Distributed Approaches for Proportional and Max-Min Fairness in Random Access Ad Hoc Networks

Xin Wang, Koushik Kar
Department of Electrical, Computer and Systems Engineering,
Rensselaer Polytechnic Institute,
110 8th Street, Troy, NY, 12180
wangx5@rpi.edu, koushik@ecse.rpi.edu

***Abstract*— We consider the question of distributed rate optimization in a wireless random access network, with the goal of achieving global fairness objectives. Since the wireless medium is broadcast in nature, a MAC (Medium Access Control) protocol is essential to reduce collisions and ensure high system throughput. For random access MAC protocols, the feasible rate region is typically a complex, non-convex and non-separable function of the MAC protocol parameters, like the transmission probabilities or back-off window sizes. These factors make the question of distributed rate optimization in a wireless random access network a complex one, and imply that cross-layer coordination is necessary to attain end-to-end fairness.**

Despite its complexity, significant progress has been made in recent years towards addressing this question for certain fairness measures. In this paper, we summarize our recent results in designing distributed approaches in a random access ad hoc network, with the goal of achieving proportional and max-min fairness at the link layer, and for the end-to-end sessions. We also pose some open questions on this topic that are currently being investigated.

I. Introduction

Channel bandwidth is a scarce resource in a wireless ad hoc network. The wireless medium is broadcast in nature, and Medium Access Control (MAC) is essential for efficient use of the network. A MAC protocol defines rules by which nodes regulate their transmission onto the shared broadcast channel. The objective is to reduce collisions, to ensure high system throughput, and to distribute the available bandwidth fairly among the competing nodes.

Fairness is a key consideration in designing MAC protocols. One of the most widely used fairness notions is the *max-min* fairness metric[1]. The objective, stated simply, is to maximize the minimum rate allocation over all links or end-users. However, the appropriateness of max-min fairness as a bandwidth-sharing objective has been questioned recently, and an alternative fairness measure, *proportional* fairness, has been proposed [3]. Rate allocations are proportionally fair if they maximize sum of the logarithmic values of the rate allocations over all links or end-users. This objective may be viewed as that of maximizing the overall utility of rate allocations, where the utility functions are logarithmic in nature.

Rate optimization in wireline networks has been extensively researched in recent years [4], [5]. In wireline networks, since the feasible rate region can be represented by a set of simple, separable, convex constraints, globally fair rates are attainable via distributed approaches. In contrast, the link rate or link capacity in a wireless ad hoc network is not a fixed quantity, and its value depends upon the MAC protocols and parameters used. In random access MAC protocols, the feasible rate region is typically a complex, non-convex and non-separable function of MAC protocol parameters, such as transmission probabilities or back-off window sizes. This implies that for random access wireless networks, developing distributed approaches that attain globally fair rates is in general a difficult problem.

Despite the complexity of the problem, significant progress has been made in recent years towards solving this question for certain important special cases. The goal of this paper is to survey our recent results on this topic, and outline some open issues. The paper is organized as follows. Section 2 describes the system model for an ad hoc network with random access. Section 3 discusses approaches for attaining proportional fairness at the link layer, as well as for end-to-end sessions. In Section 4, we show how max-min fairness can be provided at the link layer. All algorithms in this paper can be implemented in a distributed manner, and converge to the globally optimal solution. Section 5 concludes the paper with a discussion of some open questions on this topic.

II. System Model

A wireless network can be modeled as an undirected graph $G = (N, E)$, where N and E respectively denote the set of *nodes* and the set of undirected *edges*. An edge exists between two nodes if and only if they can receive each other's signals (we assume a symmetric hearing matrix). Note that there are $2|E|$ possible communication pairs, but only a subset of these may be actively communicating. The set of active communication pairs is represented by the set of *links*, L. Each link $(i, j) \in L$ is always backlogged. Without loss of generality we assume that all the nodes share a single wireless channel of unit capacity.

For any node i, the set of i's *neighbors*, $K_i = \{j : (i, j) \in L\}$, represents the set of nodes that can receive i's signals. For any node i, the set of *out-neighbors* of i, $O_i = \{j : (i, j) \in L\} \subseteq K_i$, represents the set of neighbors to which i is sending traffic. Also, for any node i, the set of *in-neighbors* of i, $I_i = \{j : (j, i) \in L\} \subseteq K_i$, represents the set of

This work was supported by the National Science Foundation under grants ECS-0330203 and CNS-0448316.

neighbors from which i is receiving traffic. A transmission from node i reaches all of i's neighbors. Each node has a single transceiver. Thus, a node can not transmit and receive simultaneously. We do not assume any capture, i.e., node j can not receive any packet successfully if more than one of its neighbors are transmitting simultaneously. Therefore, a transmission in link $(i,j) \in L$ is successful if and only if no node in $K_j \cup \{j\} \setminus \{i\}$, transmits during the transmission on (i,j).

We focus on random access wireless networks, and use the slotted Aloha model [1] for modeling interference and throughput. In this model, i transmits a packet with probability P_i in a slot. If i does not have an outgoing edge, i.e., $O_i = \phi$, then $P_i = 0$. Once i decides to transmit in a slot, it selects a destination $j \in O_i$ with probability p_{ij}/P_i, where $\sum_{j \in O_i} p_{ij} = P_i$. Therefore, in each slot, a packet is transmitted on link (i,j) with probability p_{ij}. Let $\mathbf{p} = (p_{ij}, (i,j) \in L)$ be the vector of transmission probabilities on all edges, and let $\mathbf{P}_f$ denote the feasible region for $\mathbf{p}$, i.e. $\mathbf{P}_f = \{\mathbf{p} : 0 \le p_{ij} \le 1, \forall (i,j) \in L, P_i = \sum_{j \in O_i} p_{ij}, 0 \le P_i \le 1, \forall i \in N\}$. Then, the rate or throughput on link (i,j), x_{ij}, is given by

$$x_{ij}(\mathbf{p}) = p_{ij}(1-P_j) \prod_{k \in K_j \setminus \{i\}} (1-P_k), \qquad \mathbf{p} \in \mathbf{P}_f \ . \tag{1}$$

Note that $(1-P_j)\prod_{k \in K_j \setminus \{i\}}(1-P_k)$ in (1) is the probability that a packet transmitted on link (i,j) is successfully received at j.

III. Proportional Fairness

A. Proportional Fairness at the Link Layer

The problem of attaining proportional fairness for link rates is to compute the link attempt probabilities so as to maximize the sum of the logarithmic utilities of all link rates, i.e.,

$$\mathbf{p}^* = \arg\max_{\mathbf{p} \in \mathbf{P}_f} \sum_{(i,j) \in L} \log\left(x_{ij}(\mathbf{p})\right), \tag{2}$$

where $x_{ij}(\mathbf{p})$ is given in (1).

The optimal attempt probability in any link $(i,j) \in L$, as defined in (2), is given by (details in [6])

$$p_{ij}^* = \frac{1}{|I_i| + \sum_{k \in K_i} |I_k|}. \tag{3}$$

Note that the p_{ij}^*, as given above, satisfies the constraints $0 \le p_{ij}^* \le 1$ and $0 \le P_i^* \le 1$. Also note that a node can compute its optimal attempt probabilities if it knows the number of its in-neighbors and the number of its neighbors' in-neighbors. A node can determine the latter as follows. When the network is formed, or when the network topology changes due to the joining, leaving or movement of nodes, each node broadcasts the number of its in-neighbors to all node in its transmission range. Therefore, this algorithm can be implemented in a distributed manner, with only a small amount of local information exchange.

B. Proportional Fairness for End-to-End Sessions

We now consider providing proportional fairness for end-to-end sessions in a multi-hop Aloha network. The problem can be formulated as a nonlinear optimization problem, which maximizes the utility of all sessions rates while the link capacity constraints are satisfied. The rate optimization problem can be posed as follows:

$$\begin{array}{llll} \mathbf{P}: & \max & \sum_{s \in S} \log(y_s), & \\ & \text{s.t.} & \sum_{s \in S(i,j)} y_s \le x_{ij}(\mathbf{p}) & \forall (i,j) \in L, \\ & & y_s \ge 0 & \forall s \in S, \\ & & \mathbf{p} \in \mathbf{P}_f. & \end{array} \tag{4}$$

where y_s is the rate for session $s \in S$. The first set of constraints are the link capacity constraints. Here, $S(i,j)$ denotes the set of sessions that use link (i,j), and $x_{ij}(\mathbf{p})$ represents the capacity (rate) of link (i,j), and is given by (1). The second set of constraints ensure that the session rates are non-negative, and the third ensures that $\mathbf{p}$ is feasible.

Note that the rate optimization problem couples link attempt probabilities at the link layer with end-to-end session rates at the transport layer, and is not a convex program. It is worth noting here that our end-to-end fair rate allocation question is closely related to the cross-layer optimization problem considered in [8], in which the session rates (transport layer) must be jointly optimized with the transmission powers (physical layer), to attain end-to-end fairness goals.

1) Dual-Based Algorithm: Instead of solving the problem $\mathbf{P}$ directly, we consider the version of the end-to-end proportionally fair rate optimization question where each link capacity (rate) is parameterized:

$$\begin{array}{llll} \hat{\mathbf{P}}: & \max & \sum_s \log(y_s), & \\ & \text{s.t.} & \sum_{s \in S(i,j)} y_s \le \tilde{x}_{ij} & \forall (i,j) \in L, \\ & & y_s \ge 0 & \forall s \in S. \end{array} \tag{5}$$

In the above formulation, $\tilde{x}_{ij}$, the rate on link (i,j), is assumed to be a given constant, while y_s, representing the end-to-end session rates, are variables whose values need to be determined optimally.

Note that the optimum value of $\hat{\mathbf{P}}$ is a function on $\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}}$ is the vector of all link rates in the network, i.e. $\tilde{\mathbf{x}} = \{\tilde{x}_{ij} : (i,j) \in L\}$. We define $\hat{U}(\tilde{\mathbf{x}})$ as the optimum value in $\hat{\mathbf{P}}$, i.e.,

$$\hat{U}(\tilde{\mathbf{x}}) = \max\left\{ \sum_s \log(y_s) \,\middle|\, \sum_{s \in S(i,j)} y_s \le \tilde{x}_{ij}, (i,j) \in L \right\}. \tag{6}$$

Since link rates in $\mathbf{P}$ in turn is a function of link attempt probabilities, we define $\tilde{U}(\mathbf{p}) = \hat{U}(\mathbf{x}(\mathbf{p}))$, where $\mathbf{x}(\mathbf{p}) = (x_{ij}(\mathbf{p}) : (i,j) \in L)$. Therefore $\mathbf{P}$ can be rewritten as

$$\begin{array}{lll} \tilde{\mathbf{P}}: & \max & \tilde{U}(\mathbf{p}), \\ & \text{s.t.} & \mathbf{p} \in \mathbf{P}_f. \end{array} \tag{7}$$

We solve (7) by updating $\mathbf{p}$ as below

$$p_{ij}^{(n+1)} = p_{ij}^{(n)} + \alpha \sum_{(s,t) \in L} \lambda_{st}^{*(n)} \frac{\partial x_{st}}{\partial p_{ij}}(\mathbf{p}^{(n)}), \tag{8}$$

where α is the step size, $\frac{\partial x_{st}}{\partial p_{ij}}$ is computed using the following formula

$$\frac{\partial x_{st}}{\partial p_{ij}} = \begin{cases} (1-P_t)\prod_{k\in K_t\setminus\{s\}}(1-P_k) & \text{if } t=j, s=i, \\ -p_{st}\prod_{k\in K_t\setminus\{s\}}(1-P_k) & \text{if } t=i, s\in I_t, \\ -p_{st}(1-P_t)\prod_{k\in K_t\setminus\{s\}}(1-P_k) & \text{if } t\in K_i, s\in I_t\setminus\{i\}, \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

and $\lambda_{ij}^{*(n)}$ is the optimum solution to the dual problem of $\hat{\mathbf{P}}$ when $\tilde{\mathbf{x}} = \mathbf{x}(\mathbf{p}^{(n)})$, i.e.,

$$\lambda^{(n)} = \arg\min_{\lambda\geq\mathbf{0}}\max_{\mathbf{y}} \mathscr{L}^{(n)}(\mathbf{y},\lambda). \tag{10}$$

In (10), $\lambda = (\lambda_{ij} : (i,j)\in L)$ is the vector of Lagrange multipliers for the capacity constraints on the wireless links, $\mathbf{y} = (y_s : s\in S)$ is the vector of the end-to-end session rates, and $\mathscr{L}^{(n)}(\mathbf{y},\lambda)$ is the Lagrange function of $\hat{\mathbf{P}}$ when $\tilde{\mathbf{x}} = \mathbf{x}(\mathbf{p}^{(n)})$. Note that $\mathscr{L}^{(n)}(\mathbf{y},\lambda)$ is given by

$$\mathscr{L}^{(n)}(\mathbf{y},\lambda) = \sum_{s\in S}\log(y_s) - \sum_{(i,j)\in L}\lambda_{ij}\left(\sum_{s\in S(i,j)} y_s - x_{ij}^{(n)}\right). \tag{11}$$

We then solve $\mathbf{y}^{(n)}$ from $\hat{\mathbf{P}}$ when $\tilde{\mathbf{x}} = \mathbf{x}(\mathbf{p}^{(n)})$, i.e.,

$$\mathbf{y}^{(n)} = \arg\max_{\mathbf{y}\geq\mathbf{0}}\left\{\sum_{s\in S}\log(y_s)\,\middle|\,\sum_{s\in S(i,j)} y_s \leq x_{ij}(\mathbf{p}^{(n)}), (i,j)\in L\right\}. \tag{12}$$

Intuitively, the procedure described above adjusts link attempt probabilities in the gradient direction, and then computes the optimal session rates at the given link capacities in an iterative manner. Then the link attempt probabilities are adjusted according; as a result, link capacities change in such a way that the aggregate utility of the end-to-end sessions can be increased further. If we let $\{\mathbf{p}^{(n)}(\alpha),\mathbf{y}^{(n)}(\alpha)\}$ denote the sequence of vectors of link attempt probabilities and session rates by following the iterative procedures in (8)-(12) where the step size is α, we can show that the limit point of $\{\mathbf{p}^{(n)}(\alpha),\mathbf{y}^{(n)}(\alpha)\}$ is the globally optimal solution to the problem $\mathbf{P}$, when α is sufficiently small (proof in [11]).

2) Primal-Based Algorithm: Although (4) appears to be a non-convex problem, it can be shown to be equivalent to the convex program stated below (see proof in [11]):

$$\begin{array}{rl} \max & \sum_{s\in S} z_s, \\ \text{s.t.} & \log\left(\sum_{s\in S(i,j)} e^{z_s}\right) - \log\left(x_{ij}(\mathbf{p})\right) \leq 0, \quad \forall (i,j)\in L, \\ & \mathbf{p}\in\mathbf{P}_f. \end{array} \tag{13}$$

where z_s is interpreted as the logarithm of the session rate y_s, i.e., $z_s = \log(y_s)$.

In order to solve (13), we first transfer the link capacity constraints to the objective, and then solve the unconstrained problem by the subgradient method.

Denote $\mathbf{z} = (z_s, s\in S)$, and $\mathbf{w} = (\mathbf{p},\mathbf{z})$. Define $\tilde{U}_s(\mathbf{w}) = z_s$ for session $s\in S$, and define $g_l(\mathbf{w}) = \log\left(\sum_{s\in S(i,j)} e^{z_s}\right) - \log(x_{ij}(\mathbf{p}))$ for link $l=(i,j)\in L$. (13) is rewritten as

$$\begin{array}{rl} \max & \sum_{s\in S}\tilde{U}_s(\mathbf{w}), \\ \text{s.t.} & g_l(\mathbf{w}) \leq 0 \quad \forall l\in L, \\ & \mathbf{w}\in\mathbf{W}. \end{array} \tag{14}$$

where $\mathbf{W}$ represent the region in which $\mathbf{p}\in\mathbf{P}_f$.

We can transfer the constraints into the objective by using the "exact penalty" method. The unconstrained problem is

$$\begin{array}{rl} \max & \sum_{s\in S}\tilde{U}_s(\mathbf{w}) - \kappa\sum_{l\in L}\max\{0, g_l(\mathbf{w})\}, \\ \text{s.t.} & \mathbf{w}\in\mathbf{W}, \end{array} \tag{15}$$

where κ, the "penalty scaling factor", is a positive constant. Note that the term $\kappa\max\{0,g_l(\mathbf{w})\}$ can be interpreted as the penalty associated with the violation of the capacity constraint of link l, and it is called "exact penalty" since the set of optimal solutions to (15) coincides with the set of optimal solutions to (14) when κ is greater than a certain threshold value [2].

We now solve (15) using the subgradient method. Let $p_{ij}^{(n)}$ and $z_s^{(n)}$ denote the values of p_{ij} and z_s at the nth step respectively, and $\mathbf{p}^{(n)} = (p_{ij}^{(n)}, (i,j)\in L)$. Let $\tilde{\mathbf{x}}^{(n)} = \mathbf{x}(\mathbf{p}^{(n)})$ denote the link rate vector at iteration n. For each link $(i,j)\in L$, define the "link congestion indicator" for link (i,j) at the nth iteration, $\varepsilon_{ij}^{(n)}$, as

$$\varepsilon_{ij}^{(n)} = \begin{cases} 0 & \text{if } \sum_{s\in S(i,j)} e^{z_s^{(n)}} \leq \tilde{x}_{ij}^{(n)}, \\ 1 & \text{otherwise.} \end{cases} \tag{16}$$

We update z_s and p_{ij} using their subgradients,

$$z_s^{(n+1)} = z_s^{(n)} + \gamma\left(1 - \kappa e^{z_s^{(n)}}\frac{\sum_{(i,j)\in L(s)}\varepsilon_{ij}^{(n)}}{\sum_{r\in S(i,j)} e^{z_r^{(n)}}}\right) \tag{17}$$

$$p_{ij}^{(n+1)} = p_{ij}^{(n)} - \gamma\kappa\sum_{(s,t)\in L}\frac{\varepsilon_{st}^{(n)}}{x_{st}^{(n)}}\cdot\frac{\partial x_{st}}{\partial p_{ij}}(\mathbf{p}^{(n)}) \tag{18}$$

where γ is the step size, and $\frac{\partial x_{st}}{\partial p_{ij}}$ is defined in (9).

Since e^{z_s} (= y_s) is interpreted as the rate of session s, $\frac{e^{z_s}}{\sum_{r\in S(i,j)} e^{z_r}}$ in (17) can be interpreted as the fraction of the overall traffic on link (i,j) contributed by session s.

In (18), $\frac{\partial x_{st}}{\partial p_{ij}}(\mathbf{p}^{(n)})$ depicts how the attempt probability on link (i,j) impacts the rate on link (s,t), and this term is weighted by the inverse of the rate on link (s,t).

We can show that, following the procedure based on the subgradient method, the session rates and link attempt probabilities "converge to a neighborhood around the optimum", and the size of the neighborhood becomes arbitrarily small with decreasing step-size (proof in [11]).

3) Comparison of the Algorithms: Both dual- and primal-based algorithms solve the end-to-end proportional fairness in a distributed manner. From a practical viewpoint, each algorithm has certain advantages over the other.

In the dual-based algorithm, the separation between the transport layer and the link layer is better maintained. Also note that the dual-based algorithm has embedded loops and works in two time scales. In the inner loop (at a smaller time scale), the transport layer searches for the optimal session rates and link prices (dual variables), and in the outer loop (at a larger time scale), the link layer adjusts the link attempt probabilities and thereby updates the link rates. Note that the convergence process at the transport layer (inner loop) can be time consuming in some cases, and this may slow down the overall convergence. In contrast, the primal-based algorithm shows lesser modularity than the dual-based algorithm. Updates occur at the transport layer and at the link layer at the same time scale, and information is exchanged between different layers at a faster time-scale than that in the dual-based algorithm. In terms of the overall convergence speed, however, the two algorithms are quite comparable [11].

IV. Max-Min Fairness at the Link Layer

Before discussing the approaches to provide max-min fairness at the link layer for a random access ad hoc network, we introduce in the notion of a *directed link graph*, which captures the interference relationship between all the links in an Aloha ad hoc network. As we will see later in this section, the directed link graph also determines how the max-min rate of a link compares with that that of any other link in the network.

A *directed link graph*, $G_L = (V_L, E_L)$, is a graph where each vertex stands for a link in the original network. There is a directed edge from link (i,j) to link (s,t) in the directed link graph if and only if a successful transmission on link (s,t) requires that link (i,j) be silent. If there is a directed edge from (i,j) to link (s,t), then link (i,j) and (s,t) is said to be a *neighbors* of each other. Note that the notion of a neighbor represents a symmetric relationship, i.e., if (s,t) is a neighbor of (i,j), then (i,j) is also a neighbor of (s,t)

We first focus on attaining a weaker notion of max-min fairness, in which only the minimum link rate is maximized. We then discuss an approach for maximizing the link rates in a *lexicographic* manner.

A. Maximizing the Minimum Rate at the Link Layer

1) Problem Formulation and the Equivalent Convex Program: The question of maximizing the minimum link rate can be posed as follows:

$$\begin{array}{ll} \max & x, \\ \text{s.t.} & x \le p_{ij}(1-P_j)\prod_{k\in K_j\setminus\{i\}}(1-P_k)\ \forall (i,j)\in L, \\ & \mathbf{p}\in \mathbf{P}_f, \end{array} \tag{19}$$

where x corresponds to the minimum link rate. It is worth noting that this problem is non-convex and non-separable.

Although (19) is not a convex program, it is equivalent to the following convex program [9]:

$$\begin{array}{lll} \min & \frac{1}{2}\sum_{(i,j)\in E} u_{ij}^2, & \\ \text{s.t.} & u_{ij} \le \log(x_{ij}(\mathbf{p})) \quad \forall (i,j)\in L, & \\ & u_{ij} \le u_{st} \qquad \forall (i,j)\in L \quad \forall (s,t)\in \mathfrak{L}_{ij}, & \\ & \mathbf{p}\in \mathbf{P}_f, & \end{array} \tag{20}$$

where $\mathfrak{L}_{ij}$ is the set of neighbors of link (i,j) in the directed link graph.

2) Dual Method: We solve (20) using the dual method. The Lagrangian of (20) is given by

$$\begin{aligned} \mathscr{L}(\mathbf{p},\mathbf{u},\mu,\lambda) \;=\; & \frac{1}{2}\sum_{(i,j)\in L} u_{ij}^2 + \sum_{(i,j)\in L}\sum_{(s,t)\in\mathfrak{L}_{ij}} \lambda_{ij,st}(u_{ij}-u_{st}) \\ & + \sum_{(i,j)\in L} \mu_{ij}(u_{ij}-\log(x_{ij}(\mathbf{p}))), \end{aligned} \tag{21}$$

where μ_{ij} and $\lambda_{ij,st}$ are Lagrange multipliers.

Denote $\mathbf{u} = (u_{ij}, (i,j)\in L)$, $\mu = (\mu_{ij}, (i,j)\in L)$, and $\lambda = (\lambda_{ij,st}, (i,j)\in L, (s,t)\in K_{ij})$. The dual function is $\mathscr{D}(\mathbf{p},\mathbf{u}) = \min_{\mathbf{p},\mathbf{u}} \mathscr{L}(\mathbf{p},\mathbf{u},\mu,\lambda)$, and the dual problem is $\max_{\mu,\lambda}\mathscr{D}(\mathbf{p},\mathbf{u})$. It is easily seen that at the optimum λ^* and μ^*, the corresponding $\mathbf{p}$ and $\mathbf{u}$ is the solution to the primal problem.

Minimizing (21) with respect to $\mathbf{p}$, $\mathbf{u}$, we obtain

$$p_{ij} = \frac{\mu_{ij}}{\sum_{k\in O_i}\mu_{ik} + \sum_{k\in I_i}\mu_{ki} + \sum_{k\in K_i}\sum_{l\in I_k\setminus\{i\}}\mu_{lk}} \tag{22}$$

and

$$u_{ij} = \begin{cases} -\zeta_{ij} & \text{if } \zeta_{ij} \ge 0 \\ 0 & \text{if } \zeta_{ij} < 0 \end{cases} \tag{23}$$

where $\zeta_{ij} = \mu_{ij} + \sum_{(s,t)\in\mathfrak{L}_{ij}}(\lambda_{ij,st} - \lambda_{st,ij})$

The dual problem can then be solved using gradient projection method, where the Lagrange multipliers are adjusted in the direction of the gradient $\nabla\mathscr{D}(\mathbf{p},\mathbf{u})$:

$$\lambda_{ij,st}(n+1) = \left[\lambda_{ij,st}(n) + \gamma\frac{\partial\mathscr{D}}{\partial\lambda_{ij,st}}\right]^+ = [\lambda_{ij,st}(n) + \gamma(u_{ij}-u_{st})]^+ \tag{24}$$

$$\mu_{ij}(n+1) = \left[\mu_{ij}(n) + \gamma\frac{\partial\mathscr{D}}{\partial\mu_{ij}}\right]^+ = [\mu_{ij}(n) + \gamma(u_{ij} - \log(x_{ij}(\mathbf{p})))]^+ \tag{25}$$

Here, $\gamma > 0$ is the step size, and $[z]^+ = \max\{z, 0\}$.

If we let $\{\mathbf{p}^{(n)}, \mathbf{u}^{(n)}\}$ denote the sequence generated by following the procedure stated in (22)-(25), we can show that the limit point of this sequence is an optimal solution to (20) when γ is sufficiently small [9].

3) Subgradient Method: We can also solve (20) using the subgradient method.

Define the "link congestion indicator" for link (i,j) at step n, $\varepsilon_{ij}^{(n)}$, as

$$\varepsilon_{ij}^{(n)} = \begin{cases} 0 & \text{if } u_{ij} \le \log(x_{ij}(\mathbf{p})), \\ 1 & \text{otherwise.} \end{cases} \tag{26}$$

Also define the "link relation indicator" for link (i,j) and its neighboring link (s,t) at step n, $v_{ij,st}^{(n)}$, as

$$v_{ij,st}^{(n)} = \begin{cases} 0 & \text{when } u_{ij} \le u_{st}, \\ 1 & \text{when } u_{ij} > u_{st}. \end{cases} \tag{27}$$

We can update p_{ij} and u_{ij} using their subgradients, as follows:

$$u_{ij}^{(n+1)} = u_{ij}^{(n)} - \gamma\left(u_{ij}^{(n)} + \kappa\sum_{(s,t)\in\mathfrak{L}_{ij}}\left(v_{ij,st}^{(n)} - v_{st,ij}^{(n)}\right) + \kappa\varepsilon_{ij}^{(n)}\right), \tag{28}$$

$$p_{ij}^{(n+1)} = p_{ij}^{(n)} + \gamma\kappa \left(\frac{\varepsilon_{ij}^{(n)}}{p_{ij}^{(n)}} - \frac{\sum_{k\in I_i} \varepsilon_{(k,i)}^{(n)} + \sum_{k\in N_i} \sum_{l\in I_k\setminus\{i\}} \varepsilon_{(l,k)}^{(n)}}{1-P_i^{(n)}} \right), \tag{29}$$

where κ is a positive constant, and γ is the step size.

We can show that, following the procedure based on the subgradient method given in (26)-(29), **p** and **u** will converge to a neighborhood around the optimum, and the size of the neighborhood becomes arbitrarily small with decreasing step size (see the proof in [9]).

B. Maximizing the Minimum Link Rates Lexicographically

In the commonly used notion of max-min fairness [1], the minimum rates are maximized in a lexicographic manner. More specifically, a max-min fair rate allocation algorithm should maximize the minimum rate, then maximize the second minimum rate, then maximize the third minimum rate, and so on. The notion of (lexicographic) max-min fairness can be defined more formally as follows. Let $\mathbf{x} = (x_{ij}, (i,j) \in L)$ denote the vector of rates for all links in the active communication set L (also referred to as the *allocation vector*), and $\vec{\mathbf{x}}$ be the allocation vector $\mathbf{x}$ sorted in nondecreasing order. An allocation vector $\mathbf{x}_1$ is said to be *lexicographically greater* than another allocation vector $\mathbf{x}_2$, denoted by $\mathbf{x}_1 \succ \mathbf{x}_2$, if the first non-zero component of $\vec{\mathbf{x}}_1 - \vec{\mathbf{x}}_2$ is positive. Consequently, an allocation vector $\mathbf{x}_1$ is said to be *lexicographically no less than* than another allocation vector $\mathbf{x}_2$, denoted by $\mathbf{x}_1 \succeq \mathbf{x}_2$, if $\vec{\mathbf{x}}_1 - \vec{\mathbf{x}}_2 = 0$, or the first non-zero component of $\vec{\mathbf{x}}_1 - \vec{\mathbf{x}}_2$ is positive. A rate allocation is said to be (lexicographic) max-min fair if the corresponding rate allocation vector is lexicographically no less than any other feasible rate allocation vector.

1) Background and Definitions: We use the notation $(i,j) \rightsquigarrow (s,t)$ to denote that there is a path from link (i,j) to link (s,t) in the directed link graph. We can then show the following property (see [12] for proof): if $(i,j) \rightsquigarrow (s,t)$ in the directed link graph, then $x_{ij}^* \le x_{st}^*$, where x_{ij}^* and x_{st}^* are the lexicographic max-min fair rates for link (i,j) and (s,t), respectively.

In the directed link graph $G_L = (V_L, E_L)$, a *strongly connected component* is a maximal set of vertices $C \subseteq V_L$ such that for every pair of vertices u and v in C, we have both $v \rightsquigarrow u$ and $u \rightsquigarrow v$, that is, vertices u and v are reachable from each other. Therefore, from the above property we have that all the links belonging to the same strongly connected component in the directed link graph have the same (lexicographic) max-min fair rate value at the optimum. This also implies that if the entire directed link graph consists of a single strongly connected component, then the (lexicographic) max-min fair rates can be obtained using the procedure for maximizing the minimum rate, as described in Section IV-A.

Next we show how for a general network (for which the directed link graph may consist of multiple strongly connected components), the (lexicographic) max-min rate can be obtained through solving a sequence of convex optimization problems, each of which is similar in nature to the problem discussed in Section IV-A.

Next we define the notion of a *bottleneck* link, which will be used in describing our solution procedure. Intuitively, a link is said to be a bottleneck for the problem defined in (19) if the minimum link rate in the network decreases when the capacity of that link decreases by any positive amount. More formally, a bottleneck link is defined as follows. Consider the following problem, which is formed from (19) by introducing a perturbation in the constraint corresponding to link (u,v):

$$\begin{array}{ll} \max & x, \\ \text{s.t.} & x \le p_{uv}(1-P_v)\prod_{k\in K_v\setminus\{i\}}(1-P_k) - \varepsilon, \\ & x \le p_{ij}(1-P_j)\prod_{k\in K_j\setminus\{i\}}(1-P_k) \;\; \forall (i,j) \in L\setminus\{(u,v)\}, \\ & \mathbf{p} \in \mathbf{P}_f. \end{array} \tag{30}$$

The optimal value of (30) is a function of ε, and we denote it as $U_{uv}^*(\varepsilon)$. Note that $U_{uv}^*(0)$ is actually the optimal objective value in (19). We define link (u,v) as a bottleneck link if $U_{uv}^*(\varepsilon) < U_{uv}^*(0)$ for any positive ε.

2) Multi-Step Solution Procedure: We start with solving the problem defined in (19), which maximizes the minimum link rate in the network. Refer to Section IV-A for the distributed implementation. At the optimum, we shall identify the bottleneck links, and fix their attempt probabilities (more on this bottleneck link identification step is discussed later). Let L_1 denote the set of bottleneck links identified at this step. It is worth noting that (u,v) is a bottleneck link if and only if all other links in its strongly connected component are also bottleneck links. Moreover, if (u,v) is a bottleneck link, and $(s,t) \rightsquigarrow (i,j)$, then (s,t) (and every other link in its strongly connected component) is also a bottleneck link. These facts imply that, once a bottleneck link is identified, we can identify one or more strongly connected components (in the directed link graph) that consist only of bottleneck links. As discussed shortly, this allows us to reduce the number of steps the procedure has to run, thereby speeding up the overall convergence time.

In the next step, we solve (19) assuming fixed attempt probabilities for all bottleneck links identified in the previous step, i.e. only the attempt probabilities of links in $L \setminus L_1$ are considered variables. After the optimal solution is obtained, we again identify the set of bottleneck links, denoted as L_2, and fix their attempt probabilities to their optimal values; therefore, the attempt probabilities for these links remain fixed in the subsequent steps. The value of the objective function at optimality represents the second minimum rate in the optimal (lexicographic) max-min rate vector.

The above procedure is repeated until the attempt probabilities of all links have been fixed, i.e., when every link has been identified as a bottleneck link in one of the steps.

Let $\mathbf{x}^*$ and $\mathbf{p}^*$ denote the vector of link rates and link attempt probabilities solved by the above procedure. We can then show (see [12] for the proof) that $\mathbf{x}^*$ is the (lexicographic) max-min fair rate allocation. Also, $\mathbf{p}^*$ is the unique link attempt probabilities that attain the optimal rates.

Note that the number of steps in the solution procedure mentioned above is, in the worst case, equal to the number of strongly connected components in the directed link graph.

Recall that each step requires solving a convex problem similar to (19) (which can be solved in a distributed manner using the procedures described in Section IV-A), and a bottleneck link identification procedure. It is possible to identify whether a specific link is a bottleneck or not in a fully distributed manner, using the basic definition of a bottleneck link, as described above. However, this does not directly lead to a very efficient solution procedure for identifying the set of bottleneck links. Therefore, development of an efficient distributed algorithm for identifying the bottleneck links at any step of the algorithm remains an open question.

V. Open Issues

Our research results has been able to answer several interesting questions on how globally fair rates in a random access wireless network can be attained through distributed approaches. However, there are quite a few important questions that remain unanswered, and merit further investigation.

As discussed Section IV-B, the identification of the bottleneck links is a key component of the solution procedure for finding the max-min fair rates. Therefore, in order to develop an efficient distributed solution procedure for attaining the max-min fair rates, bottleneck links must be identified efficiently. It is tempting to hypothesize here that the bottleneck links are those which are active in the optimal solution of (19), or those that have non-zero optimal dual variables. However, as we argue in [12], these hypotheses are not true. Identifying the bottleneck links based on the basic definition is computationally inefficient, as we have already discussed. Ideally, we would like to obtain the bottleneck links directly from the optimal solution of (19), with minimal additional computation. Whether this is can be done, either exactly or approximately, remains an open question.

Note that we have considered the max-min fairness question only at the link layer. The question of attaining max-min fairness for end-to-end sessions via distributed means, remains open. In this case too, we would like to have results that are similar in nature to those in the proportional fairness case. In other words, our first goal could be to obtain an iterative algorithm that operates at both the transport and link layers (requiring cross-layer coordination), and attains the globally max-min fair end-to-end session rates on convergence. This appears to be a very challenging question, particularly after considering the fact that the max-min fairness problem at the link layer is significantly more complex than the corresponding proportional fairness problem.

We have seen in Section III-A that proportional fairness at the link layer can be attained using only local (two-hop) topology information. Other than being fully distributed, the optimization algorithm in this case is very efficient, since it is an open-loop approach requiring each node to communicate with its neighborhood only once (to exchange in-degree information). The approaches for the end-to-end proportional fairness case, or those for max-min fairness, are iterative and closed-loop in nature, and could take a long time to converge. The question of whether global fairness goals can be attained using local topology information in such cases is currently under investigation. Note that in the context of the link layer, 'local topology' would typically refer to the network topology within a few hops of a node; in the context of end-to-end sessions, the same term may refer to the network topology in the neighborhood of all nodes on the end-to-end path of a session. Attaining the exact optimum under such stringent constraints on the degree of coordination/communication may be difficult; a less ambitious yet practically significant goal would be to attain a solution that is provably close the optimum.

The only fairness metrics we have considered so far are max-min and proportional fairness. While these represent the two most popular fairness metrics, it remains to be seen whether and how our results can be extended to other fairness metrics, like those that can be defined in terms of a general concave utility maximization problem [3], [4], [5].

Finally, note that the wireless transmission and interference model that we have considered here follows the simple Aloha protocol. Random access protocols that employ carrier sensing are significantly more complex to model, analyze and optimize. In a recent work [10], we have modeled (under certain simplifying assumptions) a CSMA/CA protocol with fixed back-off window sizes. We have also used some of our insights and results obtained from our study of the Aloha model, to optimize the back-off window sizes. Whether our analysis and optimization techniques can be extended further to the case of CSMA/CA with dynamic back-off window adaptations, as well as other CSMA based protocols, remains to be investigated.

References

[1] D. Bertsekas and R. Gallagher, *Data Networks*, Prentice Hall, 1992.

[2] N. Z. Shor, *Minimization Methods for Non-differentiable Functions*, Springer-Verlag, 1985.

[3] F. P. Kelly, "Charging and Rate Control for Elastic Traffic," *European Transactions on Telecommunications*, vol. 8, no. 1, 1997, pp. 33-37.

[4] F. Kelly, A. Maulloo, D. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability," *Journal of Operations Research Society*, vol. 49, no. 3, 1998, pp. 237-252.

[5] S. Low, D. E. Lapsley, "Optimization Flow Control, I: Basic Algorithm and Convergence," *IEEE/ACM Transactions on Networking*, vol. 7, no. 6, December 1999.

[6] K. Kar, S. Sarkar, L. Tassiulas, "Achieving Proportionally Fair Rates using Local Information in Multi-hop Aloha Networks," in *Proceedings of Annual Allerton Conference*, Urbana-Champaign, USA, October 2003.

[7] K. Kar, S. Sarkar, L. Tassiulas, "A Simple Rate Control Algorithm for Maximizing Total User Utility," in *Proceedings of Infocom 2001*, Anchorage, USA, April 2001.

[8] M. Chiang, "To Layer or Not To Layer: Balancing Transport and Physical Layers in Wireless Multihop Networks," in *Proceedings of Infocom 2004*, Hong Kong, China, March 2004.

[9] X. Wang, K. Kar, "Distributed Algorithms for Max-min Fair Rate Allocation in Aloha Networks," in *Proceedings of Annual Allerton Conference*, Urbana-Champaign, USA, October 2004.

[10] X. Wang, K. Kar, "Throughput Modelling and Fairness Issues in CSMA/CA Based Ad-Hoc Networks," in *Proceedings of IEEE Infocom 2005*, Miami, March 2005.

[11] X. Wang, K. Kar, "Cross-Layer Rate Control for End-to-End Proportional Fairness in Wireless Networks with Random Access," in *Proceedings of ACM MobiHoc 2005*, Urbana-Champaign, Illinois, May 2005.

[12] X. Wang and K. Kar, "Achieving Max-Min Fairness in an Ad Hoc Network with Random Access," *Technical Report*, ECSE Department, Rensselaer Polytechnic Institute, Troy, NY, Feburary 2006. Available at http://www.ecse.rpi.edu/∼koushik/TR-LMM.pdf.