

ENFORCING END-TO-END PROPORTIONAL FAIRNESS WITH BOUNDED BUFFER OVERFLOW PROBABILITIES IN AD-HOC WIRELESS NETWORKS

Nikhil Singh¹ and Ramavarapu Sreenivas²

¹Yahoo! Labs, Champaign, IL 61820, USA

²Coordinated Science Laboratory & Industrial and Enterprise Systems Engineering,
University of Illinois at Urbana-Champaign, Urbana, IL 61820

ABSTRACT

In this paper, we present a distributed flow-based access scheme for slotted-time protocols, that provides proportional fairness in ad-hoc wireless networks under constraints on the buffer overflow probabilities at each node. The proposed scheme requires local information exchange at the link-layer and end-to-end information exchange at the transport-layer, and is cast as a nonlinear program. A medium access control protocol is said to be proportionally fair with respect to individual end-to-end flows in a network, if the product of the end-to-end flow rates is maximized. A key contribution of this work lies in the construction of a distributed dual approach that comes with low computational overhead. We discuss the convergence properties of the proposed scheme and present simulation results to support our conclusions.

KEYWORDS

Wireless LAN, Access protocols, Resource management.

1. INTRODUCTION

In this paper we consider an ad-hoc wireless network [1] that carries several flows between various source-destination pairs under a slotted-time medium access control (MAC) protocol. Specifically, we are interested in a distributed scheme for the assignment of the network's resources among flows, which is fair in terms of end-to-end flow rates. We assume that each node in the network has a finite buffer assigned to each flow routed through it. In addition to the objective of fairness, we are also interested in ensuring that the buffer overflow probability at each node does not exceed a pre-determined value.

The literature contains several references to fairness and its impact on the network performance. It has been observed by many researchers that the contention control mechanism used in 802.11-MAC [2] can be inefficient [3]. In [4], [5] a list of modifications is presented, that eliminates the unfairness commonly seen in the 802.11-MAC. The literature also contains a large volume of references (cf. [6], [7], [8], for example) where it is assumed that each network flow/link is associated with a concave utility function that could be maximized. In particular, for proportional fairness, it is assumed that the utility function has the form of $\log x$, where x denotes the flow rate

[6]. It is of interest to schedule individual transmissions on the links so as to maximize the sum of the utilities of the consumers. To achieve fairness, the schemes outlined in the above-mentioned references use a penalty function that is updated by some form of feedback from the network. Using an appropriately defined cost that is implicitly dependent on the requested rates of each node within a neighbourhood, the penalty is typically the total cost of all nodes in the network. A node maximizes (its view of) a common performance function, given by the difference between the total utility and the penalty. An overview of network resource allocation through utility maximization is presented in [9].

In [10], the authors have addressed the problem of providing proportional fairness by considering joint optimization at both transport and link layers. Two algorithms are proposed for solving the problem in a distributed manner that converges to the globally optimal solutions. These results, generalized in [11], are based on the dual and the primal algorithms in convex optimization and need end-to-end feedback information to update variables maintained at the nodes. The algorithms presented in [10], [11] are oblivious of the queue dynamics of the network, which may increase delays and packet loss. Although our work is closely related to [10], [11], the problem formulation and the proposed solution differ significantly.

In [12], the solution approach uses a class of *queue backpressure random access algorithms* (QBRA), where the actual queue-lengths of the flows are used to determine any node's channel access probabilities. In this distributed algorithm, a node uses the queue-length information in a close neighbourhood to determine its channel access probability to achieve proportionally fair rates and queue stability. This scheme has the advantage that no optimization needs to be performed and nodes can achieve proportional fairness just by exchanging the queue information in the local neighbourhood. However, the frequency of exchange of this information plays a vital role in determining the performance of this algorithm. In optimization-based schemes, once the flow rates have converged to the optimum, the frequency of information exchange does not play a significant role until the network topology, or the number of flows in the network, change.

In a different approach, several policies have recently been proposed for achieving rates close to the maximum throughput region through dynamic link scheduling [13], [14], [15], [16]. These scheduling algorithms use maximal matchings in every time slot using local contention algorithms and achieve near maximal schedules. Some policies also guarantee fairness of rate allocation among different sessions.

Quality of Service(QoS) is an important issue in ad-hoc wireless networks. Service guarantees can be provided for delays, packet loss, jitter and throughput based on the application requirements. Our approach in this work is to combine the QoS guarantee in addition to providing proportional fairness. Our main contributions are as follows:

1. We derive an expression for the buffer overflow probabilities for discrete-time queues. This derivation uses the fact that there cannot be simultaneous arrivals and departures at a node within the same slot in Aloha-type networks that do not have packet capture mechanisms.
2. Using the expression for buffer overflow probabilities mentioned above, we show that an upper bound on the buffer overflow probability translates to an upper bound on the utilization or load, which can then be used as constraints in an appropriately posed convex minimization problem under convex constraints. This is a reformulation of the proportionally fair end-to-end rate allocation problem. A distributed dual approach is then used to solve this convex minimization problem using an appropriate Lagrangian function. The dual problem is solved using a projected gradient method.
3. Finally, after making some observations about the distributed implementation of the above-mentioned dual scheme, we present simulation results showing the satisfactory

performance of our proposed algorithm in terms of fairness and QoS.

The rest of the paper is organized as follows. Section 2 presents the network model that is used in the rest of the paper. We then formulate the rate control problem as a convex optimization instance with bounds on the buffer overflow probabilities at each node. In section 3, we discuss the dual-based solution approach and present a distributed implementation to achieve flow-based proportional fairness. The convergence of this algorithm to the unique global optimum is established. Section 4 contains the details of the experimental results verifying the optimality of the proposed scheme. Conclusions are provided in section 5.

2. PROBLEM FORMULATION

2.1. Wireless Network Model

We assume the following:

1. Time is divided into slots of equal duration.
2. A successful transmission in a time-slot implies collision free data transmission in that slot.
3. The transmitting nodes always have data packets to transmit (i.e. we do not consider \square the arrival rates of packets for different flows, and assume that all flows have packets to transmit at all times).
4. Nodes cannot transmit and receive packets at the same time.
5. The receipt of more than one packet within the same time-slot will result in a collision.
6. Nodes in the network have a buffer of fixed size assigned to each flow that is routed through it.
7. We also assume there is a unique route for each flow within the network (which would be the case if we used DSDV [17] as the routing protocol, for example).

Additionally, we only consider unicast flows for our derivations.

An ad-hoc wireless network carrying a collection of flows, is represented as an undirected graph $G=(V, E)$, where V represents the set of *nodes*, and $E \subseteq V \times V$ is a symmetric relationship (i.e. $(i, j) \in E \Leftrightarrow (j, i) \in E$), that represents the set of bidirectional *links*. We assume all links of the network have the same capacity, which is normalized to unity. The 1-hop neighbourhood of node $i \in V$ is represented by the symbol $N(i)$. When a node i communicates with a node $j \in N(i)$, we can represent it as an appropriate orientation of the link (i, j) in E , where i is the origin and j is the terminus. The context in which $(i, j) \in E$ is used should indicate if it is to be interpreted as a directed edge with i as origin and j as terminus. The set of flows, using a link $(i, j) \in E$ with $i(j)$ as origin (terminus), is denoted by $\mathcal{F}(i, j)$.

When node i intends to transmit data to node $j \in N(i)$ for the l -th flow ($l \in \mathcal{F}(i, j)$), it would transmit data in the appropriate time-slot with probability $p_{i,j,l}$. $\mathcal{P}_{ij} = \sum_{l \in \mathcal{F}(i,j)} p_{i,j,l}$, denotes the probability that node i transmits data to node j , and $\mathcal{P}_i = \sum_{j \in V} \mathcal{P}_{i,j}$, denotes the probability that node i will be transmitting to some node in its 1-hop neighbourhood for some flow. The probabilities $p_{i,j,l}$'s should be chosen such that \mathcal{P}_i is not greater than unity for any node $i \in V$.

2.2. Link Success Probability Expression

The probability of successful data transmission over link $(i, j) \in E$ for flow $l \in \mathcal{F}(i, j)$, denoted by

$S_{i,j,l}$, is given by the expression

$$S_{i,j,l} = p_{i,j,l} \left(1 - \sum_{(j,m) \in E, n \in \mathcal{F}(j,m)} p_{j,m,n} \right) \prod_{o \in N(j) - \{i\}} \left(1 - \sum_{(o,p) \in E, q \in \mathcal{F}(o,p)} p_{o,p,q} \right) \quad (1)$$

This is also the rate or the attainable throughput of flow l over link (i, j) .

2.3. Problem Statement

Consider an ad-hoc wireless network where there are r flows in the network. Each flow has a utility function associated with it, whose value is determined by the logarithm of the flow rate. The objective is to maximize the sum of the logarithms of the flow-rates under the operational constraints outlined below. We denote the logarithm of the rate of the l -th flow as f_l . The end-to-end proportionally fair flow control problem can be stated as

$$\max_{p_{i,j,l}} \sum_l f_l \quad (2)$$

where $(i, j) \in E$ and $l \in \{1, 2, \dots, r\}$, subject to additional constraints.

Let us assume that the l -th flow ($1 \leq l \leq r$) spans over k_l links. We use the notation $\langle l, q \rangle \in E$ to denote the l -th-flow's q -th-link, where $q \in \{1, 2, \dots, k_l\}$ is indexed in ascending order starting from the source and terminating at the destination. Thus, $\langle l, q \rangle = (i, j)$ implies the l -th-flow's q -th-link from the source has i as the source node and j as the destination node. If $\langle l, q \rangle = (i, j) \in E$ then we use the notation $S_{l,q}$ to denote $S_{i,j,l}$. The logarithm of the rate of l -th flow over link $\langle l, q \rangle$ is represented as $f_{l,q}$.

Let $\mathbf{p} = (p_{i,j,l}, 1 \leq l \leq r, 1 \leq q \leq k_l, \langle l, q \rangle \in E)$ be the vector of access probabilities of all the

flow over each link in the network and $\hat{\mathbf{f}} = (f_{l,q}, 1 \leq l \leq r, 1 \leq q \leq k_l, \langle l, q \rangle \in E)$ the vector of the logarithm of link rates of all flows.

In the case of multi-hop wireless networks, the rate of any flow is the same as the rate of the bottleneck link in that flow. The logarithm of the rate of the l -th flow is $\min\{f_{l,q} : 1 \leq q \leq k_l\}$. Hence, the problem can be stated as $\max_{p_{l,q}} \sum_l \min\{f_{l,q}, 1 \leq q \leq k_l\}$, subject to capacity

constraints, and additional constraints on the buffer overflow probabilities which is addressed in the next subsection.

2.4. Buffer Overflow Probability of a Tandem of Discrete-Time Queues

The results in reference [18] can be paraphrased as follows – for a discrete-time queue of capacity M , with a packet arrival probability p_a , and a probability p_d ($p_d > p_a$) of a packet departure from a non-empty buffer, the probability of seeing i -many packets at any time-instant in the buffer in steady state is given by the expression

$$\frac{1 - \left(\frac{p_a}{p_d}\right)}{1 - \left(\frac{p_a}{p_d}\right)^{M+1}} \left(\frac{p_a}{p_d}\right)^i \quad (3)$$

Using the time-reversibility of the underlying Markov-chain, and the mutual independence of the simultaneous states of the buffers, reference [18] also establishes that the joint stationary state probability of a tandem of discrete-time queues is the product of the distributions of each queue taken independently with an arrival probability of p_a , which is the probability of packet arrival into the first queue. This is essentially the discrete-time analogue of Jackson's result [19] involving tandems of M/M/1 queues. The key points of divergence between reference [18] and the present paper are presented below.

It should be noted that unlike the model assumed in reference [18], where arrival and departure events are permitted to occur concurrently, interference constraints in wireless networks do not permit the occurrence of certain simultaneous events. For instance, as a node cannot transmit and receive information at the same time, the simultaneous occurrence of an arrival and a departure from the discrete-time queue at the node cannot be permitted. Secondary interference constraints place additional restrictions on the set of simultaneous events that can occur among neighbouring nodes. Even when there are no restrictions on simultaneous events, reference [20] notes that it is cumbersome to use balance equations to arrive at an appropriate expression for the joint stationary probability for tandems of discrete-time queues. For situations where there are restrictions on the nature of concurrent events that can occur in a tandem of queues, such as those that model wireless networks, the joint stationary state probability of a tandem of discrete-time queues is not guaranteed to have the product-form of reference [18]. This notwithstanding, it is possible to characterize the marginal probability distribution of each queue in the tandem.

We first note that the analysis of reference [18] (cf. equations 1, 2 and the subsequent discussion of time-reversibility) applies *mutatis mutandis* to the case when utmost one packet is permitted to arrive, or depart from a single discrete-time queue of size M , along with the restriction that a simultaneous arrival and departure of a packet from the queue is not permitted. The probability of seeing i -many packets in the buffer at any time-instant in this restricted discrete-time queue is also given by equation 3. The probability of the queue of size M is non-empty is given by the expression

$$\frac{1 - \left(\frac{p_a}{p_d}\right)^M}{1 - \left(\frac{p_a}{p_d}\right)^{M+1}} \left(\frac{p_a}{p_d}\right)$$

and since the probability of a packet departure from a non-empty queue is p_d , the probability of a packet-departure from the discrete-time queue is given by

$$\frac{1 - \left(\frac{p_a}{p_d}\right)^M}{1 - \left(\frac{p_a}{p_d}\right)^{M+1}} \left(\frac{p_a}{p_d}\right) \times p_d < p_a.$$

It is not hard to see that if $M = \infty$, then the probability of a packet-departure from the discrete-

time queue is exactly equal to the probability of packet-arrival into the queue. For bounded queues ($M < \infty$) the output process of the queue is geometrically distributed with a parameter that is no greater than the input parameter p_a . Additionally, there can be no more than M -many consecutive departures, or, M -many consecutive arrivals to the discrete-time queue due to the bound on the buffer-size. We assume packets that arrive into a full-queue get dropped. This observation holds for a tandem of discrete-time queues. That is, the output process of each queue is geometrically distributed with a parameter that is no greater than that of the input to the first queue (i.e. p_a). This observation is used in establishing a bound on the buffer-overflow probabilities at each queue in a tandem of discrete-time queues in the following theorem.

Theorem 1.1: Consider a tandem of n discrete-time queues, each with buffer-size M , where at any discrete-time instant the probability of a packet-arrival into the first queue is p_a , and the probability of a packet-departure from the i -th, non-empty queue is p_{di} , ($i = 1, 2, \dots, n$). If

$p_{dj} = \min_{i=1, \dots, n} \{p_{di}\}$, and $\frac{p_a}{p_{dj}} < \frac{M}{M+1}$, then, the probability of seeing M packets in the i -th queue ($i = 1, \dots, n$) is no greater than

$$\left(\frac{1 - \frac{p_a}{p_{dj}}}{1 - \left(\frac{p_a}{p_{dj}}\right)^{M+1}} \right) \left(\frac{p_a}{p_{dj}} \right)^M$$

Proof: Suppose $\rho = \frac{p_a}{p_{dj}}$, we first note that the expression $\left(\frac{1 - \rho}{1 - \rho^{M+1}} \right) \rho^M$, increases

monotonically with respect to ρ if $\rho \leq \frac{M}{M+1}$. Let p_{ai} be the probability of a packet arrival into the

i -th queue, we know $p_{ai} \leq p_a$. If $\rho_i = \frac{p_{ai}}{p_{di}}$, since $p_{di} \geq p_{dj}$, it follows that $\rho_i \leq \rho < \frac{M}{M+1}$. The observation

follows directly from the monotonicity property mentioned above. ■

A direct consequence of theorem 1.1 is that if we are able to pick a p_a such that

$$\frac{p_a}{p_{dj}} < \left[\frac{\beta}{1 + \beta} \right]^{1/M},$$

then the buffer overflow probability at the i -th queue in the tandem of discrete-time queues will be no higher than β at all queues. In the next section, this observation is used in a convex programming solution to the problem of enforcing proportional fairness in the presence of constraints on the buffer overflow probabilities.

2.5. Problem Formulation with Buffer Overflow and Capacity Constraints

Let us assume the loss rate bounds for the l -th flow translates to each node along the flow sustaining a traffic intensity (ratio of arrival probability and departure probability at a node)

no more than $\rho_l \left(= \frac{p_a}{p_{dj}} \right)$.

Also, each link-rate in the network cannot exceed the capacity of that link given by (1). Since the logarithmic function is strictly increasing, each link constraint can be re-written as

$$f_{l,q} \leq \log(\mathcal{S}_{l,q}) \quad (5)$$

Each link constraint (5) forms a convex set over $(f_{l,q}, \mathbf{p})$. We also assume that there is a minimum achievable data-rate for each flow, i.e., $\exists \varepsilon, s.t. \varepsilon \leq f_{l,q}, \forall l, q (1 \leq l \leq r, 1 \leq q \leq k_l)$. Also, we assume that all the flows in the network have a maximum achievable data-rate i.e., $\exists \delta, s.t. f_{l,q} \leq \delta, \forall l, q (1 \leq l \leq r, 1 \leq q \leq k_l)$, (to accommodate network control traffic like routing messages, for example). We define the feasible set of access probabilities as,

$$\tilde{\mathcal{P}} = \left\{ \mathbf{p} : \sum_{\{l,q\}=(i,j), j \in N(i)} p_{l,q} \leq 1, e^\varepsilon \leq p_{l,q} \leq e^\delta, (i,j) \in E, l \in \mathcal{F}(i,j) \right\}.$$

Also, we define the *QoS region* as a set of vectors as defined by

$$\mathcal{G} = \left\{ \hat{\mathbf{f}} : \varepsilon \leq f_{l,1}, f_{l,q} \leq \delta, f_{l,1} \leq f_{l,q} + \delta_l, 2 \leq q \leq k_l \right\},$$

where $\delta_l = \log \rho_l$. The overall optimization problem can now be stated as:

$$\mathbf{V} : \max : \sum_l \min \{ f_{l,q} : 1 \leq q \leq k_l \}, \quad (6)$$

$$f_{l,q} \leq \log(\mathcal{S}_{l,q}), \forall \langle l, q \rangle \in E,$$

$$\mathbf{p} \in \tilde{\mathcal{P}}, \hat{\mathbf{f}} \in \mathcal{G}.$$

From the constraint imposed by the QoS region, we observe that for any feasible solution to \mathbf{V} , the first link will always have the lowest rate and hence it will be the bottleneck. Therefore for any feasible solution, the rate of any flow l , is same as $f_{l,1}$. We replace $f_{l,1}$ by f_l , and define the feasible set of flow rates as $\tilde{\mathcal{F}} = \{ \mathbf{f} : \varepsilon \leq f_l \leq \delta, \forall l \}$, where, $\mathbf{f} = (f_l, 1 \leq l \leq r)$, we can rewrite \mathbf{V} as the following convex optimization problem,

$$\mathbf{U} : \min \sum_l -f_l, \quad (7)$$

$$f_l \leq \log(\mathcal{S}_{l,1}), \forall \langle l, 1 \rangle \in E,$$

$$f_l \leq \log(\mathcal{S}_{l,q}) + \delta_l, \forall \langle l, q \rangle \in E, 2 \leq q \leq k_l,$$

$$\mathbf{p} \in \tilde{\mathcal{P}}, \hat{\mathbf{f}} \in \tilde{\mathcal{F}}.$$

3. SOLUTION APPROACH

3.1. Dual-based Algorithm

We can write the Lagrangian function for the problem stated in (7) as,

$$\mathcal{L}(\mathbf{f}, \mathbf{p}, \boldsymbol{\lambda}) = \sum_l -f_l + \sum_{l,1} \lambda_{l,1} (f_l - \log(\mathcal{S}_{l,1})) + \sum_{l,q(2 \leq q \leq k_l)} \lambda_{l,q} (f_l - \log(\mathcal{S}_{l,q}) - \delta_l). \quad (8)$$

Let us denote $\boldsymbol{\Lambda} = (\lambda_{l,q} : \forall l, 1 \leq q \leq k_l)$ as a vector of Lagrange multipliers. As the Slater constraint qualification is satisfied by the convex program given by (7), convex duality implies that at the optimum $\boldsymbol{\Lambda}^*$, the corresponding \mathbf{f} , \mathbf{p} are the solutions to the primal problem [21]. The dual problem can be solved using the gradient projection method similar to the scheme used in [22]. Note that the Lagrangian is separable in terms of the probabilities \mathbf{p} and the logarithm of the rates \mathbf{f} . The dual function can be stated as:

$$Q(\boldsymbol{\Lambda}) = \inf_{\mathbf{f} \in \mathcal{F}, \mathbf{p} \in \mathcal{P}} \mathcal{L}(\mathbf{f}, \mathbf{p}, \boldsymbol{\Lambda}). \quad (9)$$

The following proposition is significant for obtaining the distributed solution for the non-linear program given by (7).

Proposition 2.1: For a given $\boldsymbol{\Lambda}(\lambda_{l,q} \neq 0, \forall (l,q) \in E)$, the solution to $\inf_{\mathbf{f} \in \mathcal{F}, \mathbf{p} \in \mathcal{P}} \mathcal{L}(\mathbf{f}, \mathbf{p}, \boldsymbol{\Lambda})$ is given by:

$$p_{l,q} = \max \left\{ \min \left\{ \frac{\lambda_{l,q}}{\sum_{(y,z)=(i,k), k \in N(i)} \lambda_{y,z} + \sum_{(y,z)=(k,i), k \in N(i)} \lambda_{y,z} + \sum_{v \in N(i)} \sum_{(y,z)=(k,v), k \in N(v)-\{i\}} \lambda_{y,z}}, e^\delta \right\}, e^\varepsilon \right\}, \quad (10)$$

and,

$$f_l = \begin{cases} \varepsilon & \text{if } \sum_q \lambda_{l,q} \geq 1 \\ \delta_l & \text{if } \sum_q \lambda_{l,q} < 1 \end{cases} \quad (11)$$

Proof: Since the Lagrangian is convex with respect to \mathbf{p} , the unconstrained value of \mathbf{p} that yields the infimum of the Lagrangian is obtained by taking its derivative with respect to \mathbf{p} and equating it to zero. This results in the expression involving λ -terms in equation 10.

The denominator of this expression is essentially the sum of three terms. The first term is the sum of the Lagrange multipliers associated with all outgoing flows from node i . The second is the sum of the Lagrange multipliers associated with all incoming flows to node i . Finally, the third term is the sum of the Lagrange multipliers associated with all incoming flows to nodes in the one-hop neighborhood of node i (excluding the flows incoming from node i). It is not hard to show that this expression satisfies the constraint $0 \leq p_{l,q} \leq 1$ and $0 \leq P_i \leq 1$.

The remaining terms, and the structure of equation 10, are a direct consequence of the upper and

lower-bounds on $p_{l,q}$ ($e^\varepsilon \leq p_{l,q} \leq e^\delta$).

Also, for any l , the coefficient of f_l in the Lagrangian, is given by $(-1 + \sum_q \lambda_{l,q})$. When

$\sum_q \lambda_{l,q} \neq 1$, the infimum of Lagrangian with respect to f_l is either the lower or the upper bound of f_l , depending on the sign of its coefficient, which gives (11). When $\sum_q \lambda_{l,q} = 1$, f_l can take any value between $[\varepsilon, \delta]$. ■

The dual problem

$$\text{maximize: } Q(\Lambda) \quad (12)$$

$$\text{subject to: } \Lambda \in [\eta, \infty) \times [\eta, \infty) \times \dots \times [\eta, \infty)$$

where η is a small number close (but not equal) to zero, can now be solved using the subgradient projection method, where the Lagrange multipliers are adjusted in the direction of the subgradient

$$\lambda_{l,q}^{n+1} = \left[\lambda_{l,q}^n + \alpha^n \frac{\partial Q(\Lambda^n)}{\partial \lambda_{l,q}} \right]^+ \quad (13)$$

where $[z_{(l,q)}]^+ = \max\{\eta, z_{(l,q)}\}$ and $\frac{\partial x}{\partial y}$ denotes the subgradient of x with respect to y . The variable α^n is the step size at the n^{th} iteration that can either be a constant, or, diminishing step size that satisfies the requirements

$$\lim_{n \rightarrow \infty} \alpha^n = 0, \sum_{n=1}^{\infty} \alpha^n = \infty,$$

and the subgradient is given by,

$$\frac{\partial Q(\Lambda^n)}{\partial \lambda_{l,1}} = (f_l^n - \log(\mathcal{S}_{l,1}^n)), \quad (14)$$

$$\frac{\partial Q(\Lambda^n)}{\partial \lambda_{l,q}} = (f_l^n - \log(\mathcal{S}_{l,q}^n) - \delta_l), 2 \leq q \leq k_l, \quad (15)$$

where f_l and $\lambda_{l,q}$ are obtained from equations (10) and (11).

3.2. Convergence of the Proposed Scheme

Since $\mathbf{p} \in \tilde{\mathcal{P}}$, the norm of subgradients of the dual function given by (9), is bounded. Let Λ^* be the solution to (12) for $\eta = 0$. Also let $\hat{\Lambda}_\eta$ be the global maximum of (12) for $\eta > 0$. Let $\{\Lambda_\eta^n\}$ be the sequence generated by the subgradient method given by (13) for $\eta > 0$.

Lemma 2.2: Let $\|g(Q(\Lambda_\eta^n))\|_2 \leq G, \forall n \geq 0$. For every $\eta > 0$, $\exists \hat{\Lambda}_\eta$, such that

$$\|Q(\Lambda^*) - Q(\hat{\Lambda}_\eta)\|_2 \leq G\eta.$$

The proof follows directly from the concavity of the dual function, and the property of the subgradients [23].

Theorem 2.3: For every $\eta > 0$, every limit point of the sequence of $\{\Lambda_\eta^n\}$ obtained using the diminishing step size, is the global maximum of (12).

Proof: Chapter 2 of reference [23], presents a proof of convergence of approaches that use subgradient method involving diminishing step sizes, which can easily be extended to projected subgradient method for maximization of concave function over a convex set. ■

Theorem 2.4: For every $\eta > 0$, using the constant step size $\alpha^n = \alpha$, as $n \rightarrow \infty$,

$$Q(\hat{\Lambda}_\eta) - \liminf_{n \rightarrow \infty} (Q(\bar{\Lambda}_\eta)) \leq \frac{G\alpha^2}{2}, \text{ where } \bar{\Lambda}_\eta = \frac{1}{n+1} \sum_{i=0}^n \Lambda_\eta^i.$$

The proof of convergence of projected subgradient method with constant step size,

under assumption of bounded subgradients, is presented in the appendix for completeness.

3.3. Implementation of the Dual-Based Algorithm

The dual-based algorithm for end-to-end proportionally fair rate allocation under buffer overflow constraints in random access wireless networks can be summarized as follows:

1. Initialize the iteration count n to zero. If $\langle l, q \rangle = (i, j)$ for some flow l , node i chooses an initial value of $\lambda_{l,q}^0$ such that $0 < \lambda_{l,q}^0 < 1$.
2. Node i passes the value of $\lambda_{l,q}^n$ to the source of the l -th flow. The logarithm of the rates (f_i) are then computed by sources using (11) in $O(I)$ time.
3. Every node that the l -th flow is routed through, obtains the value of f_l^n from the source.
4. After obtaining the $\lambda_{l,q}^n$ -values from nodes within a 2-hop neighborhood, each node computes the access probability values ($p_{l,q}$) according to (10).
5. Each node increments the value of n and computes $\lambda_{l,q}^{n+1}$ by the gradient projection algorithm given by (13) in $O(I)$ time.
6. Steps 2, 3, 4 and 5 are repeated till an appropriate stopping condition is satisfied (based on heuristics or some other criteria, see discussion below).

When flows can arrive and depart in the network, constant step size is the preferred option and in this case there is no stopping criteria, i.e. the nodes continue to run the optimization algorithm without termination. The access probabilities are updated periodically and the source of each flow sets the flow rate as

$$\min \left\{ \log(\mathcal{S}_{l,1}), \min_{2 \leq q \leq k_l} (\log(\mathcal{S}_{l,q}) - \delta_l) \right\}, \quad (16)$$

to maintain the flow rates in the QoS region.

4. PERFORMANCE EVALUATION

For our simulation comparisons, we consider the example shown in figure 1, from references [10], [12]. The nodes are labeled from 1 to 6. The interference model is that each node interferes with the reception at its one-hop neighbors. For example nodes 1 and 3 cause interference at node 2; nodes 6, 5, 2 and 4 cause interference at node 3. Three end-to-end flows, namely $flow_1$, $flow_2$, and $flow_3$ are setup in this network. The source, the sinks, and the path of three flows are shown in table I.

We suppose each flow can tolerate a loss of 45 in every 100,000 packets. Additionally, we suppose each node has a buffer that can store 50 packets for each flow that is routed through it.

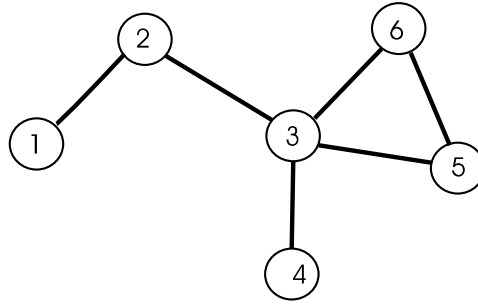


Fig. 1. An ad-hoc wireless network.

Flow	Links (Source, Sink) on the path	Observed flow rates in Matlab
$flow_1$	$\langle 1, 1 \rangle = (6, 5), \langle 1, 2 \rangle = (5, 3), \langle 1, 3 \rangle = (3, 2), \langle 1, 4 \rangle = (2, 1)$	0.0462
$flow_2$	$\langle 2, 1 \rangle = (6, 3), \langle 2, 2 \rangle = (3, 4)$	0.1123
$flow_3$	$\langle 3, 1 \rangle = (1, 2), \langle 3, 2 \rangle = (2, 3), \langle 3, 3 \rangle = (3, 4)$	0.0799

TABLE I: Path of the flows and observed flow rates in a MATLAB simulation of the network shown in figure 1.

This translates to a value of $\rho = 0.86$. For $\rho_1 = \rho_2 = \rho_3 = 0.86$, the globally optimal solutions to the problem defined in (7) was computed using the *fmincon* function in MATLAB. This solution and the solutions given by the dual-based algorithm are presented in table II. More detailed experimentation results can be found in [24], which has been excluded from this paper due to page limitations.

We have used two approaches to the dual-based algorithm outlined earlier. In the first approach we use a constant step size $\alpha^n = 5 \times 10^{-4}$ (cf. (13)), and the logarithm of the minimal achievable rate, ε was set to be -10. Figure 2(a) shows how the flow rates (cf. (16)), converge when the dual-based algorithm with fixed step size is used.

The second approach involves the use diminishing step size. In this case the step size at the n^{th} iteration $\alpha^n = 1/n$. The value of ε is set to be -10. Figure 2(b) illustrates the convergence of the flow rates (cf. (16)), when the dual-based algorithm with diminishing step sizes is used.

If $\rho_1 = \rho_2 = \rho_3 = 1$, we get the optimal solution of $U^* = -7.4897$ using MATLAB's *fmincon* function. This is higher than when $\rho_1 = \rho_2 = \rho_3 = 0.86$, but the buffer overflow will be significantly higher. To demonstrate this, we simulated the network in figure 1 in

MATLAB, using access probabilities obtained for $\rho_1 = \rho_2 = \rho_3 = 1$ and $\rho_1 = \rho_2 = \rho_3 = 0.86$, and ran the simulation for a duration of 5×10^4 time slots, where each node in the network has infinite length buffers (i.e. no packets are dropped in the simulation). For a random instance of the simulation, we plot the queue-lengths as a function of time, for flow 1 at node 5. In the plots, unit of time is a single time-slot of fixed duration. The flow rates observed for $\rho_1 = \rho_2 = \rho_3 = 0.86$ are presented in table I.

Figure 3 demonstrates the queuing performance of our algorithm at node 5. Case I is the plot of queue-length as a function of time when we use the optimal access probabilities without considering buffer overflow constraints i.e. $\rho_1 = \rho_2 = \rho_3 = 1$. Case II shows how the queue-length varies as a function of time, when the optimal access probabilities obtained by setting $\rho_1 = \rho_2 = \rho_3 = 0.86$ is used. We can observe from the plots that in case I, the queue is unstable, whereas, case II shows the queue at node 5 is stable. If a buffer size of 50 was used, then the fraction of packets transmitted that are lost, in case I will be much higher than as compared to case II.

Variables	$p_{6,5,1}$	$p_{5,3,1}$	$p_{3,2,1}$	$p_{2,1,1}$
optimum solutions	0.0881	0.2185	0.1028	0.0657
dual-based solutions with constant step sizes	0.0892	0.2165	0.1078	0.0675
dual-based solutions with diminishing step sizes	0.0882	0.2191	0.1032	0.0661
Variables	$p_{6,3,2}$	$p_{3,4,2}$		
optimum solutions	0.3388	0.1329		
dual-based solutions with constant step sizes	0.3353	0.1419		
dual-based solutions with diminishing step sizes	0.3377	0.1333		
Variables	$p_{1,2,3}$	$p_{2,3,3}$	$p_{3,4,3}$	
optimum solutions	0.1776	0.2949	0.0892	
dual-based solutions with constant step sizes	0.1875	0.2929	0.0903	
dual-based solutions with diminishing step sizes	0.1761	0.2949	0.0893	
Variables	f_1	f_2	f_3	\mathbf{U}^*
optimum solutions	0.0465	0.1143	0.0767	-7.8051
dual-based solutions with constant step sizes	0.0461	0.1109	0.0792	-7.8118
dual-based solutions with diminishing step sizes	0.0464	0.1136	0.0759	-7.8239

TABLE II: The Optimal Results and the Solution given by the Distributed Algorithm

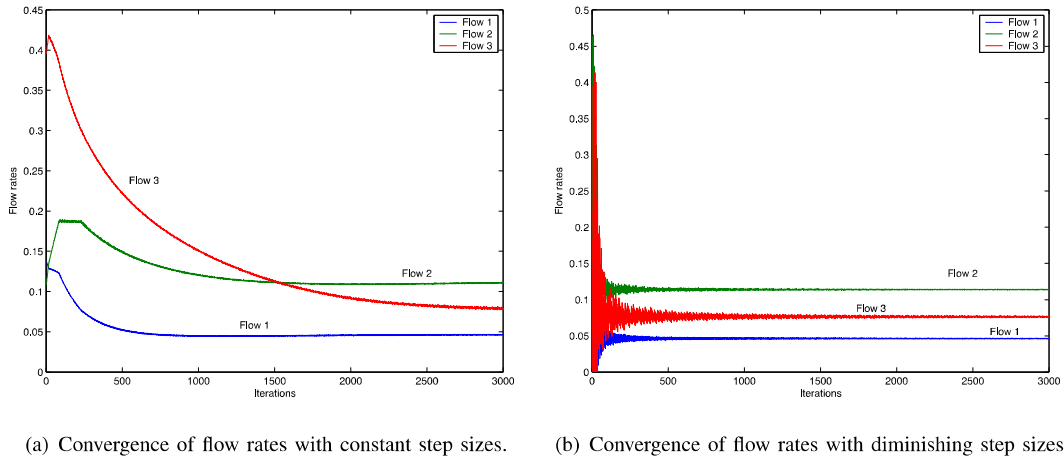


Fig. 2. The convergence of flow rates when dual-based algorithm is used.

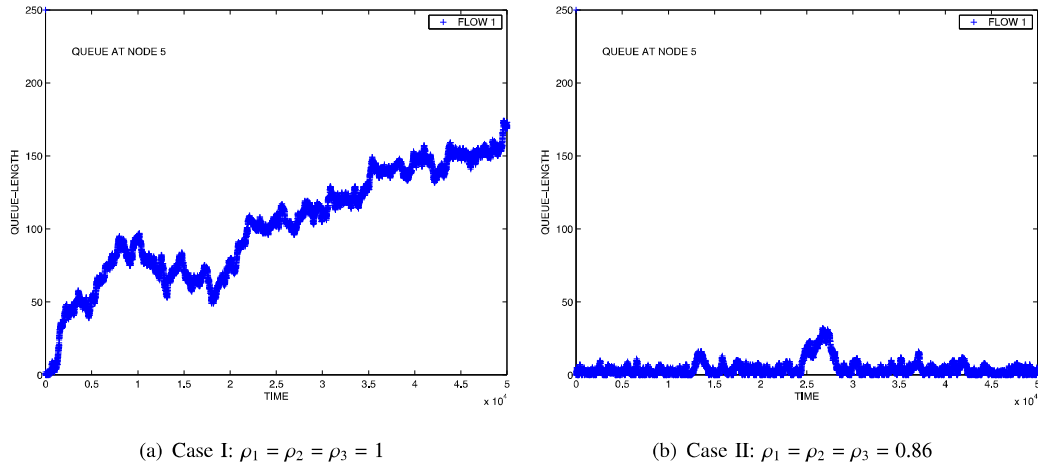


Fig. 3. The queue-length at Node 5.

5. CONCLUSION

In this paper, we proposed a distributed scheme for providing end-to-end proportionally fair flow rates in a slotted-time, multi-hop, random access network with a general network topology, with bounds on the buffer overflow probabilities at each node. After noting that each flow in the network can be viewed as a tandem of discrete-time queues, we converted the constraints on buffer overflow probabilities into appropriate constraints on the link rates, which permitted the reformulation of the original problem into an appropriately posed convex minimization problem under convex constraints. We solved this problem using an appropriately constructed Lagrange function, and discuss its convergence properties. After presenting aspects of distributed implementation of this dual-based approach, we verified the correctness of the approach using an example from the literature.

APPENDIX A: PROOF OF CONVERGENCE OF PROJECTED SUBGRADIENT METHODS WITH CONSTANT STEP SIZE

Suppose $f: \mathcal{R}^n \rightarrow \mathcal{R}$ is a concave function defined over a convex set C , having a non-empty set of maximum points M^* . To maximize f , the projected subgradient method uses the iteration $x_{k+1} = [x_k + h_{k+1}g(x_k)]^+$, where x_k is the k -th iterate, $g(x_k)$ is the subgradient of f at x_k and h_{k+1} is the step size, and for constant step size we have $h_k = h, \forall k$. We assume that norm of subgradients of f is

bounded and therefore, $\|g(x_k)\|_2 \leq G, \forall k$. Also, we define at the k -th iterate, $\bar{x} = \frac{1}{k+1} \sum_{i=0}^k x_i$.

Theorem A.1: For any $x^* \in M^*$, as $k \rightarrow \infty$ one can find a \hat{x} , such that either $f(\hat{x}) = \lim_{k \rightarrow \infty} f(x_k)$ or $f(\hat{x}) = \liminf_{k \rightarrow \infty} f(x_k)$ and $f(x^*) - f(\hat{x}) \leq G^2 h / 2$.

Proof: If $g(x_{k^*}) = 0$ for some k^* , then $f(x_{k^*}) = f(x^*), \forall k \geq k^*$ and we may take $\hat{x} = x^*$. If $g(x_k) \neq 0, \forall k$, then $x_{k+1} = [x_k + hg(x_k)]^+$. Let $z_{k+1} = x_k + h_{k+1}g(x_k)$ (without projection). Observe that

$$\|x_{k+1} - x^*\|_2 \leq \|z_{k+1} - x^*\|_2. \quad (17)$$

This is true as when we project a point onto C , we move closer to every point in C . Now,

$$\|z_{k+1} - x^*\|_2^2 = \|x_k + hg(x_k) - x^*\|_2^2 = \|x_k - x^*\|_2^2 + 2hg(x_k)^T(x_k - x^*) + h^2 \|g(x_k)\|_2^2.$$

From (17), we have

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 + 2hg(x_k)^T(x_k - x^*) + h^2 \|g(x_k)\|_2^2 \quad (18)$$

From the definition of the subgradients for concave functions we have,

$$f(x^*) \leq f(x_k) + g(x_k)^T(x^* - x_k). \quad (19)$$

From (18) and (19), we get the following inequality \square

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 + 2h(f(x_k) - f(x^*)) + h^2 \|g(x_k)\|_2^2. \quad (20)$$

Recursively from (20), we get

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_0 - x^*\|_2^2 + 2h \sum_{i=0}^k (f(x_i) - f(x^*)) + h^2 \sum_{i=0}^k \|g(x_i)\|_2^2. \quad (21)$$

Using $\|x_{k+1} - x^*\|_2^2 \geq 0$, we have,

$$2h \sum_{i=0}^k (f(x^*) - f(x_i)) \leq \|x_0 - x^*\|_2^2 + h^2 \sum_{i=0}^k \|g(x_i)\|_2^2. \quad (22)$$

By property of concave functions, we have,

$$\frac{\sum_{i=0}^k f(x_i)}{k+1} \leq f(\bar{x}), \quad (23)$$

where, $\bar{x} = \frac{1}{k+1} \sum_{i=0}^k x_i$. Thus we have,

$$\sum_{i=0}^k (f(x^*) - f(x_i)) \geq (k+1)(f(x^*) - f(\bar{x})).$$

Combining this with (22), we get the inequality

$$2h(k+1)(f(x^*) - f(\bar{x})) \leq \|x_0 - x^*\|_2^2 + h^2 \sum_{i=0}^k \|g(x_i)\|_2^2. \quad (24)$$

Given that $\|g(x_i)\| \leq G$, for all i , we have,

$$f(x^*) - f(\bar{x}) \leq \frac{\|x_0 - x^*\|_2^2 + h^2(k+1)G^2}{2h(k+1)}. \quad (25)$$

Taking the limit as $k \rightarrow \infty$, we get,

$$f(x^*) - \liminf_{k \rightarrow \infty} (f(\bar{x})) \leq G^2 h / 2. \quad (26)$$

Hence the result. ■

6. REFERENCES

- [1] M. Gast, 802.11 Wireless Networks: The Definitive Guide. Sebastapol, CA: O'Reilly & Associates, 2002.
- [2] O'Hara and A. Petrick, IEEE 802.11 Handbook: A Designer's Companion. Standards Information Network, IEEE Press, 1999.
- [3] T. Nandagopal, T.-E. Kim, X. Gao, , and V. Bharghavan, "Achieving MAC Layer Fairness in Wireless Packet Networks," in ACM Mobicom, 2000, pp. 87–98.
- [4] S. Sharma, K. Gopalan, N. Zhu, P. De, G. Peng, and T. Chiueh, "Implementation Experiences of Bandwidth Guarantees on a Wireless LAN," in ACM/SPIE Multimedia Computing and Networking (MMCN 2002), 2002.
- [5] N. Vaidya, P. Bahl, and S. Gupta, "Distributed fair scheduling in a wireless LAN," in 6th Annual International Conference on Mobile Computing and Networking, 2000.
- [6] F. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," in Journal of the Operational Research Society, vol. 49, 1998.
- [7] S.Kunniyur and R.Srikant, "End-to End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks," in INFOCOM (3), 2000, pp. 1323–1332.

- [8] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556–567, 2000.
- [9] S.Srinivas and R.Srikant, "Throughput and fairness guarantees through maximal scheduling in wireless networks," *Network Optimization and Control*, vol. 2, no. 3, pp. 271–379, 2007.
- [10] X. Wang and K. Kar, "Cross-layer rate control for end-to-end proportional fairness in wireless networks with random access," in *MobiHoc '05: Proceedings of the 6th ACM international symposium on Mobile ad hoc networking and computing*. New York, NY, USA: ACM Press, 2005, pp. 157–168.
- [11] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Utility-optimal random-access control," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 7, pp. 2741–2751, July 2007.
- [12] J.Liu and A.Stolyar, "Distributed queue length based algorithms for optimal end-to-end throughput allocation and stability in multi-hop random access networks," in *Proceedings of the 45th Annual Allerton Conference on Communication, Control, and Computing*, September 2007.
- [13] X. Lin and S. Rasool, "Constant-time distributed scheduling policies for ad hoc wireless networks," *Decision and Control, 2006 45th IEEE Conference on*, pp. 1258–1263, 13-15 Dec. 2006.
- [14] A. Gupta, X. Lin, and R. Srikant, "Low-complexity distributed scheduling algorithms for wireless networks," *INFOCOM 2007. 26th IEEE International Conference on Computer Communications*. IEEE, pp. 1631–1639, May 2007.
- [15] C.Joo and N.Shroff, "Performance of random access scheduling schemes in multi-hop wireless networks," *Signals, Systems and Computers, 2006. ACSSC '06. Fortieth Asilomar Conference on*, pp. 1937–1941, Oct.-Nov. 2006.
- [16] P. Chaporkar, K. Kar, X. Luo, and S. Sarkar, "Throughput and fairness guarantees through maximal scheduling in wireless networks," *Information Theory, IEEE Transactions on*, vol. 54, no. 2, pp. 572–594, Feb. 2008.
- [17] C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *ACM SIGCOMM'94 Conference on Communications Architectures, Protocols and Applications*, 1994, pp. 234–244.
- [18] J. Hsu and P. Burke, "Behavior of tandem buffers with geometric input and markovian output," *IEEE Trans. on Communications*, pp. 358–361, March 1979.
- [19] R. Jackson, "Queueing systems with phase type service," *Operations Research*, vol. 5, pp. 109–120, 1954.
- [20] K. Bharath-Kumar, "Discrete-time queueing systems and their networks," *IEEE Trans. on Communications*, vol. 28, no. 2, pp. 260–263, February 1980.
- [21] D. Bertsekas, *Nonlinear Programming*, Second Edition ed. Athena Scientific, 1999.
- [22] X. Wang and K. Kar, "Distributed algorithms for max-min fair rate allocation in aloha networks," in *Proceedings 42nd Annual Allerton Conference on Communication, Control, and Computing*, October 2003.
- [23] N. Z. Shor, K. C. Kiwiel, and A. Ruszcaynski, *Minimization methods for non-differentiable functions*. New York, NY, USA: Springer-Verlag New York, Inc., 1985.
- [24] N.Singh and R. Sreenivas, "Enforcing end-to-end proportional fairness with bounded buffer overflow probabilities," *Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, 1308 West Main Street, Urbana, IL 61801.*, Technical Report UILU-ENG-08-2211, August 2008.