that arise in, e.g., adaptive modulation and coding.

Since terminals contend for channel access, a transmission from T_i to T_j in time slot t is successful if a collision does not occur. In turn, this requires that: (i) Terminal i transmits to j, i.e., $q_{ij}(t) = 1$. (ii) Terminal j is silent, i.e., $q_j(t) = 0$. (iii) No other neighbor of j transmits, i.e. $q_k(t) = 0$ for all $k \in \mathcal{N}(j)$ and $k \neq i$. Recalling the definition of the interference neighborhood $\mathcal{M}_i(j)$ and that if the transmission is successful its rate is $C_{ij}\big(h_{ij}(t)p_{ij}(t)\big)$ we can express the instantaneous transmission rate from T_i to T_j in time slot t as $c_{ij}(t) := c_{ij}(\mathbf{h}_i(t)) = C_{ij}\big(h_{ij}(t)p_{ij}(t)\big)q_{ij}(t)\prod_{k\in\mathcal{M}_i(j)}[1-q_k(t)]$. Assuming an ergodic mode of operation, the link capacity can then be written as

$$c_{ij} = \mathbb{E}_{\mathbf{h}}\Bigg[C_{ij}(h_{ij}P_{ij}(\mathbf{h}_i))Q_{ij}(\mathbf{h}_i)\prod_{k\in\mathcal{M}_i(j)}[1-Q_k(\mathbf{h}_k)]\Bigg]. \quad (4)$$

Because terminals are required to make channel access and power control decisions independently of each other, $Q_{ij}(\mathbf{h}_i)$ and $P_{ij}(\mathbf{h}_i)$ are independent of $Q_{kl}(\mathbf{h}_k)$ and $P_{kl}(\mathbf{h}_k)$ for all $i \neq k$. Since $Q_k(\mathbf{h}_k) := \sum_{j\in\mathcal{N}(i)} Q_{kl}(\mathbf{h}_k(t))$ by definition, it follows that $Q_{ij}(\mathbf{h}_i)$ is also independent of $Q_k(\mathbf{h}_k)$ for all $i \neq k$. This allows us to write the expectation of the product on the right hand side of (4) as a product of expectations,

$$c_{ij} \le \mathbb{E}_{\mathbf{h}_i}\Big[C_{ij}\big(h_{ij}P_{ij}(\mathbf{h}_i)\big)Q_{ij}(\mathbf{h}_i)\Big]\prod_{k\in\mathcal{M}_i(j)}\Big[1-\mathbb{E}_{\mathbf{h}_k}\big[Q_k(\mathbf{h}_k)\big]\Big]. \quad (5)$$

Operating points are characterized by variables $\mathbf{x}_i := \big[a_{ij}^k, r_{ij}^k, c_{ij}, p_i\big]$ and functions $\mathbf{P}_i(\mathbf{h}_i) := [P_{ij}(\mathbf{h}_i), Q_{ij}(\mathbf{h}_i)]$, which are subject to box constraints $\mathcal{B}_i$ [8]. We wish to find the optimal operating point for a given wireless network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ defined as values of variables $\mathbf{x}_i$ and functions $\mathbf{P}_i(\mathbf{h}_i)$ that satisfy constraints (1) - (4) and are optimal according to certain criteria. In particular, we are interested in large rates a_i^k and low power consumptions p_i. Define then increasing concave functions $U_i^k(\cdot)$ representing rewards for accepting a_i^k units of information for flow k at T_i and increasing convex functions $V_i(\cdot)$ typifying penalties for consuming p_i units of power at T_i. The optimal network based on local CSI is then defined as the solution of the optimization problem

$$\begin{aligned}\mathsf{P} \;=\; &\max_{\{\mathbf{x}_i,\mathbf{P}_i(\mathbf{h}_i)\}\in\mathcal{B}_i} \sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i\in\mathcal{V}} V_i\left(p_i\right) \quad (6)\\ &\text{s.t.} \qquad \text{constraints (1), (2), (3), (5),}\end{aligned}$$

Our goal is to develop a distributed algorithm to solve (6) without accessing the channel pdf $m_{\mathbf{h}}(\cdot)$. This is challenging because: (i) The optimization space in (6) includes functions $Q_{ij}(\mathbf{h}_i)$ and $P_{ij}(\mathbf{h}_i)$ implying that the dimension of the problem is infinite. (ii) Since the capacity constraint (5) is non-convex and the capacity function may be even discontinuous, (6) is a non-convex optimization problem. (iii) Constraints (3) and (5) involve expectations over channel states $\mathbf{h}$ whose pdf is unknown. (iv) Since the transmission rate c_{ij} is determined not only by the transmitter but also by the receiver and its neighbors [cf. (5)], solving (6) requires a joint optimization over terminals in the network.

2.1. Problem reformulation and approximation

For reasons that will become clear in Section 3 a distributed solution of the problem in (6) is not possible because scheduling functions $Q_{ij}(\mathbf{h}_i)$ and $Q_k(\mathbf{h}_k)$ are coupled as a product in constraint (5). If we reformulate this constraint into an expression in which the terms $C_{ij}\big(h_{ij}P_{ij}(\mathbf{h}_i)\big)Q_{ij}(\mathbf{h}_i)$ and $1 - Q_k(\mathbf{h}_k)$ appear as summands instead of as factors of a product, the problem will become decomposable in the dual domain. This reformulation can be accomplished by taking logarithms on both sides of (5), yielding

$$\begin{aligned}\tilde{c}_{ij} := \log c_{ij} \le \log \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\big(h_{ij}P_{ij}(\mathbf{h}_i)\big)Q_{ij}(\mathbf{h}_i)\right] \quad (7)\\ + \sum_{k\in\mathcal{M}_i(j)} \log\left[1-\mathbb{E}_{\mathbf{h}_k}[Q_k(\mathbf{h}_k)]\right],\end{aligned}$$

where we have defined $\tilde{c}_{ij} := \log c_{ij}$. While scheduling functions of different terminals now appear as summands on the right hand side of (7), the link capacity constraint (2) mutates into the non-convex constraint $\sum_{k\in\mathcal{K}} r_{ij}^k \le e^{\tilde{c}_{ij}}$. To overcome this issue, we use the linear lower bound $e^{\tilde{c}_{ij}} \ge 1 + \tilde{c}_{ij}$ and approximate this constraint as $\sum_{k\in\mathcal{K}} r_{ij}^k \le 1 + \tilde{c}_{ij}$. Upon defining the attempted transmission rate of link (i, j) as $x_{ij} := \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\left(h_{ij}P_{ij}\left(\mathbf{h}_i\right)\right)Q_{ij}\left(\mathbf{h}_i\right)\right]$ and the transmission probability of T_i as $y_i := \mathbb{E}_{\mathbf{h}_i}\left[Q_i\left(\mathbf{h}_i\right)\right]$, the original optimization problem P is approximated by the problem

$$\begin{aligned}\mathsf{P} \ge \tilde{\mathsf{P}} \;=\; &\max_{\{\tilde{\mathbf{x}}_i,\mathbf{P}_i(\mathbf{h}_i)\}\in\tilde{\mathcal{B}}_i} \sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i\in\mathcal{V}} V_i\left(p_i\right) \quad (8)\\ \text{s.t.}\;\; & a_i^k \le \sum_{j\in\mathcal{N}(i)}\left(r_{ij}^k - r_{ji}^k\right), \quad \sum_{k\in\mathcal{K}} r_{ij}^k \le 1+\tilde{c}_{ij},\\ & \tilde{c}_{ij} \le \log x_{ij} + \sum_{k\in\mathcal{M}_i(j)} \log\left(1-y_k\right),\\ & x_{ij} \le \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\big(h_{ij}P_{ij}\left(\mathbf{h}_i\right)\big)Q_{ij}\left(\mathbf{h}_i\right)\right], \;\; y_i \ge \mathbb{E}_{\mathbf{h}_i}\left[Q_i\left(\mathbf{h}_i\right)\right],\\ & p_i \ge \mathbb{E}_{\mathbf{h}_i}\Bigg[\sum_{j\in\mathcal{N}(i)} P_{ij}\left(\mathbf{h}_i\right)Q_{ij}\left(\mathbf{h}_i\right)\Bigg],\end{aligned}$$

where we defined $\tilde{\mathbf{x}}_i := [\mathbf{x}_i, x_{ij}, y_i]$ and relaxed the definitions of attempted transmission rate and transmission probability, which we can do without loss of optimality. Problems (6) and (8) are *not* equivalent because of the linear approximation to the link capacity constraint. However, since $1 + \tilde{c}_{ij}$ is a lower bound on $e^{\tilde{c}_{ij}}$, any operating point that satisfies the constraints in (8) also satisfies the constraints in (6). In particular, the solution of (8) is feasible in (6), although possibly suboptimal. Further note that variables associated with different terminals appear as different summands of the objective and constraints of (8). This is the signature of optimization problems amenable to distributed implementations as we explain in the following section.

3. DISTRIBUTED STOCHASTIC OPTIMIZATION

Define vectors $\tilde{\mathbf{x}}$ and $\mathbf{P}(\mathbf{h})$ grouping $\tilde{\mathbf{x}}_i$ and $\mathbf{P}_i(\mathbf{h}_i)$ for all $i \in \mathcal{V}$, and introduce Lagrange multiplier $\mathbf{\Lambda} = [\lambda_i^k, \mu_{ij}, \nu_{ij}, \alpha_{ij}, \beta_i, \xi_i]^T$, where λ_i^k is associated with the flow conservation constraint, μ_{ij} with the rate constraint, ν_{ij} with the link capacity, α_{ij} with the attempted transmission rate, β_i with the transmission probability, and ξ_i with the average power. The Lagrangian for the optimization problem in (8) is given by the sum of the objective and the products of the constraints with their respective multipliers. After reordering terms, we can write the Lagrangian as

$$\mathcal{L}\left(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda}\right) = \sum_{i\in\mathcal{V}} \mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}\right) + \mathbb{E}_{\mathbf{h}_i}\Big[\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda})\Big]. \quad (9)$$

The local Lagrangian component $\mathcal{L}_i^{(1)}\big(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}\big)$ in (9) is defined as

$$\begin{aligned}\mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}\right) := &\sum_k U_i^k\left(a_i^k\right) - \lambda_i^k a_i^k + \sum_{j\in\mathcal{N}_i}\left(\lambda_i^k - \lambda_j^k - \mu_{ij}\right) r_{ij}^k\\ &+ \sum_{j\in\mathcal{N}_i}\big(\mu_{ij} - \nu_{ij}\big)\tilde{c}_{ij} + \big(\xi_i p_i - V_i(p_i)\big) + \sum_{j\in\mathcal{N}_i}\left[\nu_{ij}\log x_{ij} - \alpha_{ij}x_{ij}\right]\\ &+ \beta_i y_i + \Bigg[\sum_{k\in\mathcal{N}(i)}\Bigg(\nu_{ki} + \sum_{l\in\mathcal{N}(k), l\neq i} \nu_{lk}\Bigg)\Bigg]\log(1-y_i). \quad (10)\end{aligned}$$

Transport Layer

$$a_i^k(t) = \text{argmax}_{a_i^{\min} \le a_i^k \le a_i^{\max}} \left\{ U_i^k\left(a_i^k\right) - \lambda_i^k(t) a_i^k \right\}$$

$$\lambda_i^k(t+1) = \left[\lambda_i^k(t) - \epsilon \left[\sum_{j\in\mathcal{N}(i)} \left(r_{ij}^k(t) - r_{ji}^k(t)\right) - a_i^k(t)\right]\right]^+$$

Network Layer

$$r_{ij}^k(t) = \text{argmax}_{0 \le r_{ij}^k \le r_{ij}^{\max}} \left\{ \left(\lambda_i^k(t) - \lambda_j^k(t) - \mu_{ij}(t)\right) r_{ij}^k \right\}$$

$$\mu_{ij}(t+1) = \left[\mu_{ij}(t) - \epsilon \left[1 + \tilde{c}_{ij}(t) - \sum_{k\in\mathcal{K}} r_{ij}^k(t)\right]\right]^+$$

Link Layer

$$\tilde{c}_{ij}(t) = \text{argmax}_{0 \le \tilde{c}_{ij} \le \tilde{c}_{ij}^{\max}} \left\{\mu_{ij}(t)\tilde{c}_{ij} - \nu_{ij}(t)\tilde{c}_{ij}\right\}$$

$$\nu_{ij}(t+1) = \left[\nu_{ij}(t) - \epsilon \left[\log x_{ij}(t) + \sum_{k\in\mathcal{M}_i(j)} \log\left(1 - y_k(t)\right) - \tilde{c}_{ij}(t)\right]\right]^+$$

MAC Layer

$$p_i(t) = \text{argmax}_{0 \le p_i \le p_i^{\max}} \left\{\xi_i(t) p_i - V_i\left(p_i\right)\right\}$$
$$x_{ij}(t) = \text{argmax}_{x_{ij} \ge 0} \left\{\nu_{ij}(t) \log x_{ij} - \alpha_{ij}(t) x_{ij}\right\}$$
$$y_i(t) = \text{argmax}_{0 \le y_i \le 1} \left\{ \beta_i(t) y_i + \left[\sum_{k\in\mathcal{N}(i)} \left[\nu_{ki}(t) + \sum_{l\in\mathcal{N}(k), l\neq i} \nu_{lk}(t)\right]\right] \log\left(1 - y_i\right)\right\}$$

$$\alpha_{ij}(t+1) = \left[\alpha_{ij}(t) - \epsilon\left[C_{ij}\left(h_{ij}(t) p_{ij}(t)\right) q_{ij}(t) - x_{ij}(t)\right]\right]^+$$
$$\beta_i(t+1) = \left[\beta_i(t) - \epsilon\left[y_i(t) - q_i(t)\right]\right]^+$$
$$\xi_i(t+1) = \left[\xi_i(t) - \epsilon\left[p_i(t) - \sum_{j\in\mathcal{N}(i)} q_{ij}(t) p_{ij}(t)\right]\right]^+$$

Physical Layer

$$\{p_{ij}(t), q_{ij}(t)\} = \text{argmax}_{p_{ij}\in\left[0, p_{ij}^{\max}\right], q_{ij}\in\{0,1\}} \left\{ \sum_{j\in\mathcal{N}(i)} q_{ij}\left[\alpha_{ij}(t) C_{ij}(h_{ij}(t) p_{ij}) - \beta_i(t) - \xi_i(t) p_{ij}\right]\right\}$$

Fig. 1. Layers and layer interfaces. Layers maintain primal variables $a_i^k(t)$, $r_{ij}^k(t)$, $\tilde{c}_{ij}(t)$, $p_i(t)$, $p_{ij}(t)$ and $q_{ij}(t)$ while multipliers $\lambda_i^k(t)$, $\mu_{ij}(t)$, $\nu_{ij}(t)$, $\alpha_{ij}(t)$, $\beta_i(t)$ and $\xi_i(t)$ and auxiliary variables $x_{ij}(t)$ and $y_i(t)$ are associated with interfaces between adjacent layers. The proposed stochastic subgradient algorithm can be regarded as an information exchange mechanism among different layers.

and the local per channel component $\mathcal{L}_i^{(2)}\left(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}\right)$ as

$$\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}) := \sum_{j\in\mathcal{N}(i)} Q_{ij}\left(\mathbf{h}_i\right)\left[\alpha_{ij} C_{ij}\left(h_{ij} P_{ij}\left(\mathbf{h}_i\right)\right) - \beta_i - \xi_i P_{ij}\left(\mathbf{h}_i\right)\right]. \tag{11}$$

The dual function is then defined as the maximum of the Lagrangian (9) over the set of feasible $\tilde{\mathbf{x}}_i$ and $\mathbf{P}_i(\mathbf{h}_i)$ and the dual problem is the minimum of $g(\mathbf{\Lambda})$ over positive dual variables

$$\tilde{\mathsf{D}} = \min_{\mathbf{\Lambda}\ge 0} g(\mathbf{\Lambda}) = \min_{\mathbf{\Lambda}\ge 0} \max_{\{\tilde{\mathbf{x}}_i, \mathbf{P}_i(\mathbf{h}_i)\}\in\tilde{\mathcal{B}}_i} \mathcal{L}\left(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda}\right). \tag{12}$$

Despite being non-convex, the structure of the problem in (8) is such that $\tilde{\mathsf{P}} = \tilde{\mathsf{D}}$ as long as the fading distribution has no realization with positive probability [8]. This lack of duality gap implies that the finite dimensional convex dual problem is equivalent to the infinite dimensional non-convex primal problem. While this affords a substantial improvement in computational tractability, it does not necessarily mean that solving the dual problem is easy because evaluation of the dual function's value requires maximization of a expected value over the unknown channel distribution $m_{\mathbf{h}_i}(\mathbf{h}_i)$. Still, convexity of the dual function allows the use of descent algorithms in the dual domain. This property is exploited next to develop a stochastic subgradient descent algorithm that solves (12) using observations of instantaneous channel realizations $\mathbf{h}_i(t)$.

3.1. Stochastic Subgradient Descent

Starting from given dual variables $\mathbf{\Lambda}(t)$, for each terminal i the algorithm computes instantaneous primal variables $\tilde{\mathbf{x}}_i(t)$ and $\mathbf{P}_i(t)$ based on local channel realization $\mathbf{h}_i(t)$ in time slot t, and uses these values to update dual variables $\mathbf{\Lambda}(t+1)$. Specifically, the algorithm starts finding primal variables that optimize the summands of the Lagrangian in (9) (the operator $[\cdot]^+$ denotes projection in the positive orthant)

$$\tilde{\mathbf{x}}_i(t) = \underset{\mathbf{x}_i}{\text{argmax}} \left\{\mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}(t)\right)\right\}, \tag{13}$$

$$\mathbf{P}_i(t) = \underset{\mathbf{P}_i}{\text{argmax}} \left\{\mathcal{L}_i^{(2)}\left(\mathbf{P}_i, \mathbf{h}_i(t), \mathbf{\Lambda}(t)\right)\right\}. \tag{14}$$

Based on $\tilde{\mathbf{x}}_i(t)$ and $\mathbf{P}_i(t)$, define the stochastic subgradient $\mathbf{s}(t)$ whose components are the instantaneous constraints violation in problem $\tilde{\mathsf{P}}$; see Fig. 1. Complete the algorithm by introducing a step size ϵ and a descent update in the dual domain along the stochastic subgradient $\mathbf{s}(t)$

$$\mathbf{\Lambda}(t+1) = \left[\mathbf{\Lambda}(t) - \epsilon\mathbf{s}(t)\right]^+. \tag{15}$$

Specific expressions for the maximizations in (13) and (14) and for $\mathbf{s}(t)$ in (15) are given in Fig. 1 and explained in Section 3.2.

The proposed algorithm consists of iterative application (13)-(15). Its outcome is a time sequence of network operating points $a_i^k(\mathbb{N})$, $r_{ij}^k(\mathbb{N})$, $c_{ij}(\mathbb{N})$, $p_i(\mathbb{N})$, $q_{ij}(\mathbb{N})$ and $p_{ij}(\mathbb{N})$. It can be shown that the expected value of the stochastic subgradient $\mathbf{s}(t)$ is a subgradient of the dual function [7] which implies that, on average, $\mathbf{s}(t)$ points towards the set of optimal dual variables $\mathbf{\Lambda}^*$. Thus, it is reasonable to expect iterates of (13)-(15) to approach $\mathbf{\Lambda}^*$ in some sense. While this can be proved true and leveraged to prove almost sure convergence of primal iterates $\tilde{\mathbf{x}}(t)$ to an optimal operating point of problem (8) in an ergodic sense [7], it is not clear whether the ergodic averages of $\tilde{\mathbf{x}}(t)$ are also feasible for the original problem (6). The reason for this mismatch is that constraints in (8) are satisfied in an ergodic, i.e., time average, sense. When these ergodic limits are substituted into the capacity expression in (4) we end up with a product of ergodic limits that cannot be equated to the ergodic limit of the product. This issue is resolved in the proof of the following theorem to claim almost sure feasibility and utility yield close to P in an ergodic sense – See [8] for the proof.

Theorem 1 Consider a wireless network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with random access in the physical layer and sequences of network operating points $a_i^k(\mathbb{N})$, $r_{ij}^k(\mathbb{N})$, $c_{ij}(\mathbb{N})$, $p_i(\mathbb{N})$, $q_{ij}(\mathbb{N})$ and $p_{ij}(\mathbb{N})$ generated by algorithm (13) - (15). Assume that the second moment of the norm of the stochastic subgradient $\mathbf{s}(t)$ is finite, i.e., $\mathbb{E}_{\mathbf{h}}\left[\|\mathbf{s}(t)\|^2 \middle| \mathbf{\Lambda}(t)\right] \leq \hat{S}^2$, and that there exists a set of strictly feasible primal variables that satisfy the constraints in problem $\tilde{\mathsf{P}}$ with strict inequality. Then,
(i) **Almost sure feasibility.** The constraints (1) - (4) in problem P are almost surely satisfied in an ergodic sense.
(ii) **Almost sure near optimality.** The ergodic average of of the utility almost surely converges to a value with optimality gap smaller than $\epsilon\hat{S}^2/2$ with respect to $\tilde{\mathsf{P}}$.

As per Theorem 1, the algorithm in Fig. 1 does not find the optimal scheduling and power allocation functions $Q_{ij}(\mathbf{h}_i)$ and $P_{ij}(\mathbf{h}_i)$. Rather, it generates instantaneous schedules and power allocations based on local channel states $\mathbf{h}_i(t)$ that are optimal in an ergodic sense. This is not a drawback because the latter property is sufficient for a practical implementation. Further note that the use of constant step sizes ϵ endows the algorithm with adaptability to time-varying channel distributions.

3.2. Layers, Layer Interfaces, and message passing

The maximization of the local Lagrangian component $\mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}\right)$ required in (13) can be separated into separate maximizations with respect to a_i^k, r_{ij}^k, c_{ij}, p_i, x_{ij}, and y_i [cf. (10)]. The corresponding maximizations are shown as the transport, network, link and MAC layers in Fig. 1. Notice that each of these maximizations involves a single variable. The maximization of the local per channel component $\mathcal{L}_i^{(2)}\left(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}\right)$ required in (14) determines the optimal schedules $q_{ij}(t)$ and instantaneous power allocations $p_{ij}(t)$. This is shown as the physical layer in Fig. 1 and cannot be decomposed further. Notice however that at most one of the scheduling variables can be 1. Therefore, the maximization can be easily accomplished by determining which of the variables $q_{ij}(t)$ has to be set to 1 in order to maximize $\mathcal{L}_i^{(2)}\left(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}\right)$. If the maximum turns out negative, it is optimal to set all $q_{ij}(t)$ to 0.

The dual updates in (15) require computation of the stochastic subgradient defined as the value of the instantaneous constraint violations corresponding to the Lagrangian maximizers computed in (13) and (14). The multiplier λ_i^k is associated with the flow conservation constraint. Therefore, the corresponding subgradient is given as $\sum_{j\in\mathcal{N}(i)}\left(r_{ij}^k(t) - r_{ji}^k(t)\right) - a_i^k(t)$. This is shown as an interface between the transport and network layer. The rationale for this architectural arrangement is that the λ_i^k update uses variables associated with the transport and network layer. We can similarly determine the updates for μ_{ij} and ν_{ij} as we respectively do in the interfaces between the link and network layer and between the MAC and link layers. The updates on α_{ij}, β_i, and ξ_{ij} are slightly different because their associated constraints involve expected values with respect to the channel pdf. In this case the stochastic subgradient is *not* the constraint violation, but its instantaneous value associated with the current channel realization. E.g., the stochastic subgradient along the β_i direction is $y_i(t) - q_i(t)$ instead of the expected value of $y_i(t) - q_i(t)$. Likewise the stochastic subgradient associated with α_{ij} is $C_{ij}\left(h_{ij}(t)p_{ij}(t)\right)q_{ij}(t) - x_{ij}(t)$, and the one associated with ξ_{ij} is $p_i(t) - \sum_{j\in\mathcal{N}(i)} p_{ij}(t)q_{ij}(t)$. All of these dual updates are shown as part of the interface between the physical and MAC layers.

In Fig. 1 computation of primal and dual variables can be done locally at T_i except for $r_{ij}^k(t)$, $\lambda_i^k(t)$, $\nu_{ij}(t)$ and $y_i(t)$ that require information from neighboring terminals. This implies that a message passing mechanism among terminals is needed. At the beginning of primal iterations, T_i transmits $\lambda_i^k(t)$ and $\nu_{ij}(t)$ to its neighbors and correspondingly receives these two multipliers from each of its neighbors. It then broadcasts $\sum_{k\in\mathcal{N}(i)}\nu_{ki}(t)$ to neighboring nodes. Subsequently, T_j subtracts $\nu_{ji}(t)$ from this sum to obtain $\sum_{k\in\mathcal{N}(i),k\neq j}\nu_{ki}(t)$. Multipliers required for computing primal variables are now available and the layers proceed to update their primal variables. Primal variables are now exchanged between neighboring terminals to proceed with dual updates. Specifically, T_i passes variables $y_i(t)$ and $r_{ij}^k(t)$ to all its neighbors and proceeds to broadcast the sum of all the $y_k(t)$ received in this exchange, i.e. $\sum_{k\in\mathcal{N}(i)} y_k(t)$. Upon receiving this information, T_j adds $y_i(t)$ and subtracts $y_j(t)$ from this sum to obtain $\sum_{k\in\mathcal{M}_j(i)} y_k(t) = \sum_{k\in\mathcal{N}(i)} y_k(t) + y_i(t) - y_j(t)$. Dual updates are now performed at the interfaces and we proceed to the next iteration.

4. CONCLUSION

We developed algorithms for optimal design of wireless networks using local channel state information. Due to the time-varying nature of fading states, random access is the natural medium access choice leading to the formulation of an optimization problem for random access networks. To obtain a distributed solution, we approximated the problem so that it can be decomposed in the dual domain and developed a stochastic subgradient descent algorithm. Based on instantaneous local channel conditions, the algorithm finds network operating points that are almost surely feasible and optimal in an ergodic sense. In addition, the solution exhibits a layered architecture in which variables in each layer are computed using information from interfaces to adjacent layers.

5. REFERENCES

[1] M. Chiang, S. H. Low, R. A. Calderbank, and J. C. Doyle, “Layering as optimization decomposition,” in *Proc. IEEE*, no. 1, Cambridge, UK, Jan. 2007, pp. 255–312.

[2] 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, Aug. 2006.

[3] X. Qin and R. A. Berry, “Distributed approaches for exploiting multiuser diversity in wireless networks,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 392–413, Feb. 2006.

[4] Y. Hu and A. Ribeiro, “Adaptive distributed algorithms for optimal random access channels,” *IEEE Trans. Wireless Commun.*, Feburary 2010, submitted.

[5] X. Wang and K. Kar, “Cross-layer rate optimization for proportional fairness in multihop wireless networks with random access,” *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1548–1559, Aug. 2006.

[6] J. Lee, , M. Chiang, and A. R. Calderbank, “Utility-optimal random-access control,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2741–2751, Jul. 2007.

[7] A. Ribeiro, “Ergodic stochastic optimization algorithms for wireless communication and networking,” *IEEE Trans. Signal Process.*, vol. 58, no. 12, pp. 6369–6386, Dec. 2010.

[8] Y. Hu and A. Ribeiro, “Optimal wireless networks based on local channel state information,” *IEEE Trans. Signal Process.*, Oct. 2010, submitted.