Forty-Ninth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 28 - 30, 2011

Optimal Random Access for Wireless Networks in the Presence of Fading

Yichuan Hu and Alejandro Ribeiro
Department of Electrical and Systems Engineering, University of Pennsylvania
200 South 33rd Street, Philadelphia, PA 19104
Email: {yichuan, aribeiro}@seas.upenn.edu

***Abstract*—This paper considers 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, a problem approximation is introduced. 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 probability 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. Numerical results on a randomly generated network corroborate theoretical results.** [1]

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

I. 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., [2], and protocol design, e.g., [3], 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 in the more practical situation where, due to the presence of random fading, only local CSI is available. This restriction implies that operating variables of each terminal are selected as functions of the channels linking the terminal with neighboring nodes and further leads to the selection of random access as the natural medium access choice. Indeed, 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. 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. Operating points are characterized by external arrival rates, routes, link capacities, average power consumptions, channel access decisions, and power allocations. Our goal is to select these variables to be optimal in terms of ergodic averages.

Optimal design of multihop random access networks has been considered in [4]–[6]. Assuming that capacities of links in the network are fixed and that terminals transmit with certain probabilities without coordination, these works focus on computing terminal transmission probabilities that are optimal in some sense. E.g., distributed algorithms are proposed in [4] for achieving proportionally fair utility, and in [5] for general utility functions. To reduce algorithm complexity and increase convergence speed, several enhancements are discussed in [6]. However, optimization across fading states is not considered in any of these works.

Adapting transmission decisions to random fading states has been considered in the particular case of random multiple access protocols [7]–[10]. In this case it is known that a threshold-based policy in which terminals transmit when their channels exceed a threshold and stay silent otherwise is optimal. This was originally proved for simple collision model [7], and later extended to other scenarios with different packet reception assumptions [8]–[10]. Since these works consider single hop wireless networks they do not apply directly to the multihop wireless fading networks considered here. An existing approach to optimal multihop random access is [11] where threshold-based policies are applied in multihop random access networks. Our work differs from [11] in that: (i) While routes are fixed in [11] we consider them as variables to be optimized. (ii) While terminals in [11] are assumed to have access to the channels' probability distributions, we develop online algorithms that operate without this prior knowledge.

This paper builds on recent results showing that non-convex wireless networking optimization problems have null duality gap as long as the probability distributions of underlying fading channels have no points of strictly positive measure [12]. Given this result it is possible to develop stochastic subgradient descent algorithms in the dual domain that have been proven optimal in an ergodic almost sure sense [13]. Our goal is to apply [12] and [13] to a scenario where only local CSI is available and random access is used at the physical layer. To do so we begin by introducing an optimization problem that defines the optimal random access network (Section II). Since this problem is not amenable to distributed implementation we proceed to a suboptimal approximation through a problem that while still not convex has zero duality gap [12] (Section II-B). 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 III). Results on ergodic stochastic optimization from [13] are finally leveraged to show that the proposed algorithm yields operating points that are almost surely close to optimal (Section IV). Numerical results and concluding remarks are presented in Sections V and VI.

[1]Work in this paper is supported by ARO P-57920-NS and NSF CAREER CCF-0952867. Part of the results in this paper were presented at ICASSP 2011 [1].

II. Problem Formulation

Consider an ad-hoc wireless network consisting of J terminals denoted by $\{T_j\}_{j=1}^J$. Network connectivity is modeled as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $v \in \mathcal{V} := \{1, \ldots, J\}$ representing the J terminals and edges $e = (i,j) \in \mathcal{E}$ connecting pairs of terminals that can communicate with each other. Denote the neighborhood of terminal 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 transmission. The average rate at which k-flow packets are generated at i is denoted by a_i^k. Terminal 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. To conserve flow, exogenous rates a_i^k and endogenous rates r_{ij}^k at terminal i must satisfy

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

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

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

Unlike wireline networks where c_{ij} are fixed, link capacities in wireless networks are dynamic. Let time be divided into slots indexed by t and denote the time-varying block fading channel between i and 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 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, terminal 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 $q_{ij}(t) = 1$, terminal i transmits on link (i,j) and remains silent otherwise. Further define $q_i(t) := Q_i(\mathbf{h}_i(t)) := \sum_{j \in \mathcal{N}(i)} Q_{ij}(\mathbf{h}_i(t))$ to indicate a transmission from i to any of its neighbors. We restrict i to communicate with, at most, one neighbor per time slot implying that 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 instantaneous 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 terminal 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.,

$$p_i \geq \mathbb{E}_{\mathbf{h}_i} \Bigg[\sum_{j \in \mathcal{N}(i)} P_{ij}(\mathbf{h}_i) Q_{ij}(\mathbf{h}_i) \Bigg]. \qquad (3)$$

If terminal i transmits to node 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 signal to noise ratio (SNR) $h_{ij}(t)p_{ij}(t)$ 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.

Due to contention, a transmission from i to j at time t succeeds if a collision does not occur. In turn, this happens if: (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. $q_l(t) = 0$ for all $l \in \mathcal{N}(j)$ and $l \neq i$. Recalling the definition of interference neighborhood $\mathcal{M}_i(j)$ and that if a transmission occurs its rate is $C_{ij}\big(h_{ij}(t)p_{ij}(t)\big)$ we express the instantaneous transmission rate from i to j in time slot t as $c_{ij}(t) := c_{ij}(\mathbf{h}_i(t)) = C_{ij}\big(h_{ij}(t)p_{ij}(t)\big) q_{ij}(t) \prod_{l \in \mathcal{M}_i(j)} [1 - q_l(t)]$. Assuming an ergodic mode of operation, the capacity of link $i \to j$ can then be written as

$$c_{ij} = \mathbb{E}_{\mathbf{h}} \Bigg[C_{ij}(h_{ij} P_{ij}(\mathbf{h}_i)) Q_{ij}(\mathbf{h}_i) \prod_{l \in \mathcal{M}_i(j)} [1 - Q_l(\mathbf{h}_l)] \Bigg]. \qquad (4)$$

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 $Q_{lm}(\mathbf{h}_l)$ and $P_{lm}(\mathbf{h}_l)$ for all $i \neq l$. Since $Q_l(\mathbf{h}_l) := \sum_{m \in \mathcal{N}(l)} Q_{lm}(\mathbf{h}_l(t))$ by definition, it follows that $Q_{ij}(\mathbf{h}_i)$ is also independent of $Q_l(\mathbf{h}_l)$ for all $i \neq l$. This allows us to write the expectation of the product on the right hand side of (4) as a product of expectations,

$$c_{ij} \leq \mathbb{E}_{\mathbf{h}_i} \Big[C_{ij}\big(h_{ij} P_{ij}(\mathbf{h}_i)\big) Q_{ij}(\mathbf{h}_i) \Big] \prod_{l \in \mathcal{M}_i(j)} \Big[1 - \mathbb{E}_{\mathbf{h}_l} \big[Q_l(\mathbf{h}_l) \big] \Big], \qquad (5)$$

where we also relaxed the equality constraint to an inequality, which can be done without loss of optimality.

The operating point of a wireless network is characterized by variables a_i^k, r_{ij}^k, c_{ij}, p_i and functions $P_{ij}(\mathbf{h}_i)$, $Q_{ij}(\mathbf{h}_i)$. Besides, (1)-(3) and (5) these variables are subject to certain box constraints. Admission variables, have lower and upper bounds due to application layer requirements, i.e., $a_i^{\min} \leq a_i^k \leq a_i^{\max}$. Similarly, routing variables, link capacities, and terminal power budgets cannot be negative and are also subject to given upper bounds, i.e., $0 \leq r_{ij}^k \leq r_{ij}^{\max}$, $0 \leq c_{ij} \leq c_{ij}^{\max}$, and $0 \leq p_i \leq p_i^{\max}$. Furthermore, according to definition, $P_{ij}(\mathbf{h}_i)$ and $Q_{ij}(\mathbf{h}_i)$ can only take values from $[0, p_{ij}^{\max}]$ and $\{0,1\}$, respectively. For notational simplicity, we define vectors $\mathbf{x}_i := [a_{ij}^k, r_{ij}^k, c_{ij}, p_i]$ and $\mathbf{P}_i(\mathbf{h}_i) := [P_{ij}(\mathbf{h}_i), Q_{ij}(\mathbf{h}_i)]$ to group all the variables related to terminal i and summarize these box constraints as $\{\mathbf{x}_i, \mathbf{P}_i(\mathbf{h}_i)\} \in \mathcal{B}_i$ with

$$\begin{aligned} \mathcal{B}_i := \Big\{ & \mathbf{x}_i, \mathbf{P}_i(\mathbf{h}_i) \;:\; a_i^{\min} \leq a_i^k \leq a_i^{\max},\ 0 \leq r_{ij}^k \leq r_{ij}^{\max}, \\ & 0 \leq c_{ij} \leq c_{ij}^{\max},\ 0 \leq p_i \leq p_i^{\max}, \\ & 0 \leq P_{ij}(\mathbf{h}_i) \leq p_{ij}^{\max},\ Q_{ij}(\mathbf{h}_i) \in \{0,1\},\ Q_i(\mathbf{h}_i) \in \{0,1\} \Big\}. \end{aligned} \qquad (6)$$

A. Optimal operating point

As network designers, we wish to find the optimal operating point of the wireless network defined as a set of variables a_i^k, r_{ij}^k, c_{ij}, p_i and functions $Q_{ij}(\mathbf{h}_i)$, $P_{ij}(\mathbf{h}_i)$ that satisfy constraints (1)-(3), (5), and (6) 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 terminal i and increasing convex functions $V_i(\cdot)$ typifying penalties for consuming p_i units of power at i. The optimal network based on local CSI is then defined as the solution of

$$\mathsf{P} \;=\; \max_{\{\mathbf{x}_i,\mathbf{P}_i(\mathbf{h}_i)\}\in\mathcal{B}_i} \sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i\in\mathcal{V}} V_i(p_i) \qquad (7)$$
$$\text{s.t.} \quad \text{constraints } (1),(2),(3),(5).$$

Our goal is to develop a distributed algorithm to solve (7) without accessing the channel pdf $m_{\mathbf{h}}(\cdot)$. This is challenging because: (i) The optimization space in (7) 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, (7) is a non-convex optimization problem. (iii) Constraints (3) and (5) involve expectations over channel states $\mathbf{h}$ whose pdf is unknown. (iv) The fact that the transmission rate c_{ij} is determined not only by the transmitter but also by the receiver and its neighbors [cf. (5)] hinders the development of distributed optimization algorithms.

Notice that the number of constraints in (7) is finite. This implies that while there are infinite number of variables in the primal domain, there are a finite number of variables in the dual domain. Thus, while working in the dual domain may entail some loss of optimality due the non-convex constraints in (7), it does overcome challenge (i) because the dual function is finite dimensional. It also overcomes challenge (ii) since the dual function is always convex, while challenge (iii) can be solved by using stochastic subgradient descent algorithms on the dual function; see e.g., [14] and [13]. However, working with the dual problem of (7) does not conduce to a distributed optimization algorithm due to the coupling introduced by constraint (5). This prompts the introduction of a decomposable approximation that we pursue in the next section.

B. Problem approximation

For reasons that will become clear in Section III a distributed solution of the problem in (7) is not possible because scheduling functions $Q_{ij}(\mathbf{h}_i)$ and $Q_l(\mathbf{h}_l)$ 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 $1-Q_l(\mathbf{h}_l)$ 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

$$\tilde{c}_{ij} := \log c_{ij} \le \log \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\big(h_{ij}P_{ij}(\mathbf{h}_i)\big)Q_{ij}(\mathbf{h}_i)\right] + \sum_{l\in\mathcal{M}_i(j)} \log\left[1-\mathbb{E}_{\mathbf{h}_l}[Q_l(\mathbf{h}_l)]\right], \qquad (8)$$

where we defined $\tilde{c}_{ij} := \log c_{ij}$. While scheduling functions of different terminals now appear as summands on the right hand side of (8), 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 avoid this issue we use the linear lower bound $1+\tilde{c}_{ij} \le e^{\tilde{c}_{ij}}$ and approximate this constraint as $\sum_{k\in\mathcal{K}} r_{ij}^k \le 1+\tilde{c}_{ij}$. Upon defining the average attempted transmission rate of link (i,j) as

$$x_{ij} := \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\left(h_{ij}P_{ij}\left(\mathbf{h}_i\right)\right)Q_{ij}\left(\mathbf{h}_i\right)\right], \qquad (9)$$

and the transmission probability of terminal i as

$$y_i := \mathbb{E}_{\mathbf{h}_i}\left[Q_i\left(\mathbf{h}_i\right)\right], \qquad (10)$$

the original optimization problem P is approximated by

$$\mathsf{P} \ge \tilde{\mathsf{P}} \;=\; \max_{\{\tilde{\mathbf{x}}_i,\mathbf{P}_i(\mathbf{h}_i)\}\in\mathcal{B}_i} \sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(a_i^k\right) - \sum_{i\in\mathcal{V}} V_i\left(p_i\right) \qquad (11)$$
$$\text{s.t.} \quad a_i^k \le \sum_{j\in\mathcal{N}(i)}\left(r_{ij}^k - r_{ji}^k\right), \quad \sum_{k\in\mathcal{K}} r_{ij}^k \le 1+\tilde{c}_{ij},$$
$$\tilde{c}_{ij} \le \log x_{ij} + \sum_{l\in\mathcal{M}_i(j)} \log\left(1-y_l\right),$$
$$x_{ij} \le \mathbb{E}_{\mathbf{h}_i}\left[C_{ij}\big(h_{ij}P_{ij}\left(\mathbf{h}_i\right)\big)Q_{ij}\left(\mathbf{h}_i\right)\right], \quad y_i \ge \mathbb{E}_{\mathbf{h}_i}\left[Q_i\left(\mathbf{h}_i\right)\right],$$
$$p_i \ge \mathbb{E}_{\mathbf{h}_i}\Bigg[\sum_{j\in\mathcal{N}(i)} P_{ij}\left(\mathbf{h}_i\right)Q_{ij}\left(\mathbf{h}_i\right)\Bigg],$$

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 (7) and (11) 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 (11) also satisfies the constraints in (7). In particular, the solution of (11) is feasible in (7), although possibly suboptimal. Further note that variables associated with different terminals appear as different summands of the objective and constraints in (11). This is the signature of optimization problems amenable to distributed implementations as we explain in the next section.

III. Distributed Stochastic Learning Algorithm

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 (11) 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

$$\mathcal{L}\left(\tilde{\mathbf{x}},\mathbf{P}(\mathbf{h}),\mathbf{\Lambda}\right) = \sum_{i\in\mathcal{V}} \mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i,\mathbf{\Lambda}\right) + \mathbb{E}_{\mathbf{h}_i}\Big[\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i),\mathbf{h}_i,\mathbf{\Lambda})\Big]. \qquad (12)$$

The local Lagrangian component $\mathcal{L}_i^{(1)}\big(\tilde{\mathbf{x}}_i,\mathbf{\Lambda}\big)$ in (12) is defined as

$$\mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i,\mathbf{\Lambda}\right) := \sum_k U_i^k\left(a_i^k\right) - \lambda_i^k a_i^k + \sum_{j\in\mathcal{N}_i}\left(\lambda_i^k - \lambda_j^k - \mu_{ij}\right) r_{ij}^k$$
$$+ \sum_{j\in\mathcal{N}_i}\big(\mu_{ij}-\nu_{ij}\big)\tilde{c}_{ij} + \big(\xi_i p_i - V_i(p_i)\big) + \sum_{j\in\mathcal{N}_i}\left[\nu_{ij}\log x_{ij} - \alpha_{ij}x_{ij}\right]$$
$$+ \beta_i y_i + \Bigg[\sum_{k\in\mathcal{N}(i)}\Bigg(\nu_{ki} + \sum_{l\in\mathcal{N}(k),l\neq i}\nu_{lk}\Bigg)\Bigg]\log(1-y_i). \qquad (13)$$

Transport Layer

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

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

Network Layer

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

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

Link Layer

$$\tilde{c}_{ij}(t) = \text{argmax}_{0\le \tilde{c}_{ij}\le \tilde{c}_{ij}^{\max}} \mu_{ij}(t)\tilde{c}_{ij} - \nu_{ij}(t)\tilde{c}_{ij}, \qquad c_{ij}(t) = 1 + \tilde{c}_{ij}(t)$$

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

MAC Layer

$$p_i(t) = \text{argmax}_{0\le p_i\le p_i^{\max}} \xi_i(t)p_i - V_i(p_i)$$
$$x_{ij}(t) = \text{argmax}_{x_{ij}\ge 0}\nu_{ij}(t)\log x_{ij} - \alpha_{ij}(t)x_{ij}$$
$$y_i(t) = \text{argmax}_{0\le y_i\le 1}\beta_i(t)y_i + \left[\sum\nolimits_{j\in\mathcal{N}(i)}\left[\nu_{ji}(t) + \sum\nolimits_{l\in\mathcal{M}_i(j)}\nu_{lj}(t)\right]\right]\log(1-y_i)$$

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

Physical Layer

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

Fig. 1. Layers and layer interfaces. The stochastic subgradient descent algorithm in terms of layers and layer interfaces. Layers maintain primal variables $a_i^k(t)$, $r_{ij}^k(t)$, $\tilde{c}_{ij}(t)$, $p_{ij}(t)$, $q_{ij}(t)$ as well as auxiliary variables $p_i(t)$, $x_{ij}(t)$, and $y_i(t)$ while multipliers $\lambda_i^k(t)$, $\mu_{ij}(t)$, $\nu_{ij}(t)$, $\alpha_{ij}(t)$, $\beta_i(t)$ and $\xi_i(t)$ are associated with interfaces between adjacent layers. Primal variables can be easily computed based on multipliers from interfaces to adjacent layers and dual variables are updated using information from adjacent layers.

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

$$\mathcal{L}_i^{(2)}(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}) := \sum_{j\in\mathcal{N}(i)} Q_{ij}(\mathbf{h}_i)\left[\alpha_{ij}C_{ij}\left(h_{ij}P_{ij}(\mathbf{h}_i)\right) - \beta_i - \xi_i P_{ij}(\mathbf{h}_i)\right]. \tag{14}$$

The dual function is then defined as the maximum of the Lagrangian (12) 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

$$\tilde{\mathsf{D}} = \min_{\mathbf{\Lambda}\ge 0} g(\mathbf{\Lambda}) = \min_{\mathbf{\Lambda}\ge 0} \max_{\{\tilde{\mathbf{x}}_i, \mathbf{P}_i(\mathbf{h}_i)\}\in\tilde{\mathcal{B}}_i} \mathcal{L}\left(\tilde{\mathbf{x}}, \mathbf{P}(\mathbf{h}), \mathbf{\Lambda}\right). \tag{15}$$

The separability on per-terminal terms $\mathcal{L}_i^{(1)}\big(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}\big)$ and per-terminal and per-channel elements $\mathcal{L}_i^{(2)}\big(\mathbf{P}_i(\mathbf{h}_i), \mathbf{h}_i, \mathbf{\Lambda}\big)$ is exploited in the next section to develop a distributed stochastic subgradient descent algorithm on the dual domain that solves the dual problem (15) and, indirectly, the primal problem (11).

A. 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 (12) (the operator $[\cdot]^+$ denotes projection in the positive orthant)

$$\tilde{\mathbf{x}}_i(t) = \underset{\mathbf{x}_i}{\text{argmax}}\left\{\mathcal{L}_i^{(1)}\left(\tilde{\mathbf{x}}_i, \mathbf{\Lambda}(t)\right)\right\}, \tag{16}$$

$$\mathbf{P}_i(t) = \underset{\mathbf{P}_i}{\text{argmax}}\left\{\mathcal{L}_i^{(2)}\left(\mathbf{P}_i, \mathbf{h}_i(t), \mathbf{\Lambda}(t)\right)\right\}. \tag{17}$$

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 $\mathbf{s}(t)$

$$\mathbf{\Lambda}(t+1) = \left[\mathbf{\Lambda}(t) - \epsilon\mathbf{s}(t)\right]^+. \tag{18}$$

Specific expressions for the maximizations in (16) and (17) and for $\mathbf{s}(t)$ in (18) are given in Fig. 1.

B. Network operation, layers, and layer interfaces

]To describe the role of different variables in the network's operation it is convenient to think in terms of a layered architecture with $a_i^k(t)$ associated with the transport layer, $r_{ij}^k(t)$ with the network layer, $c_{ij}(t)$ with the link layer, $x_{ij}(t)$, $y_i(t)$, and $p_i(t)$ with the medium access (MAC) layer, and $p_{ij}(t)$ and $q_{ij}(t)$ with the physical layer; see Fig. 1.

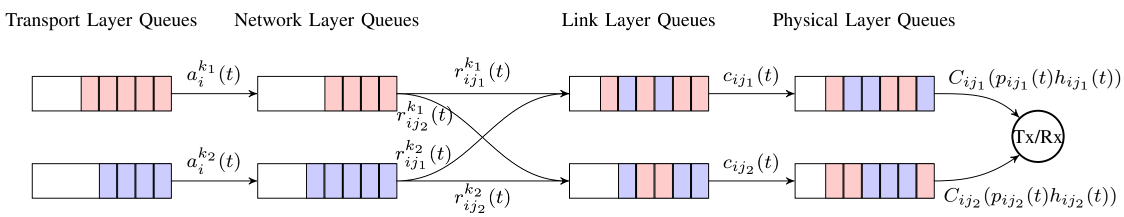


Fig. 2. Queue dynamics. Terminal i operates by controlling queues in different layers based on operating points $a_i^k(t)$, $r_{ij}^k(t)$, $c_{ij}(t)$, $p_{ij}(t)$ and $q_{ij}(t)$. In the transport layer and the network layer, each flow k has a queue. In the link layer and the physical layer, each outgoing link (i, j) maintains a queue. In this particular example, there are two flows k_1 and k_2 and there are two neighboring nodes j_1 and j_2. Packets for flow k_1 are marked red while packets for k_2 are in blue.

Variables $a_i^k(t)$, $r_{ij}^k(t)$, $c_{ij}(t)$, $p_{ij}(t)$ and $q_{ij}(t)$ determine network operation by controlling the flow of packets through queues associated with their corresponding layers; see Fig. 2. In the transport and network layers there are queues associated with each of the $|\mathcal{K}|$ flows. In the link and physical layers, queues for each of the $|\mathcal{N}(i)|$ outgoing links (i, j) are maintained. The value of $a_i^k(t)$ determines how many packets are moved from the k-flow queue in the transport layer to the k-flow queue at the network layer at time t. The number of packets transferred at time t from the k-flow network layer queue to the (i, j) queue at the link layer is determined by $r_{ij}^k(t)$. Notice that packets of a particular queue in the network layer may be distributed to different queues in the link layer. Conversely, packets in a particular queue in the link layer may come from different network layer queues, i.e., they may belong to different flows. At time t there are $c_{ij}(t)$ packets moved from the (i, j) queue at the link layer to the (i, j) queue at the physical layer.

At the physical layer queues are emptied through transmission to neighboring terminals. Resource allocation variables $q_{ij}(t)$ and $p_{ij}(t)$ determine the scheduling and transmitted power of link (i, j). If a transmission is scheduled and successful, i.e., a collision does not occur, $C_{ij}(h_{ij}(t)p_{ij}(t))$ units of information are transferred to terminal j from the (i, j) physical layer queue at terminal i. If a collision occurs, they stay at the same queue awaiting retransmission in a future time slot. When a packet is successfully decoded by terminal j it determines which flow they belong to and what destination they are heading for. If the terminal happens to be the destination, packets are forwarded to the application layer. If the terminal is not the designated destination, packets are put into a network layer queue according to their flow identifications.

Besides administering queues, layers are also responsible for updating the values of their corresponding primal variables according to (16)-(17); see Fig. 1. The transport layer updates $a_i^k(t)$, the network layer keeps track of $r_{ij}^k(t)$, while the link layer computes $\tilde{c}_{ij}(t)$ and $c_{ij}(t)$. The MAC layer updates $p_i(t)$, $x_{ij}(t)$ and $y_i(t)$, while the physical layer determines $p_{ij}(t)$ and $q_{ij}(t)$.

Computation of these primal per layer updates necessitates access to Lagrange multipliers motivating the introduction of layer interfaces to maintain and update their values. E.g., since $\lambda_{ij}^k(t)$ is associated with the flow conservation constraint that relates transport variables $a_i^k(t)$ and network variables $r_{ij}^k(t)$ it provides a natural interface between the transport and network layers. Thus, we introduce a transport-network interface tasked with computing the dual stochastic subgradient component and executing the update for $\lambda_i^k(t)$. Similarly, a network-link interface is introduced to keep track of multipliers $\mu_{ij}(t)$ and execute the corresponding update. A link-MAC interface does the proper for multipliers $\nu_{ij}(t)$. Likewise, the remaining multipliers $\alpha_{ij}(t)$, $\beta_i(t)$ and $\xi_i(t)$ provide a MAC-physical interface. Observe that primal variables are updated with information available at adjacent interfaces, while dual variable updates are undertaken with information available at adjacent layers. Their definition is thereby justified, because information is exchanged only between adjacent layers and interfaces.

We remark that MAC layer variables $x_{ij}(t)$, $y_i(t)$, and $p_i(t)$ do not affect network operation, i.e., queue dynamics, at time t. The role of these variables is to record average behaviors of the terminal to affect determination of $c_{ij}(t)$, $p_{ij}(t)$, and $q_{ij}(t)$ in subsequent time slots. This role is consistent with the definitions of p_i as the the average transmitted power [cf. (3)], x_{ij} as the average attempted transmission rate [cf. (9)], and y_i as the (average) transmission probability [cf. (10)].

C. Message passing

Most primal and dual variable updates in Fig. 1 can be done locally at terminal i. E.g., the physical layer update at terminal i requires access to multipliers $\alpha_{ij}(t)$, $\beta_i(t)$, and $\xi_i(t)$ which are available at the physical-MAC interface of terminal i. The updates for primal variables $r_{ij}^k(t)$ and $y_i(t)$, as well as duals $\lambda_{ij}^k(t)$ and $\nu_{ij}(t)$, however, necessitate access to variables of other terminals. The update of multiplier $\lambda_i^k(t)$ at the network-transport interface depends on network variables $r_{ij}^k(t)$ and $a_i^k(t)$ which are available at terminal i, but also on the variable $r_{ji}^k(t)$ available at (neighboring) terminal j. Similarly, the $r_{ij}^k(t)$ update at the network layer depends on locally available multipliers $\lambda_i^k(t)$ and $\mu_{ij}(t)$, but also on the neighboring multiplier $\lambda_j^k(t)$. The update of multiplier $\nu_{ij}(t)$ is somewhat more complex as it depends on local variables $x_{ij}(t)$ and $\tilde{c}_{ij}(t)$, 1-hop neighborhood variables $y_j(t)$, and 2-hop neighborhood variables $y_l(t)$ for all $l \in \mathcal{N}(j)$. Likewise, the update for $y_i(t)$ at the MAC layer depends on local dual variables $\beta_i(t)$, 1-hop neighborhood variables $\nu_{ji}(t)$ for all $j \in \mathcal{N}(i)$, and 2-hop neighboring variables $\nu_{lj}(t)$ for all $l \in \mathcal{N}(j)$ in the neighborhood of j for some $j \in \mathcal{N}(i)$ in the neighborhood of i. Therefore, implementation of these four updates requires sharing appropriate variables with 1-hop and 2-hop neighbors.

Given that these four updates depend on quantities available at 1-hop and 2-hop neighbors it is necessary to devise a message passing mechanism among terminals to share the necessary values. At the beginning of primal iteration, terminal i transmits $\lambda_i^k(t)$ and $\nu_{ij}(t)$ to all its neighbors $j \in \mathcal{N}(i)$. As a result, terminal i receives multipliers $\lambda_j^k(t)$ and $\nu_{ji}(t)$ from all of their neighbors $j \in \mathcal{N}(i)$.

Terminal i follows by computing and broadcasting the term $\sum_{l\in\mathcal{N}(i)} \nu_{li}(t)$ to all its neighbors $j \in \mathcal{N}(i)$. Upon receiving this information, terminal j subtracts $\nu_{ji}(t)$ from the received value to evaluate the expression $\sum_{l\in\mathcal{N}(i),l\neq j} \nu_{li}(t)$. The terms required for computing primal variables $r_{ij}^k(t)$ and $y_i(t)$ are now available at i. Since the variables necessary for the remaining primal updates are locally accessible the primal iterations associated with all the layers in Fig.1 are performed at each terminal.

After completing the layer updates, primal iterates $r_{ij}^k(t)$ and $y_i(t)$ need to be exchanged between neighbors to perform the dual updates associated with the layer interfaces in Fig.1. Terminal i starts passing variables $y_i(t)$ and $r_{ij}^k(t)$ to all its neighbors. Having received $y_j(t)$ from all $j \in \mathcal{N}(i)$ terminal i computes and broadcasts the sum $\sum_{l\in\mathcal{N}(i)} y_l(t)$ to all its neighbors. With this information in hand terminal j adds $y_i(t)$ and subtracts $y_j(t)$ from this value to evaluate $\sum_{l\in\mathcal{M}_j(i)} y_l(t) = \sum_{l\in\mathcal{N}(i)} y_l(t) + y_i(t) - y_j(t)$. Quantities necessary to update $\lambda_i^k(t)$ and $\nu_{ij}(t)$ are now available along with the terms necessary for the remaining dual updates that were locally available. The dual updates associated with the layer interfaces in Fig.1 are now performed and we proceed to the next primal iteration.

We remark that $r_{ij}^k(t)$ and $\lambda_i^k(t)$ are transmitted to 1-hop neighbors, whereas $y_i(t)$ and $\nu_{ij}(t)$ are sent to 2-hop neighbors. This latter fact holds because transmissions of a given terminal can interfere with neighbors two hops away from her.

D. Successive convex approximation

As mentioned in the problem reformulation in Sec. II-B, we a use linear lower bound to approximate the capacity constraint. In general, we can use a concave function $f_{ij}(\tilde{c}_{ij})$ which is smaller than $e^{\tilde{c}_{ij}}$ to approximate $e^{\tilde{c}_{ij}}$. As a result, instead of directly computing link capacity variable $c_{ij}(t)$, an approximated version $\tilde{c}_{ij}(t)$ is calculated in the primal iteration. In the network operation, the link capacity $c_{ij}(t) = f_{ij}\tilde{c}_{ij}(t)$ is used in the link layer. While this approximation convexifies the capacity constraint and provides a feasible solution to the original problem, it reduces the size of the feasible set of primal variables. This implies that this obtained link capacity $c_{ij}(t)$ may not be optimal to the original problem. To reduce its impact on optimality, we use different $f_{ij}(\tilde{c}_{ij})$ at different time slots and hope the approximations become better as time grows. Define then $\bar{\tilde{c}}_{ij}(t) := 1/t\sum_{u=1}^{t} \tilde{c}_{ij}(u)$ and lower bound $e^{\tilde{c}_{ij}(t+1)}$ with the first order approximation

$$e^{\tilde{c}_{ij}(t+1)} \geq e^{\bar{\tilde{c}}_{ij}(t)}\tilde{c}_{ij}(t+1) + e^{\bar{\tilde{c}}_{ij}(t)}\left[1 - \bar{\tilde{c}}_{ij}(t)\right]. \quad (19)$$

Notice that the right hand side of (19) is a linear function of $\tilde{c}_{ij}(t+1)$ and thus concave. We can then choose $f_{ij}^{(t+1)}(\tilde{c}_{ij}) = e^{\bar{\tilde{c}}_{ij}(t)}\tilde{c}_{ij} + e^{\bar{\tilde{c}}_{ij}(t)}\left[1 - \bar{\tilde{c}}_{ij}(t)\right]$ to approximate $e^{\tilde{c}_{ij}}$ at time slot $t+1$.

IV. Feasibility and Optimality

Solving the optimization problem in (7) entails finding optimal variables $\mathbf{x}_i^*$, and power allocations $\mathbf{P}_i^*(\mathbf{h}_i)$ that satisfy problem constraints and offer optimal yield P. This would require knowledge of the channels' probability distributions and a joint optimization among terminals. To overcome these restrictions and develop an adaptive distributed solution, we reformulated the problem as in (11) entailing a performance degradation to $\tilde{\mathsf{P}} \leq \mathsf{P}$. This reformulation permits introduction of the dual stochastic subgradient descent algorithm, defined by recursive application of (16)-(18), that produces a sequence of network operating points $\mathbf{x}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ – as well as sequences of auxiliary variables $x_{ij}(\mathbb{N})$ and $y_i(\mathbb{N})$ – which given results in [13] are expected to be almost surely feasible and give a utility yield close to $\tilde{\mathsf{P}}$ in an ergodic sense. Notice however, that since (16)-(18) descends on the dual function of the reformulated problem, feasibility holds with respect to the constraints in (11). Our main intent here is to show that sequences of operating points $\mathbf{x}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ generated by (16)-(18) are also feasible for the optimization problem in (7). Specifically, our goal is to prove the following theorem.

Theorem 1 *Consider a wireless network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ using random access at the physical layer so that ergodic link capacities are as given in* (5). *Let $a_i^k(\mathbb{N})$, $r_{ij}^k(\mathbb{N})$, $c_{ij}(\mathbb{N})$, $p_i(\mathbb{N})$, $q_{ij}(\mathbb{N})$ and $p_{ij}(\mathbb{N})$ be sequences of network operating points generated by the stochastic descent algorithm in* (16)-(18) *and denote as $\bar{a}_i^k$, $\bar{r}_{ij}^k$, $\bar{c}_{ij}$, and $\bar{p}_i$ the corresponding ergodic limits of $a_i^k(\mathbb{N})$, $r_{ij}^k(\mathbb{N})$, $c_{ij}(\mathbb{N})$, and $p_i(\mathbb{N})$. Assume the following hypotheses: (h1) The second moment of the norm of the stochastic subgradient $\mathbf{s}(t)$ is finite, i.e., $\mathbb{E}_{\mathbf{h}}\left[\|\mathbf{s}(t)\|^2 \middle| \mathbf{\Lambda}(t)\right] \leq \hat{S}^2$. (h2) There exists a set of strictly feasible primal variables that satisfy the constraints of the reformulated optimization problem in* (11) *with strict inequality. (h3) The dual function $g(\mathbf{\Lambda})$ of the reformulated problem as defined in* (15) *has a unique minimizer $\mathbf{\Lambda}^*$. It then holds:*

(i) Near feasibility of physical layer constraints. *There exists a function $M(\epsilon)$ with $\lim_{\epsilon\to 0} M(\epsilon) = 0$ such that the average transmission rate constraint in* (5) *is almost surely satisfied with feasibility gap smaller than $M(\epsilon)$ in an ergodic sense, i.e.,*

$$\bar{c}_{ij} \leq \lim_{t\to\infty} \frac{1}{t}\sum_{u=1}^{t}\left[C_{ij}(h_{ij}(u)p_{ij}(u))q_{ij}(u)\prod_{k\in\mathcal{M}_i(j)}[1 - q_k(u)]\right] + M(\epsilon), \quad a.s. \quad (20)$$

(ii) Feasibility of upper layer constraints. *The flow conservation constraint in* (1), *the link capacity constraint in* (2) *and the average power constraint in* (3) *are almost surely satisfied in an ergodic sense, i.e.,*

$$\bar{a}_i^k \leq \sum_{j\in\mathcal{N}(i)}\left[\bar{r}_{ij}^k - \bar{r}_{ji}^k\right], \qquad \sum_{k\in\mathcal{K}} \bar{r}_{ij}^k \leq \bar{c}_{ij}, \qquad a.s., \quad (21)$$

$$\bar{p}_i \geq \lim_{t\to\infty}\frac{1}{t}\sum_{u=1}^{t}\sum_{j\in\mathcal{N}(i)} p_{ij}(u)q_{ij}(u), \qquad a.s. \quad (22)$$

(iii) Utility yield. *The utility yield of the ergodic averages of sequences $a_i^k(\mathbb{N})$ and $p_i(\mathbb{N})$ converges to a value within $\epsilon\hat{S}^2/2$ of $\tilde{\mathsf{P}}$, i.e.,*

$$\tilde{\mathsf{P}} - \left[\sum_{i\in\mathcal{V},k\in\mathcal{K}} U_i^k\left(\bar{a}_i^k\right) - \sum_{i\in\mathcal{V}} V_i\left(\bar{p}_i\right)\right] \leq \frac{\epsilon\hat{S}^2}{2}, \quad a.s. \quad (23)$$

Proof: See [15].

The feasibility results in (21) for the flow conservation and rate constraints are identical to (1) and (2). As such they imply that the ergodic limits $\bar{a}_i^k$, $\bar{r}_{ij}^k$, $\bar{c}_{ij}$ obtained from recursive application of (16)-(18) satisfy these constraints with probability 1. Notice that these limits may be different for different realizations of the

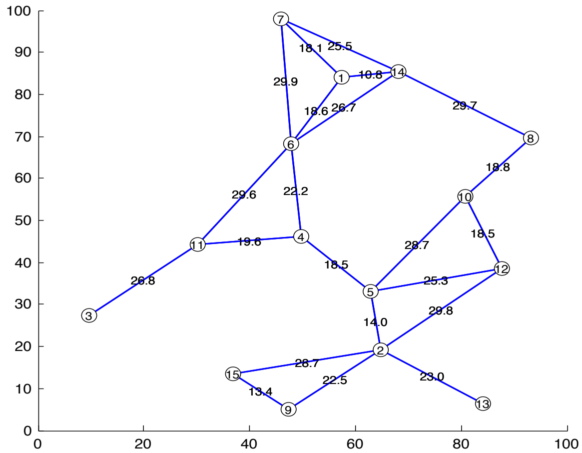


Fig. 3. Connectivity graph of a network used in simulation. The numbers on each edge shows the distance (in meters) between two communicating terminals.

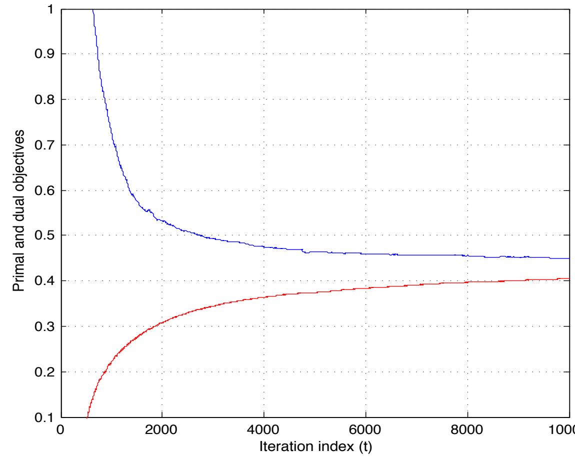


Fig. 4. Optimality of the proposed algorithm. As time grows, primal and dual objectives approach each other.

algorithm's run. Nonetheless, constraints (1) and (2) are satisfied for almost all runs. The feasibility result in (20) for the link capacity constraint, however, is not identical to (5). The difference is not only the presence of the $M(\epsilon)$ feasibility gap, but the fact that (5) involves an expectation over channel realizations whereas (20) does not. In fact, asides from the $M(\epsilon)$ constant, (20) is stronger than (5). The feasibility result in (20) states that even though sequences $\mathbf{x}_i(\mathbb{N})$ and $\mathbf{P}_i(\mathbb{N})$ may not be ergodic, the possibly different ergodic limits in the right and left hand sides of (20) satisfy the stated inequality. This implies that operating the network using variables $\mathbf{x}_i(t)$ and $\mathbf{P}_i(t)$ as generated by (16)-(18) results in long-term feasibility in that all packets are (almost surely) delivered to their corresponding destinations. Further notice that the power feasibility result in (22) is not identical to the corresponding power constraint in (3) because (3) involves an expected value whereas (22) does not. The same comments stated for the comparison of (20) and (5) extend naturally.

The utility yield result in (23) states that the long term performance of the network, as determined by average end-to-end rates $\bar{a}_i^k$ and powers $\bar{p}_i$, is close to the optimal yield $\tilde{\mathsf{P}}$ of the reformulated problem. The gap between $\tilde{\mathsf{P}}$ and the attained yield can be controlled by reducing ϵ. Notice that reducing the step size ϵ also reduces the feasibility gap $M(\epsilon)$ in (20). We also remark that the use of constant step sizes ϵ endows the algorithm with adaptability to time-varying channel distributions. This is important in practice because wireless channels are non-stationary due to user mobility and environmental dynamics.

V. Numerical Results

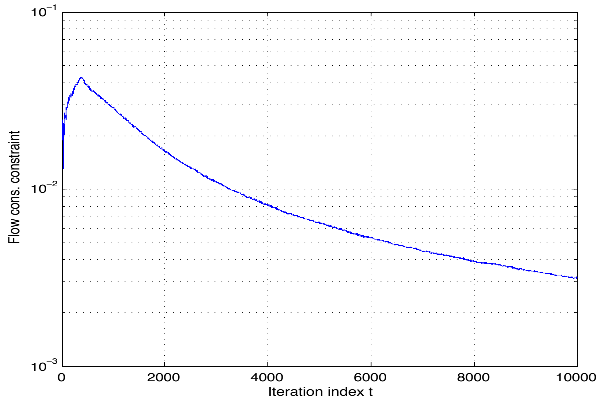
We illustrate performance of the proposed algorithm by implementing and simulating it over a network with $n = 15$ terminals randomly placed in a square with side $L = 100$ meters. Terminals can communicate with neighbors whose distances are within 30 meters. Numerical experiments here utilize the realization of this random placement shown in Fig. 3. Channel gains $h_{ij}(t)$ are Rayleigh distributed with mean $\bar{h}_{ij}$ and are independent across links and time. The average channel gain $\bar{h}_{ij} := \mathbb{E}[h_{ij}]$ follows an exponential pathloss law, $\bar{h}_{ij} = \alpha d_{ij}^{-\beta}$ with d_{ij} denoting the distance in meters between T_i and T_j and constants $\alpha = 10^{-1}\text{m}^{-1}$ and $\beta = 2.5$. Assume the use of capacity achieving codes so that the instantaneous transmission rate takes the form

$$C_{ij}(h_{ij}(t)p_{ij}(t)) = \log\left(1 + \frac{h_{ij}(t)p_{ij}(t)}{N_0}\right), \tag{24}$$

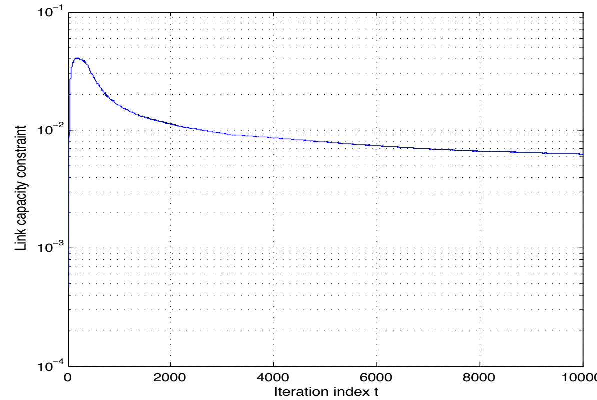
where N_0 is the channel noise set to $N_0 = 10^{-4}$ for all links. Fading channels are generated as i.i.d. There are two flows supported by the network, one from T_1 to T_2 and the other from T_8 to T_{11}. For each flow the minimum and maximum amount of information to be delivered are constrained by $a_i^{\min} = 0.1$ bits/s/Hz and $a_i^{\max} = 1$ bits/s/Hz for all nodes i. The routing and link capacity variables are bounded by $r_{ij}^{\min} = c_{ij}^{\min} = 0$ bits/s/Hz and $r_{ij}^{\max} = c_{ij}^{\max} = 1$ bits/s/Hz. The maximum average power consumption per terminal and maximum instantaneous power consumption per terminal are set to 2, i.e., $p_i^{\max} = p_{ij}^{\max} = 2$. Our objective is to maximize total amount of information delivered by the network, i.e., $U_i^k(a_i^k) = a_i^k$ and $V_i(p_i) = 0$. We set $\epsilon = 0.02$ and the simulation is conducted for 10^4 time slots. Successive convex approximation is used.

To show optimality of the algorithm we compare ergodic primal and dual objectives, as shown in Fig. 4. As time grows, the convergence of the proposed algorithm is observed as the primal and dual values approach each other. By Theorem 1, the algorithm is almost surely near optimal in the sense that the ergodic average of the utility almost surely converges to a value with optimality gap smaller than $\epsilon\hat{S}^2/2$ with respect to the optimal objective. Indeed, this is true as shown in Fig. 4 that the gap between primal and dual values becomes a small constant (about 0.05) as t increases.

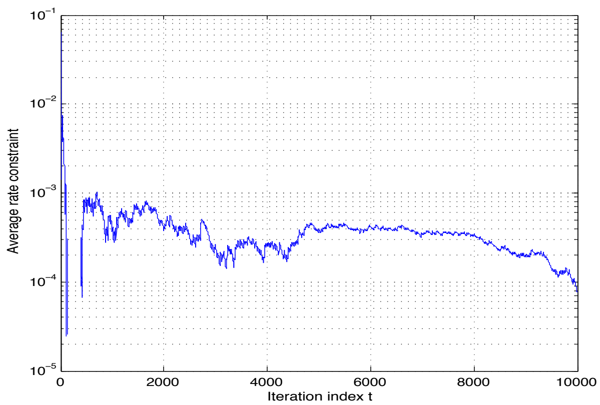
Fig. 5 shows feasibility of the proposed algorithm in terms of constraint violations. Specifically, we compute average violations of the flow conservation, link capacity, average rate and average power constraints, respectively. If these values are nonnegative, it means the corresponding constraints are satisfied in an average sense. As we can see, after about 500 steps all constraints are satisfied within 10^{-2} tolerance. The average rate constraint takes the longest time to be satisfied (see Fig. 5 (c)). This is because the transmission rate on link $T_i \to T_j$ depends not only on schedules and powers of T_i but also on those of T_j and its neighbors. This requires information to be received from, and propagated to, 2-hop networks.



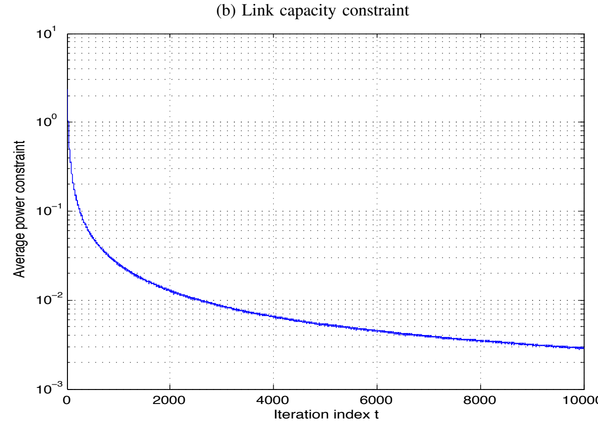
(a) Flow conservation constraint



(b) Link capacity constraint



(c) Average rate constraint



(d) Average power constraint

Fig. 5. Feasibility of the proposed algorithm.

VI. Conclusions

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. The solution exhibits a layered architecture in which variables in each layer are computed using information from interfaces to adjacent layers. The algorithm is fully distributed in that all operations necessary to achieve optimal operation are based on local information and information exchanges between neighbors.

References

[1] Y. Hu and A. Ribeiro, "Optimal wireless networks based on local channel state information," in *Proc. ICASSP'11*, Prague, Czech Republic, May 2011.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] A. H. Mohsenian-Rad, J. Huang, M. Chiang, and V. W. S. Wong, "Utility-optimal random access: Reduced complexity, fast convergence, and robust performance," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 898–911, Feb. 2009.

[7] Y. Yu and G. B. Giannakis, "Opportunistic medium access for wireless networking adapted to decentralized csi," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1445–1455, Jun. 2006.

[8] S. Adireddy and L. Tong, "Expoiting decentralized channel state information for random access," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 537–561, Feb. 2005.

[9] 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.

[10] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: an optimal stopping approach," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 205–222, Jan. 2009.

[11] G. Miao, G. Y. Li, and A. Swami, "Decentralized optimization for multichannel random access," *IEEE Trans. Commun.*, vol. 57, no. 10, pp. 3012–3023, Oct. 2009.

[12] A. Ribeiro and G. B. Giannakis, "Separation principles of wireless networking," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4488 – 4505, Sep. 2010.

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

[14] Y. Hu and A. Ribeiro, "Adaptive distributed algorithms for optimal random access channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2703–2715, Aug. 2011.

[15] ——, "Optimal wireless networks based on local channel state information," *IEEE Trans. Signal Process.*, Oct. 2010, submitted. [Online]. Available: http://www.seas.upenn.edu/ỹichuan/preprint/ranet_main.pdf