

Utility-Optimal Random-Access Control

Jang-Won Lee, *Member, IEEE*, Mung Chiang, *Member, IEEE*, and A. Robert Calderbank, *Fellow, IEEE*

Abstract—This paper designs medium access control (MAC) protocols for wireless networks through the network utility maximization (NUM) framework. A network-wide utility maximization problem is formulated, using a collision/persistence-probabilistic model and aligning selfish utility with total social welfare. By adjusting the parameters in the utility objective functions of the NUM problem, we can also control the tradeoff between efficiency and fairness of radio resource allocation. We develop two distributed algorithms to solve the utility-optimal random-access control problem, which lead to random access protocols that have slightly more message passing overhead than the current exponential-backoff protocols, but significant potential for efficiency and fairness improvement. We provide readily-verifiable sufficient conditions under which convergence of the proposed algorithms to a global optimality of network utility can be guaranteed, and numerical experiments that illustrate the value of the NUM approach to the complexity-performance tradeoff in MAC design.

Index Terms—Wireless network, medium access control (MAC), mathematical programming/optimization, network utility maximization, network control by pricing.

I. INTRODUCTION

A. Motivation

IF contentions among transmissions on different links in wireless networks are not appropriately controlled, a large number of collisions may occur, resulting in waste of resources such as bandwidth and energy, as well as loss of system efficiency and degradation of resource sharing fairness. The important issue of contention resolution in the wireless medium access control (MAC) protocol is studied in this paper. To rigorously and systematically explore alternative designs of MAC protocols, we need a tool that incorporates both efficiency and fairness metrics in the most flexible way. In this paper we use the framework of Network Utility Maximization (NUM), which provides a rigorous method for analyzing MAC performance and designing MAC protocols. This approach complements the well-established queuing-theoretic stochastic

analysis, with a focus on attaining optimality with respect to general utility objectives for long-lived flows.

In wireless networks, especially in the multi-hop ones, there may not be any central controller, thus we focus on random access MAC protocols that can be implemented in a distributed way. Currently, the Distributed Coordination Function (DCF) is the standardized MAC protocol in IEEE 802.11 [2]. However, it has been concluded by many researchers that DCF and its Binary Exponential Backoff (BEB) mechanism for contention resolution can be inefficient and unfair in face of location-dependent contentions (e.g., [3]). Various new algorithms have been developed to tackle these issues (e.g., [3]–[9]).

Due to the inadequate feedback mechanism in the BEB protocol, neither convergence nor social welfare optimality can be assured [1], [10]. This motivates the following question: what kind of new distributed algorithms will be provably convergent to the global optimum of total network utility? After formulating a probabilistic-modeled NUM problem for wireless MAC, we develop optimal algorithms to solve the NUM problem, and these algorithms are then turned into random access MAC protocols. Through this design approach, optimality with respect to prescribed user utilities, which in turn determine protocol efficiency and fairness, is guaranteed.

B. Review: Basic NUM and Congestion Control

Since the seminal paper by Kelly *et al.* [11] in 1998, the basic NUM problem has been extensively studied for rate allocation and congestion control in wired networks (e.g., [12]–[15]). Consider a communication network with L logical links, each with a fixed capacity of c_l bps, and S sources (i.e., end users), each transmitting at a source rate of x_s bps. Each source s emits one flow, using a fixed set $L(s)$ of links in its path, and has a utility function $U_s(x_s)$. Each link l is shared by a set $S(l)$ of sources. NUM, in its basic version for wired networks, is the following problem of maximizing the network utility $\sum_s U_s(x_s)$, over the source rates \mathbf{x} , subject to linear flow constraints $\sum_{s \in S(l)} x_s \leq c_l$ for all links l :

$$\begin{aligned} & \text{maximize} && \sum_s U_s(x_s) \\ & \text{subject to} && \sum_{s \in S(l)} x_s \leq c_l, \quad \forall l, \\ & && \mathbf{x}^{\min} \preceq \mathbf{x} \preceq \mathbf{x}^{\max}. \end{aligned} \quad (1)$$

Making the standard assumption on the concavity of the utility functions, problem (1) is a simple concave maximization of decoupled terms under linear constraints. Recent work in congestion control literature (e.g., in [15]) has shown that TCP congestion control protocols can be thus reverse engineered: TCP variants are implicitly solving the basic NUM (1) for different utilities using Lagrange dual variables as link congestion prices.

Manuscript received December 6, 2005; revised April 4, 2006 and June 23, 2006; accepted July 3, 2006. The associate editor coordinating the review of this paper and approving it for publication was M. Zorzi. This work was supported by Yonsei University research fund of 2005 and NSF Grants CCF-0430487, CCF-0440443, CNS-0417607, CNS-0427677, and CCF-0448012. The earlier version of this paper was presented at IEEE Infocom 2006 [1].

J.-W. Lee is with the Department of Electrical and Electronic Engineering, Yonsei University, 134 Shinchon-dong, Seodaemun-gu, Seoul, Korea 120-749 (email: jangwon@yonsei.ac.kr).

M. Chiang is with the Electrical Engineering Department and also affiliated with the Program in Applied and Computational Mathematics and with the Computer Science Department at Princeton University, Princeton, NJ 08544, USA (email: chiangm@princeton.edu).

A. R. Calderbank is with the Electrical Engineering Department, the Mathematics Department, and the Program in Applied and Computational Mathematics at Princeton University, Princeton, NJ 08544, USA (email: calderbk@princeton.edu).

Digital Object Identifier 10.1109/TWC.2007.05991.

In the NUM framework, there are at least two interpretations of each source's utility function. It can be interpreted as the level of satisfaction attained by a user as a function of resource allocation. Each user may have a different utility function depending on its type of service. By maximizing the network utility (*i.e.*, the sum of all user utilities), we maximize the social welfare of the system. Efficiency of resource allocation algorithms can be measured by the achieved network utility. Utility functions can also be interpreted as the 'knobs' to control the tradeoff between efficiency and fairness. Different shapes of utility functions lead to different types of fairness. For example, a family of utility functions parameterized by $\alpha \geq 0$ is proposed in [13]:

$$U^\alpha(x) = \begin{cases} (1 - \alpha)^{-1} x^{1-\alpha}, & \text{if } \alpha \neq 1 \\ \log x, & \text{otherwise} \end{cases} \quad (2)$$

In particular, if we set $\alpha = 0$, NUM reduces to system throughput maximization. If $\alpha = 1$, proportional fairness among competing users is attained; if $\alpha = 2$, then harmonic mean fairness; and if $\alpha \rightarrow \infty$, then max-min fairness.

Recently, NUM has also been used to develop MAC protocols in wireless multi-hop networks in [3], [6]–[9]. The paper by Nandagopal *et al.* [3] is the first one that extends Kelly's wired NUM model to wireless MAC design by proposing the contention graph construction. In [8], [9], max-min fairness (*i.e.*, $\alpha \rightarrow \infty$ in (2)) and proportional fairness (*i.e.*, $\alpha = 1$ in (2)) are considered, respectively. The NUM framework for more general utility functions are studied in [3], [6], [7] through deterministic approximations. Extensive discussion of and comparison with all related work can be found in Sections I.C and IV. Also, we refer readers to [16] for a survey for the NUM framework and layering as optimization decomposition.

C. Summary of Results

We formulate a new NUM problem in terms of the persistence probabilities of each node and link, and then develop utility-optimal random-access protocols. It turns out that, except for the special case of logarithmic utility, the resulting NUM problem is non-convex and non-separable. Convexity and separability properties are essential to develop a distributed and optimal algorithm for a NUM problem. Despite these difficulties, we show how to develop distributed MAC protocols in which each link adjusts its own persistence probability based on only its local information and limited message passing. We also provide readily-verifiable sufficient conditions on utility functions' curvatures under which convergence of the proposed algorithms to a global optimality can be proved.

Compared with all the previous studies on MAC protocols by using NUM, our approach has following distinguishing features. In contrast to [3], [6], [7] where contention is modeled using 'deterministic approximation' and well-developed algorithms for NUM in wired networks are used, we directly model contentions among transmissions through a collision and persistence probability model. It will be shown that the deterministic approximation approach [3], [6], [7] cannot accurately model contention in the random access protocol. In [3], a distributed algorithm considering collision probabilities

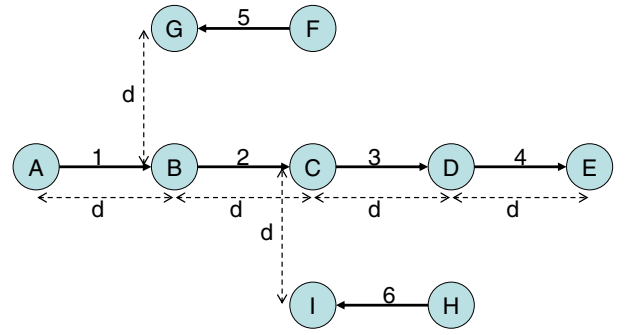


Fig. 1. A network topology graph.

that approximately solves the NUM problem is proposed. However, it is not rigorously shown how closely it can approximate the optimal solution. In contrast to [8], [9] where only max-min fairness and proportional fairness (*i.e.*, only special types of utility functions such as log utility functions) are considered, we consider general types of utility functions, which facilitates the provisioning of different types of services and different efficiency-fairness tradeoffs.

After developing new contention-based MAC protocols based on NUM distributed solutions and a new idea of 'contention pricing', we compare the performance of these algorithms with the existing MAC standards and recent proposals in the literature.

- Our random access MAC protocols (proposed in Section III).
- Random access protocol based on the known deterministic approximation.
- Random access BEB protocol in the current IEEE 802.11 standard.

We show that our protocol provides better efficiency-fairness tradeoffs than both the deterministic approximation and BEB protocols.

II. SYSTEM MODEL AND NOTATION

We consider an ad-hoc network represented by a directed graph $G(V, E)$, *e.g.*, as in Figure 1, where V is the set of nodes and E is the set of logical links. We define $L_{out}(n)$ as the set of outgoing links from node n , $L_{in}(n)$ as the set of incoming links to node n , Tx_l as the transmitter node of link l , and Rx_l as the receiver node of link l . We assume that the transmission of a link is interfered from the transmission of another link, if the distance between the receiver node of the former and the transmitter node of the latter is less than some threshold (*e.g.*, d). We define $N_{to}^I(l)$ as the set of nodes whose transmissions¹ cause interference to the receiver of link l , excluding the transmitter node of link l , (*i.e.*, Tx_l), and $L_{from}^I(n)$ as the set of links whose transmissions get interfered from the transmission of node n , excluding outgoing links from node n (*i.e.*, $l \in L_{out}(n)$). Hence, if the transmitter of link l and a node in set $N_{to}^I(l)$ transmit data simultaneously, the transmission of link l fails. If node n and the transmitter

¹We say that node n transmits data if one of its outgoing links transmits data.

of a link l in set $L_{from}^I(n)$ transmit data simultaneously, the transmission of link l also fails.

Each node has a contention resolution protocol based on the transmission persistence probability. Each node n transmits data with a probability P^n . When it determines to transmit data, it chooses one of its outgoing links with a probability $q_l, \forall l \in L_{out}(n)$, such that $\sum_{l \in L_{out}(n)} q_l = 1$, and transmits data on the chosen link. Hence, there is no collision among links that have the same transmitter node. Consequently, link $l, l \in L_{out}(n)$, transmits data with a probability $p_l = P^n q_l$ such that $\sum_{l \in L_{out}(n)} p_l = P^n, \forall n$. We call q_l and p_l conditional persistence probability and persistence probability of link l , respectively. Each node adjusts its persistence probability and persistence probabilities of its outgoing links based on its adopted algorithm to be developed in the next section.

III. DESIGN OF UTILITY-OPTIMAL RANDOM-ACCESS MAC PROTOCOL

In this section, we propose a new NUM formulation and develop enhanced MAC protocols as distributed solutions to such a network-wide maximization problem. We will show that limited message passing of ‘contention prices’ can provide the optimal coordination to align the selfish non-cooperative nature of the BEB protocol into a new MAC protocol that always converges, and to the global optimum of network utility.

Each link has a utility function $U_l(x_l)$, an increasing non-linear function of its average data rate x_l , which is in turn a function of persistence probabilities. We would like to discover how each link should adjust its own persistence probability so as to globally maximize the total network utility. The average capacity $C_l(\mathbf{p}, \mathbf{P})$ on link l is obtained as:

$$C_l(\mathbf{p}, \mathbf{P}) = c_l p_l \prod_{k \in N_{to}^I(l)} (1 - P^k), \quad \forall l, \quad (3)$$

where c_l is a fixed data rate of link l . The NUM problem can be formulated as follows, over the optimization variables of $(\mathbf{x}, \mathbf{p}, \mathbf{P})$:²

$$\begin{aligned} & \text{maximize} && \sum_l U_l(x_l) \\ & \text{subject to} && x_l \leq C_l(\mathbf{p}, \mathbf{P}) (= c_l p_l \prod_{k \in N_{to}^I(l)} (1 - P^k)), \quad \forall l \\ & && \sum_{l \in L_{out}(n)} p_l = P^n, \quad \forall n \\ & && x_l^{min} \leq x_l \leq x_l^{max}, \quad \forall l \\ & && 0 \leq P^n \leq 1, \quad \forall n, \quad 0 \leq p_l \leq 1, \quad \forall l \end{aligned} \quad (4)$$

where x_l is the average data rate of link l , and x_l^{min} and x_l^{max} are the average data rate, constraints on minimum and maximum average data rates of link l , respectively. The first constraint states that the average data rate of a link must be less than or equal to the average capacity of the link. The second constraint states that the sum of persistence probabilities of the outgoing links of a node must be equal to the persistence probability of the link.

The objective of this problem is to obtain the optimal data rate \mathbf{x} and the optimal persistence probabilities \mathbf{p} for links, and \mathbf{P} for nodes so as to maximize the network utility. For log

utilities, problem (4) is readily shown to be a decomposable convex optimization [8]. However, in general, it is non-convex and non-separable, thus difficult to be distributively solved for global optimality. We show that under a readily-verifiable sufficient condition on curvatures of the utility functions, it can be turned into a convex and separable optimization problem.

We first take the log of both sides of the first constraint in problem (4) and a log change of variables and constants: $x'_l = \log x_l$, $x'_l{}^{max} = \log x_l^{max}$, $x'_l{}^{min} = \log x_l^{min}$, $U'_l(x'_l) = U_l(e^{x'_l})$, and $c'_l = \log c_l$. This reformulation turns the problem into:

$$\begin{aligned} & \text{maximize} && \sum_{l \in L} U'_l(x'_l) \\ & \text{subject to} && c'_l + \log p_l + \sum_{k \in N_{to}^I(l)} \log(1 - P^k) - x'_l \geq 0, \quad \forall l \\ & && \sum_{l \in L_{out}(n)} p_l = P^n, \quad \forall n \\ & && x'_l{}^{min} \leq x'_l \leq x'_l{}^{max}, \quad \forall l \\ & && 0 \leq P^n \leq 1, \quad \forall n, \quad 0 \leq p_l \leq 1, \quad \forall l. \end{aligned} \quad (5)$$

Note that problem (5) is now separable but still may not be a convex optimization problem, since the objective $U'_l(x'_l)$ may not be a (strictly) concave function, even though $U_l(x_l)$ is a (strictly) concave function. However, the following lemma provides a sufficient condition for its concavity. Define

$$g_l(x_l) = \frac{d^2 U_l(x_l)}{dx_l^2} x_l + \frac{dU_l(x_l)}{dx_l}.$$

Lemma 1: If $g_l(x_l) < 0$, $U'_l(x'_l)$ is a strictly concave function of x'_l .

Proof: Since $x_l = e^{x'_l}$,

$$\begin{aligned} \frac{d^2 U'_l(x'_l)}{dx'^2_l} &= \frac{d^2 U_l^x(x_l)}{dx_l^2} \left(\frac{dx_l}{dx'_l} \right)^2 + \frac{dU_l(x_l)}{dx_l} \frac{d^2 x_l}{dx'^2_l} \\ &= e^{x'_l} \left(\frac{d^2 U_l(x_l)}{dx_l^2} x_l + \frac{dU_l(x_l)}{dx_l} \right) \\ &= e^{x'_l} g_l(x_l). \end{aligned}$$

Hence, if $g_l(x_l) < 0$, $U'_l(x'_l)$ is a strictly concave function of x'_l . ■

Remark 1: The condition of $g_l(x_l) < 0$ is equivalent to:

$$\frac{d^2 U_l(x_l)}{dx_l^2} < -\frac{dU_l(x_l)}{dx_l}.$$

Since utility functions are increasing, $\frac{dU_l(x_l)}{dx_l}$ has a positive value. The above inequality states that the utility function needs to be not just strictly concave (i.e., $\frac{d^2 U_s(x_l)}{dx_l^2} < 0$), but with a curvature that is bounded away from 0 by as much as $\frac{dU_l(x_l)}{dx_l}$, i.e., the application needs represented by the utility functions must be elastic enough.

Lemma 2: For the utility functions (2) parameterized by $\alpha \geq 0$, $g_l(x_l) > 0$ if $\alpha < 1$, $g_l(x_l) = 0$ if $\alpha = 1$, and $g_l(x_l) < 0$ if $\alpha > 1$.

Proof: First, consider the case $\alpha = 1$. Then, since $U_l(x) = \log(x_l)$, we can easily show that $g_l(x_l) = \frac{d^2 U_l(x_l)}{dx_l^2} x_l + \frac{dU_l(x_l)}{dx_l} = 0$. We now consider the case $\alpha \neq 1$. Then, since $U_l(x_l) = (1 - \alpha)^{-1} x_l^{1-\alpha}$,

$$\begin{aligned} g_l(x_l) &= \frac{d^2 U_l(x_l)}{dx_l^2} x_l + \frac{dU_l(x_l)}{dx_l} = -\alpha x_l^{-(1+\alpha)} x_l + x_l^{-\alpha} \\ &= (1 - \alpha) x_l^{-\alpha}. \end{aligned}$$

²Similar models have recently been considered in [8], [9] for a restricted class of utility functions, which indeed can be recovered as special cases of the general results in this section.

Hence, if $\alpha > 1$, $g_l(x_l) < 0$ and if $\alpha < 1$, $g_l(x_l) > 0$. ■

Hence, in this type of utility functions, if $\alpha > 1$, $U'_l(x'_l)$ becomes a strictly concave function. Throughout this section, we will assume that the condition in Lemma 1 is satisfied.

We now solve problem (5) by using a dual decomposition approach [17]. We first write down the Lagrangian function associated with problem (5) as follows, where λ_l is the Lagrange multiplier on link l with an interpretation of ‘contention price’:

$$\begin{aligned}
L(\boldsymbol{\lambda}, \mathbf{x}', \mathbf{p}, \mathbf{P}) &= \sum_{l \in L} U'_l(x'_l) \\
&+ \sum_{l \in L} \{ \lambda_l (c'_l + \log p_l + \sum_{k \in N_{to}^I(l)} \log(1 - P^k) - x'_l) \} \\
&= \sum_{l \in L} \{ U'_l(x'_l) - \lambda_l x'_l \} + \sum_{l \in L} \lambda_l \log p_l \\
&+ \sum_{l \in L} \lambda_l \sum_{k \in N_{to}^I(l)} \log(1 - P^k) + \sum_{l \in L} \lambda_l c'_l \\
&= \sum_{n \in N} \sum_{l \in L_{out}(n)} \{ U'_l(x'_l) - \lambda_l x'_l \} + \sum_{n \in N} \sum_{l \in L_{out}(n)} \lambda_l \log p_l \\
&+ \sum_{n \in N} \sum_{m \in L_{from}^I(n)} \lambda_m \log(1 - P^n) + \sum_{l \in L} \lambda_l c'_l. \quad (6)
\end{aligned}$$

Note that in this Lagrangian, we do not need to relax the second constraint in problem (5). By definition, the Lagrange dual function is

$$\begin{aligned}
Q(\boldsymbol{\lambda}) &= \max_{\substack{\sum_{l \in L_{out}(n)} p_l = P^n, \forall n \\ \mathbf{x}'^{min} \preceq \mathbf{x}' \preceq \mathbf{x}'^{max} \\ \mathbf{0} \preceq \mathbf{p} \preceq \mathbf{1} \\ \mathbf{0} \preceq \mathbf{P} \preceq \mathbf{1}}} L(\boldsymbol{\lambda}, \mathbf{x}', \mathbf{p}, \mathbf{P}). \quad (7)
\end{aligned}$$

The dual problem is thus formulated as

$$\min_{\boldsymbol{\lambda} \succeq \mathbf{0}} Q(\boldsymbol{\lambda}). \quad (8)$$

To solve the dual problem, we first consider problem (7). This maximization of the Lagrangian over $(\mathbf{x}', \mathbf{p}, \mathbf{P})$ can be conducted in parallel at each node n :

$$\max_{x'_l \leq x'_l \leq x'_l} \{ U'_l(x'_l) - \lambda_l x'_l \}, \quad \forall l \in L_{out}(n) \quad (9)$$

and

$$\begin{aligned}
&\text{maximize} \quad \sum_{l \in L_{out}(n)} \lambda_l \log p_l + \sum_{k \in L_{from}^I(n)} \lambda_k \log(1 - P^n) \\
&\text{subject to} \quad \sum_{l \in L_{out}(n)} p_l = P^n \\
&\quad 0 \leq p_l \leq 1, \quad \forall l \in L_{out}(n) \\
&\quad 0 \leq P^n \leq 1. \quad (10)
\end{aligned}$$

To obtain the solutions of problem (10), we first define k_n as

$$k_n = \sum_{l \in L_{out}(n)} \lambda_l + \sum_{l \in L_{from}^I(n)} \lambda_l, \quad \forall n.$$

Then, for a given $\boldsymbol{\lambda}$, the $\mathbf{P}(\boldsymbol{\lambda})$ and $\mathbf{p}(\boldsymbol{\lambda})$ that maximize problem (10) are obtained in closed-forms as³ (see Appendix)

$$P^n(\boldsymbol{\lambda}) = \begin{cases} \frac{\sum_{l \in L_{out}(n)} \lambda_l}{\sum_{l \in L_{out}(n)} \lambda_l + \sum_{l \in L_{from}^I(n)} \lambda_l}, & \text{if } k_n \neq 0 \\ \frac{|L_{out}(n)|}{|L_{out}(n)| + |L_{from}^I(n)|}, & \text{if } k_n = 0 \end{cases}, \quad \forall n$$

and

$$p_l(\boldsymbol{\lambda}) = \begin{cases} \frac{\lambda_l}{\sum_{l \in L_{out}(n)} \lambda_l + \sum_{l \in L_{from}^I(n)} \lambda_l}, & \text{if } k_n \neq 0 \\ \frac{1}{|L_{out}(n)| + |L_{from}^I(n)|}, & \text{if } k_n = 0 \end{cases}, \quad \forall l \in L_{out}(n), \forall n.$$

We can now solve the dual problem (8) by using a subgradient projection algorithm⁴ at each link l , i.e., at each node n such that $l \in L_{out}(n)$, through the following iterations indexed by t :

$$\begin{aligned}
\lambda_l(t+1) &= [\lambda_l(t) - \alpha(t) (c'_l + \log p_l(t) \\
&+ \sum_{k \in N_{to}^I(l)} \log(1 - P^k(t)) - x'_l(t))]^+, \\
&\quad \forall l \in L_{out}(n), \quad \forall n, \quad (11)
\end{aligned}$$

where $\alpha(t)$ is a step size and $x'_l(t)$, and $p_l(t)$ and $P^k(t)$ are the solutions to problems (9) and (10), respectively, with $\boldsymbol{\lambda}(t)$. Hence, problems (9), (10), and (11) can be solved in each node n in a distributed way with only local information, i.e., $\lambda_l(t)$, $\forall l \in L_{from}^I(n) \cup L_{out}(n)$ and $P^m(t)$, $\forall m \in N_{to}^I(l)$, $\forall l \in L_{out}(n)$. The proposed MAC protocol can be implemented through the following algorithm with very simple local computational steps, and a limited amount of explicit message passing in Step 3.2, which we will later remove through additional improvements and heuristics.

Distributed Algorithm 1 for Utility-Optimal MAC Protocol

- 1:** Each node n constructs its *local* interference graph to obtain sets $L_{out}(n)$, $L_{in}(n)$, $L_{from}^I(n)$, and $N_{to}^I(l)$, $\forall l \in L_{out}(n)$.
- 2:** Each node n sets $t = 0$, $\lambda_l(1) = 1$, $\forall l \in L_{out}(n)$, $P^n(1) = \frac{|L_{out}(n)|}{|L_{out}(n)| + |L_{from}^I(n)|}$, and $p_l(1) = \frac{1}{|L_{out}(n)| + |L_{from}^I(n)|}$, $\forall l \in L_{out}(n)$.
- 3:** *Locally* at each node n , iterate through t ⁵:
 - 3.1:** Set $t \leftarrow t + 1$.
 - 3.2:** Inform $\lambda_l(t)$ to all nodes in $N_{to}^I(l)$, $\forall l \in L_{out}(n)$, and $P^n(t)$ to Tx_l , $\forall l \in L_{from}^I(n)$.
 - 3.3:** Set $k_n(t) = \sum_{l \in L_{out}(n)} \lambda_l(t) + \sum_{l \in L_{from}^I(n)} \lambda_l(t)$ and $\alpha(t) = \frac{1}{t}$.

³If $k_n = 0$ then $P^n(\boldsymbol{\lambda})$ and $p_l(\boldsymbol{\lambda})$, $\forall l \in L_{out}(n)$, can be any feasible solutions that satisfy the constraints of problem (10).

⁴Since the solution to problem (10) may not be unique, $Q(\boldsymbol{\lambda})$ is not everywhere differentiable. Hence, we use a subgradient projection algorithm. A subgradient of a (possibly nondifferentiable) function $f: \mathbf{R}^n \rightarrow \mathbf{R}$ is any vector \mathbf{g} such that for any \mathbf{x} in the domain of f , we have $f(\mathbf{y}) - f(\mathbf{x}) \geq \mathbf{g}^T(\mathbf{y} - \mathbf{x})$ for all \mathbf{y} in the domain of f .

⁵Note that we do not put any termination conditions of the algorithm. Hence, the algorithm will be performed indefinitely, while converging to the optimal persistence probabilities. However, we can easily add termination conditions of the algorithm such as Heuristic 2 in Remark 5.

3.4: Solve the following problems to obtain $P^n(t+1)$, and $x'_l(t+1)$, $p_l(t+1)$, and $\lambda_l(t+1)$, $\forall l \in L_{out}(n)$:

$$P^n(t+1) = \begin{cases} \frac{\sum_{l \in L_{out}(n)} \lambda_l(t)}{\sum_{l \in L_{out}(n)} \lambda_l(t) + \sum_{l \in L_{from}^I(n)} \lambda_l(t)}, & \text{if } k_n(t) \neq 0 \\ \frac{|L_{out}(n)|}{|L_{out}(n)| + |L_{from}^I(n)|}, & \text{if } k_n(t) = 0 \end{cases},$$

$$p_l(t+1) = \begin{cases} \frac{\lambda_l(t)}{\sum_{l \in L_{out}(n)} \lambda_l(t) + \sum_{l \in L_{from}^I(n)} \lambda_l(t)}, & \text{if } k_n(t) \neq 0 \\ \frac{1}{|L_{out}(n)| + |L_{from}^I(n)|}, & \text{if } k_n(t) = 0 \end{cases},$$

$$x'_l(t+1) = \underset{x'_{l \min} \leq x' \leq x'_{l \max}}{\operatorname{argmax}} \{U'_l(x'_l) - \lambda_l(t)x'_l\},$$

and

$$\lambda_l(t+1) = [\lambda_l(t) - \alpha(t)(c'_l + \log p_l(t) + \sum_{m \in N_{to}^I(l)} \log(1 - P^m(t)) - x'_l(t))].$$

3.5: Each node n sets the conditional persistence probability of each of its outgoing links $q_l(t) = p_l(t)/P^n(t)$, $\forall l \in L_{out}(n)$.

3.6: Each node n decides if it will transmit data with a probability $P^n(t)$. If it decides to transmit data, it chooses to transmit on one of its outgoing links with a probability $q_l(t)$, $\forall l \in L_{out}(n)$.

Remark 2: Note that the above algorithm is conducted at each node n to calculate P^n , p_l , λ_l , and x'_l for its *outgoing* link l (i.e., $\forall l \in L_{out}(n)$). Hence, the above algorithm is conducted at the *transmitter node* of each link. If we assume two nodes within interference range can communicate with each other (e.g., if nodes within distance $2d$ in Fig. 1 can establish a communication link), each node requires information from nodes within two-hop distance from it. To calculate P^n and p_l for its outgoing link l (i.e., $\forall l \in L_{out}(n)$), node n needs λ_m from the transmitter node Tx_m of link m that is interfered from the transmission of node n (i.e., from Tx_m , $\forall m \in L_{from}^I(n)$). Note that Tx_m is within two-hop distance from node n . For example, in Fig. 1, node D needs λ_2 from the transmitter of link 2 (node B), since link 2 is in set $L_{from}^I(D)$, and node B is at two-hop distance from node D. Also, to calculate λ_l for its outgoing link l (i.e., $\forall l \in L_{out}(n)$), node n needs P^m from node m from whose transmission its outgoing link l is interfered (i.e., from $m \in N_{to}^I(l)$, $\forall l \in L_{out}(n)$). For example, in Fig. 1, node B needs P^D from node D, since its outgoing link, link 2 is interfered from the transmission of node D, i.e., $D \in N_{to}^I(2)$ and $2 \in L_{out}(B)$, and node D is at two-hop distance from node B.

Alternatively, if λ_l and x'_l for each link l are calculated at its *receiver* node Rx_l instead of its transmitter node Tx_l , it is possible to implement an algorithm to solve (5) in which each node requires information from nodes only within *one-hop* distance.

Distributed Algorithm 2 for Utility-Optimal MAC Protocol

1: Each node n constructs its *local* interference graph to obtain sets $L_{out}(n)$, $L_{in}(n)$, $L_{from}^I(n)$, and $N_{to}^I(l)$, $\forall l \in L_{out}(n)$.

2: Each node n sets $t = 0$, $\lambda_l(1) = 1$, $\forall l \in L_{in}(n)$, $P^n(1) = \frac{|L_{out}(n)|}{|L_{out}(n)| + |L_{from}^I(n)|}$, and $p_l(1) = \frac{1}{|L_{out}(n)| + |L_{from}^I(n)|}$, $\forall l \in L_{out}(n)$.

3: *Locally* at each node n , iterate through t :

3.1: Set $t \leftarrow t + 1$.

3.2: Inform $\lambda_l(t)$ to node k , $\forall k \in N_{to}^I(l)$ and node Tx_l , $\forall l \in L_{in}(n)$, $P^n(t)$ to Tx_l , $\forall l \in L_{from}^I(n)$, and $p_l(t)$, $\forall l \in L_{out}(n)$ to each node Rx_l .

3.3: Set $k_n(t) = \sum_{l \in L_{out}(n)} \lambda_l(t) + \sum_{l \in L_{from}^I(n)} \lambda_l(t)$ and $\alpha(t) = \frac{1}{t}$.

3.4: Solve the following problems to obtain $P^n(t+1)$ and $p_l(t+1)$, $\forall l \in L_{out}(n)$, and $x'_l(t+1)$ and $\lambda_l(t+1)$, $\forall l \in L_{in}(n)$:

$$P^n(t+1) = \begin{cases} \frac{\sum_{l \in L_{out}(n)} \lambda_l(t)}{\sum_{l \in L_{out}(n)} \lambda_l(t) + \sum_{l \in L_{from}^I(n)} \lambda_l(t)}, & \text{if } k_n(t) \neq 0 \\ \frac{|L_{out}(n)|}{|L_{out}(n)| + |L_{from}^I(n)|}, & \text{if } k_n(t) = 0 \end{cases},$$

$$p_l(t+1) = \begin{cases} \frac{\lambda_l(t)}{\sum_{l \in L_{out}(n)} \lambda_l(t) + \sum_{l \in L_{from}^I(n)} \lambda_l(t)}, & \text{if } k_n(t) \neq 0 \\ \frac{1}{|L_{out}(n)| + |L_{from}^I(n)|}, & \text{if } k_n(t) = 0 \end{cases},$$

$$x'_l(t+1) = \underset{x'_{l \min} \leq x' \leq x'_{l \max}}{\operatorname{argmax}} \{U'_l(x'_l) - \lambda_l(t)x'_l\},$$

and

$$\lambda_l(t+1) = [\lambda_l(t) - \alpha(t)(c'_l + \log p_l(t) + \sum_{m \in N_{to}^I(l)} \log(1 - P^m(t)) - x'_l(t))].$$

3.5: Each node n sets the conditional persistence probability of each of its outgoing links $q_l(t) = p_l(t)/P^n(t)$, $\forall l \in L_{out}(n)$.

3.6: Each node n decides if it will transmit data with a probability $P^n(t)$. If it decides to transmit data, it chooses to transmit on one of its outgoing links with a probability $q_l(t)$, $\forall l \in L_{out}(n)$.

Remark 3: In Algorithm 2, all message passing can be done within one-hop distance, at the expense of additional message passing of λ_l and p_l between transmitter and receiver nodes of link l . Steps in Algorithm 2 are conducted at each node n to calculate P^n and p_l for its outgoing link l (i.e., $\forall l \in L_{out}(n)$), and λ_m and x'_m for its *incoming* link m (i.e., $\forall m \in L_{in}(n)$). Hence, P^n of the transmitter node of link l and p_l for link l are calculated at the transmitter node of link l (i.e., Tx_l), and λ_l and x'_l for link l are calculated at the *receiver node* of link l (i.e., Rx_l). To calculate P^n and p_l for its outgoing link l (i.e., $\forall l \in L_{out}(n)$), each node n now needs λ_m from the receiver node Rx_m of link m that is interfered from the

transmission of node n (i.e., from Rx_m , $\forall m \in L_{from}^I(n)$) and also λ_l from the receiver node Rx_l of its outgoing link l (i.e., from Rx_l , $\forall l \in L_{out}(n)$). These Rx_m and Rx_l are at one-hop distance from node n . For example, in Figure 1, node D needs λ_2 from the receiver node of link 2 (node C), since link 2 is in set $L_{from}^I(D)$, and node C is at one-hop distance from node D.

Similarly, to calculate λ_l for its incoming link l (i.e., $\forall l \in L_{in}(n)$), node n needs P^m from node m from whose transmission its incoming link l is interfered (i.e., from $\forall m \in N_{to}^I(l)$, $\forall l \in L_{in}(n)$), and also p_l from the transmitter node Tx_l of its incoming link l (i.e., from Tx_l , $\forall l \in L_{in}(n)$), and they are at one-hop distance from node n . For example, in Figure 1, node C, which is the receiver node of link 2, needs P^D from node D, since its incoming link, i.e., link 2, is interfered from the transmission of node D, i.e., $D \in N_{to}^I(2)$, $2 \in L_{in}(C)$, and node D is at one-hop distance from node C.

Remark 4: The number of message passing required in each of the above two algorithms depends on the network topology. The average numbers of message passing in each iteration for Algorithm 1 and Algorithm 2, M_1 and M_2 , are obtained as

$$M_1 = H \sum_n \left\{ |L_{from}^I(n)| + \sum_{l \in L_{out}(n)} |N_{to}^I(l)| \right\},$$

where H is the average number of hops that each message traverses, $1 \leq H \leq 2$, and

$$M_2 = \sum_n \left\{ |L_{from}^I(n)| + |L_{out}(n)| + \sum_{l \in L_{in}(n)} |N_{to}^I(l)| + |L_{in}(n)| \right\}.$$

Since $\sum_n \sum_{l \in L_{out}(n)} |N_{to}^I(l)| = \sum_n \sum_{l \in L_{in}(n)} |N_{to}^I(l)|$ and $\sum_n |L_{out}(n)| = \sum_n |L_{in}(n)| = |L|$, the difference between M_1 and M_2 is obtained by

$$M_1 - M_2 = (H - 1) \sum_n \left\{ |L_{from}^I(n)| + \sum_{l \in L_{out}(n)} |N_{to}^I(l)| \right\} - 2|L|.$$

For the network in Figure 1, $M_1 = 49$ and $M_2 = 40$. It is important to notice that both M_1 and M_2 only grow linearly with the number of interfering nodes and links.

Remark 5: There are several heuristics that can substantially reduce the amount of message passing. In practice, these heuristics will be used for protocol implementation. Performance of these heuristics is found to come close to the benchmarks established by the optimal algorithms, as shown by simulations in Section IV.

- **Heuristic 1:** In (11), $\lambda_l(t+1)$ is determined by the difference between its desired data rate $x'_l(t)$ and the experienced data rate $c'_l + \log p_l(t) + \sum_{k \in N_{to}^I(l)} \log(1 - P^k(t))$ of link l . Hence, if a node for each link (the transmitter in Algorithm 1 and the receiver in Algorithm 2) can measure its experienced data rate, messages $P^k(t)$ and $p_l(t)$ do not have to be exchanged among the nodes. Message

passing overhead is substantially reduced, e.g., we now have $M_1 = 23$ and $M_2 = 20$ in the earlier example.

- **Heuristic 2:** Each node needs to transmit messages only if the difference between the current and previous values for each variable exceeds some threshold. For example, node n transmits $\lambda_l(t)$ only if $|\lambda_l(t) - \lambda_l(t-1)| > \epsilon$. In addition, each node needs to update its variables only when it receives new values for variables from its neighbor nodes. By doing this, as the algorithm converges to the optimum, the amount of message passing will be reduced and eventually there will be no message passing and the algorithm terminates.
- **Heuristic 3:** In Algorithm 2, since message passing is conducted within one-hop distance, each node may broadcast its information through data packets by piggybacking to its neighboring nodes. If a node receives information successfully, it can update its information accordingly and perform the algorithm with the updated information. However, some nodes may not receive information successfully due to collision. In this case, those nodes may perform the algorithm with outdated information.

We now prove the optimality and convergence of Algorithms 1 and 2. For a rigorous proof, we first need the following technical condition to have a unique solution to problem (10) at the optimal dual solution. At the optimal dual solution λ^* ,

$$\sum_{l \in L_{out}(n)} \lambda_l^* + \sum_{l \in L_{from}^I(n)} \lambda_l^* \neq 0, \quad \forall n.$$

A sufficient condition that satisfies the above assumption is that for each node n , there exists at least one link l , $l \in L_{out}(n) \cup L_{from}^I(n)$ such that $x'_l < x'_l{}^{max}$. Hence, it is easily satisfied in most cases.⁶

Theorem 1: Algorithms 1 and 2 converge to the optimal dual solution λ^* that solves problem (8). Furthermore, at λ^* , solutions to problems in (9) and (10): \mathbf{x}^* , \mathbf{p}^* , and \mathbf{P}^* are the optimal solutions to problem (5).

Proof: Due to space limitation, we sketch the proof outline. By Danskin's theorem [18],

$$\frac{\partial Q(\lambda)}{\partial \lambda_l} = c'_l + \log p_l + \sum_{k \in N_{to}^I(l)} \log(1 - P^k) - x'_l.$$

Hence, (11) is a subgradient algorithm for the dual problem as derived in the steps preceding the description of Algorithm 1. There thus exists a step size $\alpha(t)$ (e.g., $\alpha(t) = 1/t$) that guarantees $\lambda(t)$ to converge to the optimal dual solution λ^* [19].⁷

Problem (5) is a convex optimization problem, and by the assumption above, each of problems (9) and (10) has a unique solution at the dual optimal solution λ^* . Hence, from Property 6.5 in [20], \mathbf{x}^* , \mathbf{p}^* , and \mathbf{P}^* are the optimal solutions to problem (5). ■

⁶Note that we do not need this assumption, if we add a penalty term, such as $-\beta \sum_n (P^n)^2$ for a small $\beta > 0$. This makes the objective function of problem (5) strictly concave and, thus, the solution to problem (10) unique.

⁷In practice, a constant step size (i.e., $\alpha(t) = \alpha$, $\forall t$) might be more desirable than a diminishing step size. In this case, $\lambda(t)$ converges to a neighbor of λ^* [19], providing an approximately optimal solution.

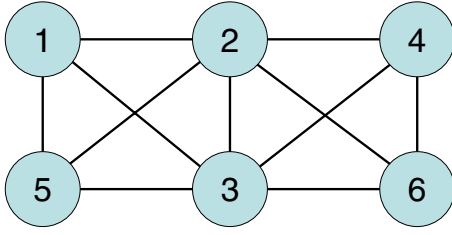


Fig. 2. A contention graph based on the network in Figure 1.

Remark 6: In the proof of Theorem 1, it is implicitly assumed that the algorithm is performed synchronously. However, it can be shown that the asynchronous dual based algorithm still converges to the optimal solution with minor conditions [17].

IV. NUMERICAL EXAMPLES

A. Probabilistic Model vs. Deterministic Model

In this and the next subsections, we provide simulation results for the proposed algorithms in Section III, considering the network in Figure 1. In this subsection, we first study the accuracy of our probabilistic NUM formulation based on collision and persistence probabilities, and compare it with that of the deterministic approximation approach recently studied in [3], [6], [7].

We first briefly summarize the deterministic approximation approach. We refer readers to [6], [7] for more details. The key idea of this approach is to introduce a contention graph [3], where each vertex represents a link in the network and two vertices are connected with an edge if transmissions from the links in the network corresponding to those two vertices interfere with each other. Figure 2 shows a contention graph for the network in Figure 1 assuming that transmissions from two links interfere with each other, if the distance between the transmitter node of one link and the receiver node of the other link is less than $2d$. Hence, if two vertices in the contention graph is connected, then the links in the network corresponding to those vertices cannot transmit data simultaneously without collision. In other words, only one link in the same maximal clique in the contention graph can transmit data successfully at a time. Hence, in the deterministic approximation, each maximum clique is defined as a resource with a finite capacity that is shared by the links belonging to the clique. Capacity of a clique is defined as the maximum value of the sum of time fractions that each link in the clique can transmit data without collision. Consequently, a NUM problem has been formulated as follows, with capacity constraint C_{CL_i} at each maximal clique CL_i :

$$\begin{aligned} & \text{maximize } \sum_l U_l(x_l) \\ & \text{subject to } \sum_{l \in L(CL_i)} \frac{x_l}{c_l} \leq C_{CL_i} \quad \forall i, \\ & \quad \mathbf{x}^{\min} \preceq \mathbf{x} \preceq \mathbf{x}^{\max}, \end{aligned} \quad (12)$$

where x_l is the average data rate of link l , c_l is the data rate of link l , $L(CL_i)$ is a set of links that are in maximal clique CL_i , and C_{CL_i} is the capacity of CL_i , ($0 \leq C_{CL_i} \leq 1$). The above problem is a separable and convex optimization problem, and similar to the basic NUM problem in (1). Hence, one can

easily solve it by using the same algorithm used to solve the basic NUM problem (1) for TCP congestion control.

However, there are several drawbacks in this deterministic approximation. First, in this approach to wireless MAC, each maximal clique generates feedback information to solve the problem. The maximal clique is an artificially constructed entity that does not have a physical controller. Second, in general we do not know the capacity of the maximal clique *a priori*. Third, it is assumed that the average data rate of each link depends only on the fraction of the time that it transmits data. This is implicitly assuming that no collision occurs, if the constraint in problem (12) is satisfied. However, this can happen only when transmissions of links are properly scheduled (probably by a central controller) to avoid collisions. For a random-access protocol, where the time fraction that each link transmits is mapped into persistence probability or backoff window size, there exists a non-negligible probability of collision even when the constraint in problem (12) is satisfied.⁸ Hence, the actual performance of the resulting random-access protocol can be quite different from the analytical solution obtained by the deterministic approximation (*i.e.*, by solving problem (12)).

In the first example, we consider proportional fairness, *i.e.*, the utility function of each link l is defined as $U_l(x_l) = \log(x_l)$. We set $c_l = 10$ (Mbps), $x_l^{\min} = 0$, and $x_l^{\max} = c_l$ for each link l . We assume that nodes and links are synchronized. We first present the performance of the deterministic approximation approach. To solve problem (12), we need to know the capacity of each clique to ensure the feasibility of the problem, which is in general difficult to know. However, in this simple example, we can easily know that the capacity of each clique is one. In Table I, we first provide the solutions of problem (12) with $C_{CL_i} = 1, \forall i$, which are denoted as ‘analysis’. We also provide simulation results of the random-access MAC protocol with the deterministic approximation. Since x_l/c_l is the time fraction that link l must transmit data to achieve the average data rate x_l , x_l/c_l is interpreted in [7] as the persistence probability of link l . We provide simulation results for the resulting random-access MAC protocol in Table I, which are denoted as ‘simulation’. The results show that there is a large difference between the solution to problem (12) and the performance of the random-access MAC protocol. This discrepancy arises due to the ignorance of collisions in the analytic model of the deterministic approximation approach. This implies that the deterministic approximation model is appropriate only for the collision-free scheduling-based protocol that requires a central controller.

Even though we can easily know that the capacity of each clique is one in this simple example, it is difficult to know for general networks. However, it is known that if the capacity of the clique is set to be less than $2/3$, it is sufficient to guarantee schedulability [6]. Hence, in Table II, we provide the solutions of problem (12) with $C_{CL_i} = 2/3, \forall i$, which can be used as the capacity of the clique in general cases. The results show that we have not only a large difference between analysis

⁸In [7], a scheduling-based algorithm that does not require the capacity of the clique and guarantees no collision is also proposed. However, to be implemented in a distributed way, it requires a major approximation, in which each link rounds the obtained solution to either zero or its maximum capacity.

TABLE I
DETERMINISTIC APPROXIMATION WITH $C_{CL_i} = 1$.

Link	1	2	3	4	5	6	Total
x_l (analysis)	3.35	1.67	1.67	3.35	3.35	3.35	16.74
x_l (simulation)	1.55	0.41	0.49	2.23	1.55	1.55	7.78
$U_l(x_l)$ (analysis)	1.21	0.51	0.51	1.21	1.21	1.21	5.86
$U_l(x_l)$ (simulation)	0.44	-0.89	-0.71	0.80	0.44	0.44	0.52

TABLE II
DETERMINISTIC APPROXIMATION WITH $C_{CL_i} = 2/3$.

Link	1	2	3	4	5	6	Total
x_l (analysis)	2.41	1.2	1.2	2.41	2.41	2.41	12.04
x_l (simulation)	1.42	0.46	0.52	1.83	1.42	1.42	7.07
$U_l(x_l)$ (analysis)	0.88	0.19	0.19	0.88	0.88	0.88	3.90
$U_l(x_l)$ (simulation)	0.35	-0.77	-0.65	0.60	0.35	0.35	0.23

TABLE III
PERFORMANCE OF THE PROPOSED RANDOM ACCESS PROTOCOL (ALGORITHM 1).

Link	1	2	3	4	5	6	Total
p_l	0.5	0.25	0.20	0.25	0.25	0.25	
x_l (analysis)	2.25	0.84	0.84	1.88	0.75	1.13	7.69
x_l (simulation)	2.25	0.84	0.85	1.88	0.75	1.13	7.70
$U_l(x_l)$ (analysis)	0.81	-0.17	-0.17	0.63	-0.29	0.12	0.93
$U_l(x_l)$ (simulation)	0.81	-0.17	-0.17	0.63	-0.29	0.12	0.93

and simulation results but also much lower performance than the results with the true capacity of the clique. The results from Tables I and II imply that due to two major drawbacks: ignoring collisions and no knowledge of the capacity of the clique, the deterministic approximation is not the appropriate approach to study a random-access-based MAC protocol.

In Table III, we now summarize the performance of our random-access protocol. We show the average data rate x_l , the achieve utility $U_l(x_l)$, and the optimal persistence probability p_l of each link. We compare analysis results, which are solutions to problem (4), with simulation results, in which each node transmits randomly according to persistence probabilities obtained by the proposed protocol. The results from simulation are very close to those from analysis, showing that our probabilistic approach provides an accurate model for a random-access-based MAC protocol.

B. Efficiency-Fairness Comparison

This subsection focuses on comparing the efficiency-fairness tradeoffs of various protocols: our random-access protocols (Algorithm 1 and Algorithm 2), the deterministic approximation protocol, and the window-based BEB protocol. We also show how the desired efficiency-fairness tradeoff can be achieved by appropriately adjusting the parameters of utility functions. The results show that our protocols provide much better efficiency-fairness tradeoffs than both the deterministic approximation and BEB protocols.

- In Algorithm 1 and the deterministic approximation, we assume that each node always has up-to-date information to perform the algorithm in each time-slot. For the

deterministic approximation, we set the capacity of each clique to be one, which is the true capacity of the clique, thus favoring the deterministic model in the simulation.

- In Algorithm 2, each node n broadcasts its information (i.e., $P^n(t)$, $p_l(t)$, $\forall l \in L_{out}(n)$, and $\lambda_l(t)$, $\forall l \in L_{in}(n)$) through data packets by piggy-backing to its neighboring nodes (Heuristic 3 in Remark 5). Due to collision, some nodes may not receive it successfully. Hence, in Algorithm 2 simulation, some nodes actually carry out the algorithm with outdated information.
- Window-based BEB protocol's performance highly depends on the choice of maximum and minimum window sizes, W_l^{max} and W_l^{min} . Hence, we first simulate the BEB protocol with various values for maximum and minimum window sizes. We present the performance of the BEB protocol when average-performance parameters are chosen: $W_l^{max} = 20$ and $W_l^{min} = 10$.

In this experiment, the utility function for each link l , $U_l(x_l)$ is in the following standard form of concave utility parameterized by α in (2), shifted such that $U_l(x_l^{min}) = 0$ and $U_l(x_l^{max}) = 1$:

$$U_l(x_l) = \frac{x_l^{(1-\alpha)} - x_l^{min(1-\alpha)}}{x_l^{max(1-\alpha)} - x_l^{min(1-\alpha)}}.$$

We set $x_l^{min} = 0.5$ and $x_l^{max} = 5$, $\forall l$, and compare the network utility, and the tradeoff curve between rate and fairness for each protocol, varying the value of α from 1 to 4.

In Figure 3, we compare trajectories of persistence probabilities of links 1 and 3 in Algorithm 1 and Algorithm 2 with $\alpha = 1.5$. The result shows that even though the

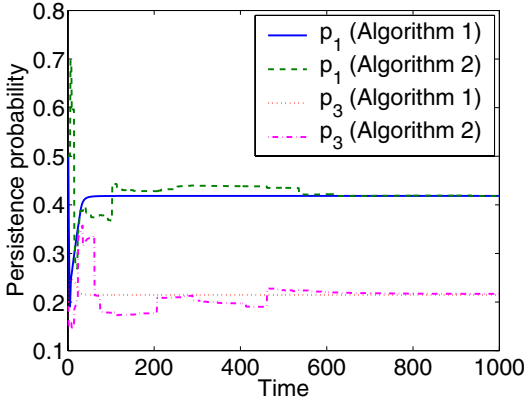


Fig. 3. Comparison of the trajectories of persistence probabilities in two versions of the proposed random access protocols.

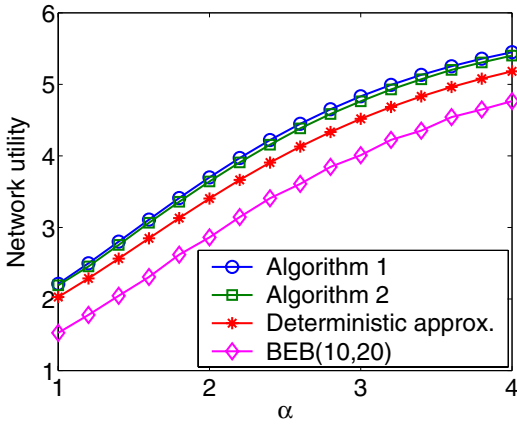


Fig. 4. Comparison of network utilities across four contention based protocols.

persistence probability of the link in Algorithm 2 converges slower than that in Algorithm 1, they converge to almost the same value. Hence, we expect that Algorithm 2, with substantial savings in message passing overhead, provides almost the same performance as that of Algorithm 1, as confirmed in the next set of results.

In Figure 4, we compare the network utility achieved by each protocol. This figure shows that our two protocols (Algorithm 1 and Algorithm 2 with reduced message passing overhead) provide almost the same network utility. It also shows that our protocols outperform the other protocols, *i.e.*, BEB and deterministic approximation protocols.

In Figure 5, we show the tradeoff curve of rate and fairness for each protocol. To compare fairness, we use the following fairness index $f(\mathbf{x})$ defined in [21]:

$$f(\mathbf{x}) = \frac{(\sum_l x_l)^2}{|L| \sum_l x_l^2}, \quad (13)$$

where x_l is the average data rate achieved by link l and $|L|$ is the number of links in the network. A higher value of $f(\mathbf{x})$ implies a higher degree of fairness. For each protocol shown in the graph, the area to the left and below of the tradeoff curve is the achievable region and the area to the right and above of the tradeoff curve is the infeasible region. Operating on the boundary of the achievable region, *i.e.*, the Pareto optimal

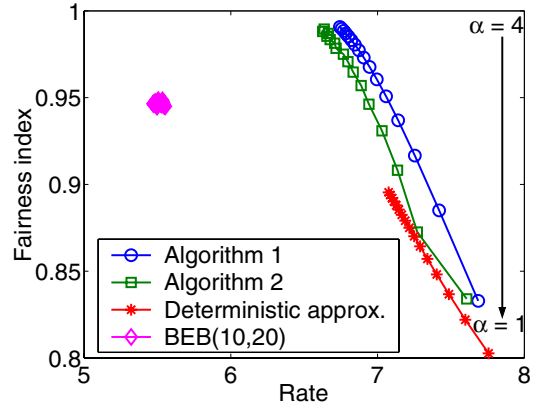


Fig. 5. Comparison of rate-fairness tradeoff across four contention based protocols.

tradeoff curve,⁹ is the best. Since the BEB protocol is a static protocol, it always provides the same efficiency and fairness regardless of the choice of utility functions. Hence, we cannot control the efficiency-fairness tradeoff in the BEB protocol. The figure shows that our protocols not only provide a higher fairness index but also have a much wider dynamic range of tradeoff than deterministic approximation and BEB protocols.

V. CONCLUSION AND FUTURE WORK

We have developed a NUM framework to achieve desired efficiency and fairness and their tradeoff by appropriately adjusting the utility function and solving the resulting NUM problem. Unlike recent publications on MAC NUM, we explicitly model collision and persistence probabilities and allow general utility functions. The resulting NUM problem is coupled and non-convex optimization, and we show how to distributively solve it for global optimality despite such difficulties. The solution then leads to two distributed random-access MAC protocols.

We also have compared the performances of our protocols with those of the deterministic approximation protocol and the standard BEB protocol, showing that both of our protocols can provide not only a higher network utility and a larger fairness index, but also a wider dynamic range of the tradeoff curve between efficiency and fairness. Performance guarantee of convergence to the global optimum of the NUM formulation is rigorously proved for the proposed algorithms, and simplifying heuristics are then developed based on the optimal algorithms.

There are several directions to extend our work. For example, stochastic versions of MAC NUM formulations and solutions need to be developed to incorporate the arrival statistics of packets and sessions. Then MAC protocols can be analyzed and designed using both stochastic stability results in traditional queuing models and utility-optimality results in this paper. Furthermore, using the ‘layering as NUM

⁹Let $r(\mathbf{p}, \mathbf{P})$ and $f(\mathbf{p}, \mathbf{P})$ be the total rate and fairness index achieved with persistence probabilities \mathbf{p} and \mathbf{P} . An efficiency-fairness (total rate-fairness index) vector, $(r(\mathbf{p}^*, \mathbf{P}^*), f(\mathbf{p}^*, \mathbf{P}^*))$ is said to be the Pareto optimal tradeoff, if there is no other feasible persistence probabilities \mathbf{p} and \mathbf{P} such that $r(\mathbf{p}, \mathbf{P}) \geq r(\mathbf{p}^*, \mathbf{P}^*)$ and $f(\mathbf{p}, \mathbf{P}) > f(\mathbf{p}^*, \mathbf{P}^*)$, or $r(\mathbf{p}, \mathbf{P}) > r(\mathbf{p}^*, \mathbf{P}^*)$ and $f(\mathbf{p}, \mathbf{P}) \geq f(\mathbf{p}^*, \mathbf{P}^*)$. In other words, at the Pareto optimal efficiency-fairness tradeoff, we cannot increase efficiency (or fairness) without decreasing fairness (or efficiency).

decomposition' approach and formulating utilities as functions of end-to-end rates, the methodologies to tackle coupling and non-convexity in this paper can be readily extended to investigate the interactions among transport (e.g., end-to-end rate allocation), network (e.g., routing), link (e.g., medium access), and physical (e.g., power and coding level control) layers in wireless multi-hop networks. Indeed, the extension to jointly optimal end-to-end rate control and per-hop contention control has been recently investigated [9], [22].

APPENDIX

Consider the following optimization problem.

$$\begin{aligned} & \text{maximize} \quad \sum_{l \in L_{out}(n)} \lambda_l \log p_l + \sum_{k \in L_{from}^I(n)} \lambda_k \log (1 - P^n) \\ & \text{subject to} \quad \sum_{l \in L_{out}(n)} p_l = P^n \\ & \quad 0 \leq p_l \leq 1, \quad \forall l \in L_{out}(n) \\ & \quad 0 \leq P^n \leq 1. \end{aligned}$$

Let

$$\begin{aligned} f^n(\mathbf{p}) = & \sum_{l \in L_{out}(n)} \lambda_l \log p_l \\ & + \sum_{k \in L_{from}^I(n)} \lambda_k \log (1 - \sum_{l \in L_{out}(n)} p_l), \end{aligned}$$

which is the objective function of the above optimization problem taking into account the first constraint. Let \mathbf{p}^* and P^{n*} be the optimal solution of the problem. Then, from the first order condition that $\frac{\partial f^n(\mathbf{p}_n^*)}{\partial p_l} = 0$, $l \in L_{out}(n)$, we have

$$p_l^* = \frac{\lambda_l (1 - \sum_{m \in L_{out}(n)} p_m^*)}{\sum_{k \in L_{from}^I(n)} \lambda_k}. \quad (14)$$

By summing up both sides for $l \in L_{out}(n)$, we have

$$\sum_{l \in L_{out}(n)} p_l^* = \frac{\sum_{l \in L_{out}(n)} \lambda_l (1 - \sum_{m \in L_{out}(n)} p_m^*)}{\sum_{k \in L_{from}^I(n)} \lambda_k}.$$

Hence,

$$P^{n*} = \sum_{l \in L_{out}(n)} p_l^* = \frac{\sum_{k \in L_{out}(n)} \lambda_k}{\sum_{k \in L_{from}^I(n)} \lambda_k + \sum_{k \in L_{out}(n)} \lambda_k}$$

and by (14),

$$p_l^* = \frac{\lambda_l}{\sum_{k \in L_{from}^I(n)} \lambda_k + \sum_{l \in L_{out}(n)} \lambda_l}, \quad l \in L_{out}(n).$$

ACKNOWLEDGEMENT

We thank Steven Low and Lijun Chen at Caltech and Koushik Kar at RPI for helpful discussions on the topic.

REFERENCES

- [1] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Utility-optimal medium access control: reverse and forward engineering," in *Proc. IEEE INFOCOM 2006*.
- [2] IEEE, *Wireless LAN medium access control (MAC) and physical layer (PHY) specifications*, June 1999, IEEE Standard 802.11.
- [3] T. Nandagopal, T.-E. Kim, X. Gao, and V. Bharghavan, "Achieving MAC layer fairness in wireless packet networks," in *Proc. ACM MobiCom 2000*, pp. 87–98.

- [4] X. L. Huang and B. Bensaou, "On max-min fairness and scheduling in wireless ad-hoc networks: analytical framework and implementation," in *Proc. ACM MobiHOC 2001*, pp. 221–231.
- [5] L. Tassiulas and S. Sarkar, "Maxmin fair scheduling in wireless networks," in *Proc. IEEE INFOCOM 2002*, vol. 2, pp. 763–772.
- [6] Z. Fang and B. Bensaou, "Fair bandwidth sharing algorithms based on game theory frameworks for wireless ad-hoc networks," in *Proc. IEEE INFOCOM 2004*, vol. 2, pp. 1284–1295.
- [7] L. Chen, S. Low, and J. Doyle, "Joint congestion control and media access control design for ad hoc wireless networks," in *Proc. IEEE INFOCOM 2005*, vol. 3, pp. 2212–2222.
- [8] K. Kar, S. Sarkar, and L. Tassiulas, "Achieving proportional fairness using local information in Aloha networks," *IEEE Trans. Autom. Control*, vol. 49, no. 10, pp. 1858–1862, October 2004.
- [9] X. Wang and K. Kar, "Cross-layer rate control for end-to-end proportional fairness in wireless networks with random access," in *Proc. ACM MobiHoc 2005*, pp. 157–168.
- [10] J.-W. Lee, A. Tang, J. Huang, M. Chiang, and A. R. Calderbank, "Reverse engineering MAC: a non-cooperative game model," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 6, Aug. 2007 (to appear).
- [11] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *J. Operational Research Society*, vol. 49, no. 3, pp. 237–252, March 1998.
- [12] R. Srikant, *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.
- [13] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
- [14] R. J. La and V. Anantharam, "Utility-based rate control in the Internet for elastic traffic," *IEEE/ACM Trans. Networking*, vol. 10, no. 2, pp. 272–286, April 2002.
- [15] S. H. Low, "A duality model of TCP and queue management algorithms," *IEEE/ACM Trans. Networking*, vol. 11, no. 4, pp. 525–536, Aug. 2003.
- [16] M. Chiang, S. H. Low, A. R. Calderbank, J. C. Doyle, "Layering as optimization decomposition: a mathematical theory of network architecture," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.
- [17] S. H. Low and D. E. Lapsley, "Optimization flow control-I: basic algorithm and convergence," *IEEE/ACM Trans. Networking*, vol. 7, no. 6, pp. 861–874, Dec. 1999.
- [18] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.
- [19] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions*. Springer-Verlag, 1985.
- [20] M. Minoux, *Mathematical Programming: Theory and Algorithms*. Wiley, 1986.
- [21] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," Tech. Rep., Sept. 1984.
- [22] J.-W. Lee, M. Chiang, and A. R. Calderbank, "Jointly optimal congestion and medium access control based on network utility maximization," *IEEE Commun. Lett.*, vol. 10, no. 3, pp. 216–218, March 2006.



Jang-Won Lee (S'02-M'04) received his B.S. degree in Electronic Engineering from Yonsei University, Seoul, Korea in 1994, M.S. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Taejeon, Korea in 1996, and Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, USA in 2004. In 1997-1998, he was employed with Dacom R&D Center, Taejeon, Korea. In 2004-2005, he was a Postdoctoral Research Associate in the Department of Electrical Engineering at

Princeton University, Princeton, NJ, USA. Since September 2005, he has been an assistant professor in the School of Electrical and Electronic Engineering at Yonsei University, Seoul, Korea. His research interests include resource allocation, QoS and pricing issues, optimization, and performance analysis in communication networks.



Mung Chiang (S'00–M'03) received the B.S. (Hon.) degree in electrical engineering and mathematics, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1999, 2000, and 2003, respectively. He is an Assistant Professor of Electrical Engineering, and an affiliated faculty of Applied and Computational Mathematics and of Computer Science at Princeton University. He conducts research in the areas of optimization of communication systems, analytic foundations of network architectures, algorithms in

broadband access, and information theory stochastic models of communication. Prof. Chiang has been awarded a Hertz Foundation Fellow and received the Stanford University School of Engineering Terman Award, the SBC Communications New Technology Introduction Contribution Award, the National Science Foundation CAREER Award, the Young Investigator Award from the Office of Naval Research, and the Princeton University Howard B. Wentz Junior Faculty Award. He is a coauthor of the IEEE Globecom Best Student Paper Award, and one of his papers became a Fast Breaking Paper in Computer Science in 2006 by ISI citation data. He is the Lead Guest Editor of the *IEEE Journal on Selected Areas in Communications* (Special Issue on Nonlinear Optimization of Communication Systems), a Guest Editor of the *IEEE Transactions on Information Theory*, and *IEEE/ACM Transactions on Networking* (Joint Special Issue on Networking and Information Theory), an Editor of *IEEE Transactions on Wireless Communications*, the Program Co-Chair of the 38th Conference on Information Sciences and Systems, and the co-editor of Springer's book series on Control and Optimization of Communication Systems.



A. Robert Calderbank (M89 - SM97 - F98) received the BSc degree in 1975 from Warwick University, England, the MSc degree in 1976 from Oxford University, England, and the PhD degree in 1980 from the California Institute of Technology, all in mathematics.

He is currently Professor of Electrical Engineering and Mathematics at Princeton University where he directs the Program in Applied and Computational Mathematics. Dr. Calderbank joined Bell Telephone Laboratories as a Member of Technical

Staff in 1980, and retired from AT&T in 2003 as Vice President of Research. Dr. Calderbank has research interests that range from algebraic coding theory and quantum computing to the design of wireless and radar systems.

Dr. Calderbank served as Editor in Chief of the *IEEE Transactions on Information Theory* from 1995 to 1998, and as Associate Editor for Coding Techniques from 1986 to 1989. He was a member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996. Dr. Calderbank was honored by the IEEE Information Theory Prize Paper Award in 1995 for his work on the Z4 linearity of Kerdock and Preparata Codes (joint with A.R. Hammons Jr., P.V. Kumar, N.J.A. Sloane, and P. Sole), and again in 1999 for the invention of space-time codes (joint with V. Tarokh and N. Seshadri). He is a recipient of the IEEE Millennium Medal, and was elected to the National Academy of Engineering in 2005.