Provable Algorithms for Joint Optimization of Transport, Routing and MAC layers in Wireless Ad Hoc Networks

V. S. Anil Kumar
Virginia Bio-Informatics
Institute and Dept. of
Computer Science
Virginia Tech
Blacksburg, VA 24061
akumar@vbi.vt.edu

Madhav V. Marathe
Virginia Bio-Informatics
Institute and Dept. of
Computer Science
Virginia Tech
Blacksburg, VA 24061
mmarathe@vbi.vt.edu

Srinivasan Parthasarathy
Next Generation Distributed
Systems
IBM T. J. Watson Research
Center
Yorktown Heights, NY 10598
spartha@us.ibm.com

Aravind Srinivasan
Department of Computer
Science and Institute for
Advanced Computer Studies
University of Maryland
College Park, MD 20742
srin@cs.umd.edu

ABSTRACT

Given a wireless network and a collection of source-destination pairs $\{(s_i, t_i)\}$, what is the maximum end-to-end rate (throughput) at which the network can transfer data from the sources to their corresponding destinations? The problem is non-trivial to solve in the case of wireless networks due to interference. It is additionally complicated when taking into account TCP like transport protocols.

Here, we present near-optimal provably good polynomial-time routing and scheduling algorithms for solving these and other throughput maximization problems in wireless ad hoc networks. We also present distributed algorithms for *simultaneously* optimizing a large class of throughput related objectives with *fixed* routes and schedules. We consider a wide variety of conflict-graph based models with both primary and secondary wireless interference constraints. Our techniques can accommodate a variety of routing constraints such as low energy, low hop-count, etc. as well as incorporate wireless technologies such as multiple channels and directional antennas.

Categories and Subject Descriptors

C.2.2 [**Computer-Communication Networks**]: Network Protocols—*Routing Protocols*; F.2.3 [**Theory of Computation**]: ANALYSIS OF ALGORITHMS AND PROBLEM COMPLEXITY—*Tradeoffs between Complexity Measures*

General Terms

Algorithms, Theory

Keywords

Cross-layer design, interference, end-to-end scheduling, wireless networks, throughput maximization

1. INTRODUCTION

This paper describes algorithmic approaches for optimizing rate-related objectives in wireless ad hoc networks. In other words, given a collection of source-destination pairs $\{(s_i, t_i)\}$, what is the maximum rate (throughput) at which the network can transfer data from the sources to their corresponding destinations? For a wired network, some of these constraints can be formulated easily as a simple linear program (L), but this problem is non-trivial to solve in the case of wireless networks due to interference. The problem is further complicated when one tries to optimize non-linear throughput objectives that are related to realistic transport protocols such as TCP.

In the usual OSI protocol stack model, the above problem of transmitting packets for each source-destination pair is broken down into sub-problems, the most important of which are: (i) choosing routes for each such pair - a protocol like AODV chooses some sort of (single) shortest path for each pair, (ii) MAC scheduling of the packets along these paths - this resolves contention, and determines who sends at which time slot, (iii) actual transmission of the packets on the physical channel, and (iv) choosing actual rates of transmission for each pair - this is achieved dynamically by a TCP like protocol, which uses feedback from the network to regulate the flow. While this modularity is useful in designing the network, it is almost impossible to determine the quality of the performance of such protocols, and how to improve the performance. In fact, there is a significant interaction between protocols at different layers, and plugging in optimal protocols for each layer does not lead to optimal overall performance [3]. This motivates the study of unified cross layer aware protocols and associated measures and is currently an active topic of research [20, 33].

In this paper, we build on the recent work in [19, 13, 21, 34, 18, 16, 25] and present provably good near-optimal algorithms (those that are optimal to within a small factor) that compute end-to-end/link-level rates, routes and MAC-layer schedules for optimizing throughput related objectives in wireless networks. Our main contributions are the following three. First, Given a wireless network and a set of k-connections and a concave utility U associated with each end-to-end rate, we present an approximation algorithm for jointly determining the routes, end-to-end link rates and schedules for maximizing the sum of individual utilities subject to wireless interference. Our approach exploits geometric properties of wireless interference/signal propagation and provides first constant-factor/logarithmic approximation factor guarantees for a wide range of communication models. These models incorporate features of realistic networks such as connectivity loss due to occlusions and scheduling conflicts due to primary and secondary interference. Second, we also consider centralized/distributed algorithms for computing a *fixed* set of routes and a *fixed* schedule for *simultaneously* optimizing a large class of throughput related objectives. From a protocol perspective, this has the desirable consequence that our *fixed* routing and scheduling algorithm, in conjunction with *different* transport protocols, *simultaneously optimizes various aggregate utility functions encoded by these transport protocols.* Third, our approach makes it easy to incorporate any convex/linear constraints of end-to-end rates and link rates (e.g. low-energy, end-to-end fairness, low hop-count) in our framework with no loss in our approximation guarantees.

As noted above, a key feature of the our framework is the ability to incorporate secondary interference (i.e., interference between non-adjacent but proximate links in the network). The difficulty in modeling this in optimization programs can be seen from the fact that none of the existing utility maximization approaches provide provable approximation guarantees in the presence of secondary interference. To quote from a very recent work of Chen, Low, Chiang and Doyle [5], *" The conflict graph for the network with secondary interference is more complicated. We can ... formulate a utility optimization problem for the system and carry out cross-layer design ...... However, the scheduling problem will be much more difficult. It can be shown that it is equivalent to a maximum weight independent set problem, which is NP-hard for general graphs. It is easy to design some heuristic algorithm but is hard to bound its performance. However, due to the broadcast nature of wireless channel, it may be possible ... This will be part of our future work."* This paper takes the first step in addressing the issue by providing polynomial time algorithm with bounded worst case performance guarantees for the joint optimization problem for a wide variety of conflict-graph models with secondary interference, by exploiting the geometry in wireless networks.

We start by describing our models and notation and survey related work in Section 2. We follow this with our algorithm for utility maximization which achieves in wireless network with multi-path routing under a variety of communication/interference models (Section 3). We then describe our centralized and distributed algorithms for simultaneous utility maximization (Sections 4 and 4.1). Due to lack of space we omit the proofs of some of our claims.

2. BACKGROUND

Physical & MAC layers: Our physical+MAC layer wireless model consists of two parts: (i) a *radio-propagation* model and (ii) an *interference* model. The radio-propagation model determines which (ordered) pairs of nodes in the network can form a communication link and hence, specifies the directed communication graph $G = (V, E)$. Links are rate-limited and a link $e \in E$ has a fixed maximum channel-capacity of $c(e)$ bits/second. Time is synchronized and slotted; without loss of generality (w.l.o.g.), we assume that the duration of a time slot is one second. The interference model determines the associated *conflict graph* $H = (E, R)$ which specifies the interference relation between pairs of links in G: an undirected edge between two vertices in H implies that the corresponding links in G interfere with each other and hence cannot be active in the same time slot. In general, the conflict graph captures two types of interference relations: *primary* conflicts between links that are adjacent to a common node in G; *secondary* conflicts between proximal but non-adjacent links in G. For any link $e \in E$, we let $I(e)$ denote the the set of all links which interfere with e.

Given the network G and the associated conflict graph H, a feasible schedule S specifies a binary variable $X(e, t)$ such that (i) $X(e, t) = 1$ if and only if link e is active at time t, and (ii) if $X(e, t) = 1$, then, for all other edges e' conflicting with e, $X(e', t) = 0$. A link-utilization vector $\vec{x}$ for a schedule specifies a value $x(e)$ for each link e; this is the fraction of time for which link e is active during the schedule. The link-utilization vector $\vec{x}$ is feasible if and only if there is a feasible schedule S such that every link e is active for $x(e)$ fraction of the slots in S.

We model wireless communication networks as geometric intersection graphs. Specifically, we consider three graph-classes: disk graphs, quasi unit-disk graphs, and (r, s)-civilized graphs. A *disk graph* [22] is specified by a set of points V, with a disk $D(v)$ of radius $r(v)$ centered at each $v \in V$. The directed graph $G = (V, E)$ induced by these disks is the following: the set of nodes is V and a (directed) link (u, v) is present if $v \in D(u)$. A unit disk graph is a restriction of the disk graph wherein all disks have the same radii. An undirected graph $G = (V, E)$ is a quasi unit-disk graph [23, 24] parameterized by value $\rho \in [0, 1]$, if the vertices of G can be laid out in R^2 such that the following conditions hold: (i) every pair of points with distance at most ρ has a link between them; (ii) no pair of points with distance greater than 1 has a link between them. Notice that if the distance is strictly between ρ and 1, then the link may or may not exist. An undirected graph $G = (V, E)$ is said to be (r, s)-civilized [22] (where, the parameters $r, s > 0$ with $r < s$), if it can be embedded in R^2 such that for any pair of points u, v, their distance $d(u, v) \geq s$, and for any link $(u, v) \in E$, $d(u, v) \leq r$. Disk graphs allow for unidirectional edges and varying transmission radii but do not model loss of links due to occlusions. Both quasi unit-disk graphs and (r, s)-

civilized graphs allow for loss of links due to occlusions in a controlled manner. All three models generalize the standard unit-disk graph model in different ways. Finally, we note that while we described the graph models in terms of their two-dimensional layouts, the models and our results naturally generalize to three-dimensions as well.

Network & Transport layer: We use a multi-commodity flow model for network traffic: we have a set of k connections (possibly routed through multiple paths). The pair (s_i, t_i) represents the (source, sink) nodes of connection i. Let $\mathcal{P}_i$ denote the set of all paths between s_i and t_i in G. For connection i, and a path $p_i \in \mathcal{P}_i$, $f(p_i)$ denotes the rate at which data is routed across p_i by connection i. Thus, the total end-to-end rate f_i for connection i equals $\sum_{p_i \in \mathcal{P}_i} f(p_i)$. The data routed through p_i induces data-rate $f(p_i)$ on all the links in p_i. Thus, for any link $e \in E$, the total link-rate $l(e) = \sum_i \sum_{(p_i \in \mathcal{P}_i) \& (e \in p_i)} f(p_i)$.

We model the transport layer using the utility maximization framework introduced by Kelly et al. [21]. The intuition underlying this framework is that TCP congestion control algorithms can be viewed as distributed primal-dual algorithms which implicitly maximize the aggregate *concave* utility functions of the network connections. The concave utility function $U(f_i)$ for each connection i is specified (implicitly) by the TCP algorithm and the function value depends only on the end-to-end data-rate f_i.

Thus, the cross-layer network utility maximization problem can now be formally stated as follows. Given a communication graph $G = (V, E)$ the associated conflict-graph $H = (E, R)$, and a set of k connections, let Π denote the set of all feasible link-utilization vectors. The joint utility maximization problem seeks the optimal solution to the following (convex) program.

$$\max \sum_i U(f_i) \tag{1}$$

$$\forall i \in \{1, \ldots, k\},\ f_i \quad = \quad \sum_{p \in \mathcal{P}_i} f(p) \tag{2}$$

$$\forall e \in E, x(e) \quad = \quad \frac{\sum_i \sum_{(p \in \mathcal{P}_i) \& (p \ni e)} f(p)}{c(e)} \tag{3}$$

$$\vec{x} \in \Pi \tag{4}$$

$$\forall p \in \bigcup_i \mathcal{P}_i,\ f(p) \quad \geq \quad 0 \tag{5}$$

The first two constraints connect the end-to-end rates with the link utilization vector x and the third constraint requires that $\vec{x} \in \Pi$. Unfortunately, for most network/interference models, it is not possible to exactly characterize the link-utilization region Π using a polynomial sized convex program (unless P=NP). One of the main goals of this work is to obtain approximate, but provably good characterizations of the region Π using linear constraints in polynomial time.

Related work: There has been substantial work in recent years on various aspects of cross layer design and analysis in wire line and wireless networks. The seminal work by Kelly *et al.* [21] showed that in a wire line network, TCP congestion control algorithms can be viewed as distributed primal-dual algorithms that seek to maximize sum of aggregate concave utilities. Recently, extensions of this basic idea have been considered in the context of several applications including cross-layer optimization for wire line and wireless ad hoc networks, analysis of stability and optimality as a function of time-varying network conditions, etc (including loads, link capacity, etc.): see [19, 5, 14, 25, 14, 6, 15, 21, 34] and the references therein.

However, with the exception of Chen *et al.* [5], and Lin and Shroff [25], none of these consider the joint optimization of MAC, routing and transport layers together. Both [5] and [25] consider conflict-graph models with primary interference in their optimization framework. Further, Lin and Shroff [25] study the impact of suboptimal scheduling policies on the performance and distributed convergence of the joint optimization problem. However, both of these do not address the issue of obtaining provably good performance guarantees under more general conflict-graph models with secondary interference.

Chiang [6] considers the joint optimization of Transport+MAC layers, with aggregate concave utilities modeling the Transport layer objective and link-level power allocation vector being the optimization variable which determines maximum link capacities through SIR constraints. However, [6] does not consider the effect of MAC layer scheduling. Yi and Shakkottai [35] consider joint optimization of Transport+MAC under primary interference. The work of Chen, Low and Doyle [19] considers the joint optimization of Transport+MAC layer scheduling with rate-limited channels and conflict-graphs involving secondary interference, and is more closely related to our work. However, [19] does not consider routing and does not provide provably good performance guarantees for various conflict-graph models.

The issue of characterizing achievable rate regions (either exactly or approximately) is closely related to cross-layer optimization of throughput objectives in ad hoc networks. The first attempts towards this can be be traced back to Hajek and Sasaki [12], and to Baker *et al.* [2]. Several recent results have emerged for characterizing achievable rate regions using linear/convex programming techniques under various communication and conflict graph assumptions [18, 32, 13, 16]. Our results in Section 3 make use of linear constraints for the Tx, Protocol, and D2 interference models from [18] which yield approximate but provably good characterizations of rate-regions in the presence secondary interference. Additionally, characterizations of rate-regions in other generalized network/interference models are also obtained.

3. INTERFERENCE MODELING FOR UTILITY MAXIMIZATION

Given a communication graph $G = (V, E)$, an associated conflict graph $H = (E, R)$, a set of k connections, and a concave utility U for each connection which is a function of the end-to-end data rate, our goal is to jointly route and schedule these connections such that the aggregate utilities of all the connections is maximized. In this section, we present our solution to this problem for settings with multi-path routing. Our broad approach is to model this problem as a mathematical program, where the objective function is the aggregate utility and the constraints are the feasibility of link-flows. Define the feasible link-utilization region Π for a given graph G and an associated conflict graph H as the set of all feasible link-utilization vectors. The main question which we need to address is the following: how do we express the feasible link-utilization region Π efficiently so that it can be incorporated in the mathematical program.

The key driver behind our approach is the observation that, *for a wide class of radio-propagation and interference models, the associated link-utilization region can be modeled (approximately) using necessary and sufficient conditions that are linear.* This observation enables us to formulate and solve the network optimization problem approximately, but with *provably good performance guarantees.* While such performance guarantees are not possible for arbitrary communication/interference models, we show that this is indeed the case for a wide range of geometric radio-propagation models, and geometric/graph-theoretic interference models. Of course, these geometric modeling assumptions result in a loss of generality to certain extent but the models presented below can capture several important features of real-world scenarios. In particular, we note that physical obstructions, which prevent communication along links in G can be easily modeled by setting the link capacity of the obstructed links as zero.

3.1 Basic Approach

We illustrate our basic approach using the Tx-model. In the Tx-model, nodes are assumed to be embedded in the X-Y plane. Each node u has a transmission range $r(u)$ and a link (u, v) exists only if v is within distance $r(u)$ of u. Two nodes u and w can transmit simultaneously if and only if the distance between the two nodes $d(u, w)$ is such that $d(u, w) > (1 + \Delta)(r(u) + r(v))$, where $\Delta > 0$ is a fixed constant independent of the network. For a link $e = (u, v)$, let $I(e)$ denote the set of links which conflict with e: i.e., $I(e)$ consists of all links $e' = (p, q)$ such that $e' \neq e$ and $d(u, p) \leq (1 + \Delta)(r(u) + r(p))$. Let $I_{\geq}(e = (u, v)) \doteq \{e' = (p, q) : (e' \in I(e))\ \&\ (r(p) \geq r(u))\}$ (i.e., e' conflicts with e and the source node of e' has a range greater than or equal to the range of source of e).

Let $\vec{x}$ be a link-utilization vector. Kumar *et al.* [18] showed the following necessary and sufficient conditions for the feasibility of the link-utilization vector $\vec{x}$ under the Tx-model:

$$\forall e \in E,\ x(e) + \sum_{e' \in I_{\geq}(e)} x(e') \leq 5 \text{ (necess. cond.)} \quad (6)$$

$$\forall e \in E,\ x(e) + \sum_{e' \in I_{\geq}(e)} x(e') \leq 1 \text{ (suff. cond.)} \quad (7)$$

Inductive Scheduling: For the sake of completeness, we also present the inductive scheduling algorithm from [18] which schedules any feasible vector $\vec{x}$ satisfying (7). Consider a time window with W slots such that $x(e) \cdot W$ is integral for all links e. Process the links in the decreasing order of the range of their source nodes. Let the current link being processed be e. Allocate any set of $x(e) \cdot W$ slots in the time window which have not already been allocated to any of the links in $I_{\geq}(e)$. This schedule can now be repeated periodically, with a period of W time slots. Since $\vec{x}$ satisfies (7), it is easy to see that this algorithm yields a conflict-free schedule with the desired link-rates.

We now formulate the joint network optimization problem for the Tx-model as follows:

$$\begin{aligned} \max \sum_i U(f_i) & \\ \forall i \in \{1, \ldots, k\},\ f_i &= \sum_{p \in \mathcal{P}_i} f(p) \\ \forall e \in E, x(e) &= \frac{\sum_i \sum_{(p \in \mathcal{P}_i) \& (p \ni e)} f(p)}{c(e)} \\ \forall e \in E,\ x(e) + \sum_{e' \in I_{\geq}(e)} x(e') &\leq 1 \\ \forall e \in E,\ x(e) &\in [0, 1] \\ \forall p \in \bigcup_i \mathcal{P}_i,\ f(p) &\geq 0 \end{aligned}$$

Since the utility function is concave, the joint optimization problem is a convex program and can be solved optimally in polynomial time[1]. Let OPT denote the optimal solution value of the convex program. Let OPT^* denote the optimal value of $\sum_i U(f_i)$ when it is maximized over the space of all the feasible link-utilization vectors under Tx-model (and not just the space of vectors which satisfy (7)). We have:

LEMMA 1. *$OPT \geq OPT^*/5$.*

PROOF. Let $\vec{x}^*$ be the link-utilization vector and $\vec{f}^*$ be the end-to-end rate vector which achieve the solution value OPT^*. Let $\vec{y} = \frac{\vec{x}^*}{5}$ and $\vec{g} = \frac{\vec{f}^*}{5}$ (i.e., each component of $\vec{y}$ and $\vec{g}$ is one-fifth of the corresponding components in $\vec{x}^*$ and $\vec{f}^*$). For commodity i, let $f_i^* = \sum_{p \in \mathcal{P}_i} f^*(p)$ and $g_i = \sum_{p \in \mathcal{P}_i} g(p)$. Observe that, for all i, $g_i = \frac{f_i^*}{5}$. Crucially, since the utility function U is concave, $U(g_i) = U(\frac{f_i^*}{5}) \geq \frac{U(f_i^*)}{5}$. Since x^* is a feasible link-utilization vector, it satisfies (6) and y satisfies (7). Hence, $(\vec{y}, \vec{g})$ represents a feasible solution for the convex program. The proof can be concluded by noting that $OPT \geq \sum_i U(g_i) \geq \sum_i \frac{U(f_i^*)}{5} \geq \frac{OPT^*}{5}$. $\square$

Kumar *et al.* [18] also provide necessary and sufficient constraints for link-utilization similar to (6) and (7) for the Protocol model [11] and Distance-2 model [17] of interference, while Alicherry *et al.* [1] provide such conditions for a combined Tx and Distance-2 type conflict model. Hence, the sufficient conditions for link-utilization feasibility from all of these models can be incorporated into the convex program to obtain a constant factor performance guarantee for the joint network optimization problem. Note also that fairness constraints such as end-to-end rates of each pair of connections is at least α (where $\alpha \leq 1$ is the end-to-end fairness index), can be easily expressed using linear constraints; Such constraints can directly be accommodated in our above framework. In general, other linear constraints (average energy consumed by a connection, average hop-length for a connection) can also be incorporated above, leading to

---

[1] While the convex program as presented could be exponential in size, the program can be expressed in polynomial size using standard flow-conservation constraints and the added link feasibility conditions. We ignore the additive error ϵ in the solution of the convex program: ϵ can be made arbitrarily small.

constant-factor approximations for the energy/fairness/hop-count constrained optimization problems.

3.2 The q-inductive Transmission Model

All the transmission models discussed above share a general property which allows their link-utilization regions to be expressed (approximately) using a small set of linear inequalities. We now identify this key property and show how to approximately express the link-utilization region for any transmission model which satisfies this property. This enables us to (approximately) characterize the link-utilization region for a very wide class of transmission models as shown later in this section.

Given the communication graph G and the conflict graph H, we say that the pair (G, H) satisfies the **q-induction property** if there exists a total ordering $\succ$ on the links of the graph such that for all $e \in E$, the maximum number of links in $I_{\succ}(e)$ which can be active simultaneously is at most q. Here $I_{\succ}(e)$ denotes the set of edges e' such that $e' \in I(e)$ and $e' \succ e$. By abstracting the proof in the preceding subsection, we get the following general theorem.

THEOREM 1. *Suppose a given graph $G = (V, E)$ and its associated conflict graph $H = (E, I)$ have a q-inductive total ordering $\succ$ on E. For any feasible link-utilization vector $\vec{x}$, the following necessary condition must be satisfied.*

$$\forall e \in E,\ x(e) + \sum_{e' \in I_{\succ}(e)} x(e') \quad \leq \quad q \tag{8}$$

On the other hand, any vector $\vec{x}$ satisfying the following condition:

$$\forall e \in E,\ x(e) + \sum_{e' \in I_{\succ}(e)} x(e') \quad \leq \quad 1 \tag{9}$$

can be scheduled. Hence, the joint optimization problem for any instance with the q-inductive property can be solved in polynomial time with a performance guarantee of $O(q)$ for any concave utility function U.

PROOF. In any feasible schedule, $\forall e \in E,\ \forall t,\ X(e, t) + \sum_{e' \in I_{\succ}(e)} X(e', t) \leq q$. This follows from the fact that during slot t, either e can be active and all the edges in the set $I_{\succ}(e)$ inactive, or e can be inactive and at most q links from the set $I_{\succ}(e)$ be active. The necessary condition now follows by averaging both sides of this expression over all time slots. The inductive scheduling algorithm presented earlier can be used to produce a feasible schedule which satisfies all the link-utilization demands, whenever condition (9) holds. Hence, (9) represents a sufficient condition for feasibility. The performance guarantee q for the solution of the joint optimization convex program follows as a corollary as in the proof of Lemma 1. ☐

3.3 Unified framework for conflict modeling

Ramanathan [28] introduced a unified framework for the study of resource assignment problems in wireless networks under a wide-variety of graph-theoretic conflict graph models (the resources to be assigned being time-slots, frequencies, or codes). This framework identifies eleven atomic conflicts underlying most assignment problems, and an assignment problem is characterized by a combination of these conflicts. We now describe this unified framework for conflict-modeling briefly and show how to express the link-utilization regions for all assignment problems that are characterized by a combination of these conflicts models.

A conflict is a symmetric relation between two vertices or two links in a graph. A conflict imposes a restriction that the entities (nodes/links) which conflict with each can not be active during the same time slot in any schedule. In Ramanathan's framework, the constraints are classified according to whether they are between vertices or edges, the graph-theoretic separation between them, and whether it is a transmitter and/or a receiver based constraint. Specifically, constraint c is denoted using the syntax $c =< \epsilon >^{<s>}_{<d>}$, where $\epsilon \in \{V, E\}$, $s \in \{0, 1\}$, and $d \in \{tr, tt, rr, rt\}$. Here, ϵ is the entity (node (V) or link (E)) being constrained, s is the forbidden separation between two nodes or edges, and d qualifies the separation by specifying its direction with respect to transmitter (t) or receiver (r).

Assignment problems are characterized by a combination of conflicts, i.e., a conflict set; e.g., $C = \{E^0_{rr}, E^0_{tt}, E^0_{tr}, E^1_{tr}\}$ characterizes a half-duplex TDMA link scheduling problem, where two links can be simultaneously active implies that, there is no link from the transmitter of one link to the receiver of the other link (E^1_{tr}), and the two active links are not incident on the same node $(E^0_{tt}, E^0_{rr}, E^0_{tr})$. If the nodes are capable of full duplex communications (i.e., can transmit and receive simultaneously), then E^0_{tr} constraint can be removed.

Let $\succ$ be an arbitrary total ordering of the edges in the network. Given a edge-based conflict set C, let $I(e)$ denote the set of all edges which conflict with edge e under the conflict set C. Let $I_{\succ}(e)$ be the subset of links e' in $I(e)$ such that $e' \succ e$. Hence, by Theorem 1, (9) represents a sufficient condition for the feasibility of the a link-utilization vector x. We note that if the conflict set is vertex based, then, a similar sufficient condition holds for any node-utilization vector x. This allows us to (approximately) express link-utilization regions and obtain solutions for the joint network optimization problem under a large class of of assignment constraints in Ramanathan's framework. In general, if the communication graph G is allowed to be arbitrary, the overall performance ratio for this technique would be proportional to the maximum degree of the communication graph G. However, for many practical graph classes such as (r, s)-civilized graphs [22] and quasi unit-disk graph models [23, 24], which model realistic communication scenarios, the performance guarantee for most assignment problems in Ramanathan's framework can be shown to be dependent only on the protocol parameters (such as r and s), and independent of network parameters such as size and maximum degree, as described in the following theorem.

THEOREM 2. *Let $\mathcal{G}$ denote one of the graph classes: disk, quasi unit disk or (r, s)-civilized and let I denote one of following interference models: (T/F)-DMA broadcast, Distance-2matching, Cellular, (T/F)-DMA link and handshake. Then for each tuple (g, i) such that $g \in \mathcal{G}$ and $i \in I$, there exists a constant $q(g, i)$ such that (g, i) is $q(g, i)$ inductive. The specific values are summarized in Table 2.*

Table 1: The approximation factors for different combinations of Interference models and different conflict graph models of the physical layer

	(T/F)-DMA broadcast (V_{tr}^0, V_{tt}^1)	Distance-2 matching $(E_{xy}^0, E_{xy}^1, \forall x, y)$	Cellular V_{tr}^0	(T/F)-DMA link $(E_{rr}^0, E_{tr}^0, E_{tt}^0, E_{tr}^1)$	Handshake $(E_{tt}^0, E_{xy}^1, \forall x, y)$
Disk	$O(1)$	$O(1)$	$O(1)$	Δ	$O(1)$
Quasi unit disk	$(\frac{4}{d}+1)^2$	$(\frac{4}{d}+1)^2$	Δ	Δ	$O(1)$
(r,s)-civilized	$(\frac{4s}{r}+1)^2$	$(\frac{4s}{r}+1)^2$	Δ	Δ	$O(1)$

PROOF. The proof consists of showing that each entry (i, j) in the table yields a $q(i, j)$-inductive transmission model in the table; the result then follows by applying Theorem 1. The proof is very similar for each case. Thus due to lack of space, we will only discuss the case of disk graphs here and the (T/F)-DMA broadcast model.

For $e = (u, v)$, define $r(e) = \max\{r(u), r(v)\}$. For the (T/F)-DMA broadcast scheduling model, we consider the inductive ordering on the vertices - they are ordered $v_1 \prec v_2 \prec \ldots$ in non-decreasing order of radii. Consider any node v. There it is not possible to have active nodes w_1, w_2 such that (w_1, v) and (w_2, v) are edges, since this violates the V_{tt}^1 constraint. Therefore, at most one of the in-neighbors of v can be active at a time. Next, suppose $A = \{w_i : (v, w_i) \in E, v \prec w_i\}$. Therefore, A is the set of active out-neighbors of v that occur after it in the inductive order, and so $r(w_i) \geq r(v)$ for each i. ¿From the V_{tt}^1 constraint, we must have that for any $w_i, w_j \in A$, $d(w_i, w_j) > r(v)$. Therefore, $|A| \leq (\frac{r}{r/2})^2 + 1 = 5$. ☐

Remark While it is intuitively obvious that interference constraints can be expressed as node and edge coloring constraints, we should point out that it is not at all obvious if these constraints can be linearized. Theorem 2 is doing precisely that, and shows that in most cases, the approximation factor for the coloring problem translates to similar bounds for the utility maximization. Finally, we remark that while our constant factors in the approximation guarantees are not very close to one, they are the only known *rigorous* performance guarantees that hold for the wide variety of models considered here.

4. SIMULTANEOUS UTILITY MAXIMIZATION

So far we have studied the optimization of a specific utility function and have described algorithms to compute rates which approximate the objective within an constant factor. The specific utility functions of most interest are naturally the ones that model TCP. However, as many papers [20, 19, 33, 26] show, different variants of TCP can be modeled by different utility functions. In this context, Cho and Goel [7] raise the following important question: *is it possible to obtain a rate vector that is good with respect to all these variants?* By extending the results of [7, 9] on wire line networks, we show in this section that such simultaneous approximations are indeed possible for wireless networks, and for a large class of utility functions, called the *canonical utility functions*, which are defined below. From a practical viewpoint, computing such a rate vector in a distributed manner is a very important issue, and again, we extend the results of [7] and describe a distributed algorithm for computing the rates.

We first recall the following definitions and notation from Goel and Meyerson [9]. For any end-to-end flow vector $\vec{f} = \langle f_1, f_2, \ldots, f_k \rangle$, denote the i-th smallest component of $\vec{f}$ by $f_{(i)}$. Define $P_j(\vec{f}) = \sum_{i=1}^{j} f_{(i)}$. This is the j-th prefix of vector $\vec{f}$, the sum of its j smallest coordinates.

Definition 1. [9] Given two k-dimensional vectors $\vec{f}$ and $\vec{g}$, $\vec{f}$ is said to be α-supermajorized by $\vec{g}$ if $\alpha P_j(\vec{f}) \geq P_j(\vec{g})$ for all $j \leq k$. This is denoted by $\vec{f} \prec^{\alpha} \vec{g}$. A vector is said to be globally α-fair if it is α-supermajorized by any other feasible vector.

We omit the adjective "global" for the sake of brevity. We will be considering concave utility functions; we will say that a concave utility function U is *canonical* if $U(0) = 0$, it is symmetric and non-decreasing in any argument. A resource allocation problem is one that involves maximizing a canonical utility function over a convex set. The following Theorem from [9] establishes the significance of Definition 1.

THEOREM 3. *[9] A feasible solution $\vec{f}$ is a simultaneous α-approximation for a resource allocation problem if and only if $\vec{f}$ is α-fair.*

Given a convex set of k-dimensional vectors Π, the above theorem implies that it suffices to find a fair vector, in order to simultaneously optimize all canonical utility functions. However, it is not even obvious that such globally fair solutions exist for small values of α. The following surprising result from [9] gives a non-trivial bound on α for which such approximations exist for any convex set Π. Define $P_j^* = \max_{x \in \Pi} P_j(x)$.

THEOREM 4. *([9]) For any nonnegative convex program, there exists an $O(\log \frac{P_n^*}{nP_1^*})$-fair solution. Moreover, such a solution can be computed in polynomial time by solving n convex programs.*

We now show how the above result yields a logarithmic approximation to the utility maximization problem. Let $R = \frac{\max_e c(e)}{\min_e c(e)}$.

THEOREM 5. *Given a graph $G = (V, E)$, an associated conflict graph $H = (E, I)$ which has a q-inductive ordering on the edge set E, and a set of k connections, there exists a link-utilization vector $\vec{x}$ and end-to-end rate vector $\vec{f}$ such that $(\vec{x}, \vec{f})$ is feasible and simultaneously approximates any canonical utility function to within a factor of $O(q \log knR)$.*

PROOF. We start with the convex program in Theorem 1 with the necessary conditions (8), and consider the convex set consisting of feasible end-to-end rate vectors for that program. Applying Theorem 4 on this convex set, we get a solution $\vec{f}$ that is globally $O(\log \frac{P_k^*}{kP_1^*})$-fair. First, observe that $P_k^* = \max_x\{\sum_i x_i\}$, which is simply the maximum total end-to-end throughput. Trivially, we have $P_k^* \leq n^2(\max_e c(e))$. To lower bound P_1^*, consider any feasible flow that sends $f_i = (\min_e c(e))/kn^2$ on each connection i. Let $x(e) = \sum_i \sum_{(p \in \mathcal{P}_i)\&(p \ni e)} f(p)/c(e)$ be the link-utilization defined by such a flow. Then, $x(e) \leq 1/n^2$, since there are k flows, and therefore, the vector $\vec{x}$ will satisfy the necessary conditions for feasibility for all links. Since $P_1(\vec{f}) = (\min_e c(e))/kn^2$ for this flow, we have $P_1^* \geq (\min_e c(e))/kn^2$, and therefore, Theorem 4 implies that there is a solution $(\vec{x}, \vec{f})$ that satisfies all the feasibility conditions (8), and is $O(\log knR)$-fair. By Theorem 1, it follows that the link-utilization vector $\frac{\vec{x}}{q}$ satisfies the suffient condition for scheduling: in this case, for any j, $P_j(\frac{\vec{x}}{q}) = \frac{P_j(\vec{x})}{q}$; therefore, the vector $(\frac{\vec{x}}{q}, \frac{\vec{f}}{q})$ is a simultaneous $O(q \log knR)$-approximation for any canonical utility function. ☐

4.1 Distributed Algorithms

From an algorithmic point of view, a more important question is to compute the rate vector in Theorem 5. In the case where the feasible solutions satisfy some set of linear constraints, denoted by $Ax \leq c$, Goel and Meyerson [9] showed that P_s^* can be computed by the following linear program: minimize λ_s subject to: (i) $Af \leq \lambda_s c$, and (ii) $\forall S \subseteq \{1, \ldots, k\}$ such that $|S| = s$, $\sum_{i \in S} f_i \geq 1$. If λ_s^* is the optimum for this program, it can be shown that $\lambda_s^* = 1/P_s^*$. By solving n such programs, it is possible to find a solution that gives the bounds of Theorem 4. However, this is not feasible in practice. Cho and Goel [7] develop a much more efficient algorithm for the simultaneous optimization problem in the context of bandwidth optimization in wire line networks - they develop a dual-update algorithm for computing the solution to any individual LP, and then show how to combine all the n different dual-update algorithms into a single one. We show how their framework can be modified to compute the rate vector of Section 4 in the presence of wireless interference constraints.

In this section, we assume that we are given a fixed path p_i corresponding to connection i. We also assume the unit disk model here. We can simplify the formulation in Section 3 to one using just the variables $f(p_i)$. The Cho and Goel result uses the framework of Plotkin, Shmoys and Tardos (PST) [27] to solve the above LP combinatorially. Using the PST framework requires making repeated calls to the following program

$$\beta_s(y) = \max P_s(x)$$
$$Cx = 1, x \geq 0,$$

where $C = y^t A$ represents the dual costs - the cost C_i of flow on p_i is defined as $\sum_{e:N_{\geq}(e) \cap p_i \neq \phi} y(e)$. We briefly discuss how this dual program can be solved in a distributed manner in a wireless setting. For any s, optimal solutions to this program can be characterized very easily [7]: $\forall i, j, C_i \leq C_j \Rightarrow x_i \geq x_j$, and there is a value γ such that $\forall i, x_i \in \{0, \gamma\}$. Clearly, finding the optimum solution x requires finding the index i_0 such that $x_i = \gamma, \forall i \leq i_0$ and $x_i = 0, \forall i > i_0$. Given the dual edge costs, $y(e)$, the quantities C_i are easy to compute locally - since we have assumed a unit disk graph model, every edge $e \in p_i$ can collect this information from all edges $e' \in N(e)$, where $N(e)$ is the set of edges interfering with e. While the optimum solution x^s for the dual program for each s is different, Cho and Goel show that one can instead use the solution $\bar{x}$ that dominates all the x^s's, i.e., $\bar{x}_i = \max_s\{x_i^s\}$, with a small penalty. By scaling the solution appropriately, we will assume that we have an approximate solution such that for each edge e, $\sum_{p_i : e \in p_i} x(p_i) \leq c(e)$, and there exists an edge $e \in E$ s.t. $\sum_{p_i : e \in p_i} x(p_i) = c(e)$.

We now describe how to modify the algorithm of [7] for computing the rates f. We are given a parameter ϵ, which can be a constant. We maintain dual variables $y_t(e)$ for each edge e, and each iteration t; initially, $y_0(e) = \delta/c(e)$, where $\delta = m^{-1/\epsilon}$, m being the number of edges. In phase t, we keep track of $D(t) = \sum_e y_t(e)c(e)$, and the phases are performed while $D(t) < 1$. The steps in one such phase t are:

1. Using the dual edge length function, y_t, we compute the scaled dual solution $x(t)$, as described above.

2. $x = x + x(t)$

3. For each edge e, we update its length $y_{t+1}(e) = y_t(e)(1 + \epsilon \sum_{i:N(e) \cap p_i \neq \phi} x(p_i)/c(e))$

The last step above is the only difference from [7], who only consider the flow through edge e, and not the flow on all interfering edges. The above algorithm is clearly local and distributed in the sense described earlier - each flow agent only needs information from edge agents on edges close to the path, and from other flow agents. Let T be the final iteration when this procedure stops. The arguments of [7] give us the following results on the performance of this algorithm.

LEMMA 2. *Let G be a unit disk graph with n nodes and m edges. Let there be k connections, with path p_i specified for connection i. Then: the flow $x(T)/\log_{1+\epsilon} 1/\delta$ is feasible ,i.e., it satisfies all the feasibility conditions of Section 3, the number of iterations of the algorithm is $T = O(m \log m)$, and the final solution is an $O(\log n + \log R)$-majorized solution, where $R = \frac{\max_e c(e)}{\min_e c(e)}$. Thus, these flow rates approximate any canonical utility function within an $O(\log^{O(1)} n)$ factor.*

5. CONCLUSION

We have studied the questions of designing near optimal algorithms for joint rate control, scheduling and routing in multi-hop wireless networks. The algorithms were based on non-linear programming formulations and provided provable worst-case performance guarantees for a number of interference and physical layer models. In addition the techniques were extensible to handle other linear constraints on the user rates; examples include, bounds on the path length, energy consumption etc. A number of questions remain open. First, it would be interesting to design fully-distributed algorithms for the problems considered in this paper; they will form the first step towards designing cross layer aware protocols with provable *worst case* performance guarantees. Second, most of the work presented here assumed a multi-path routing; extensions of our results to single-path routing is of interest.

Acknowledgments. The research of V. S. A. Kumar, and M. V. Marathe was supported in part by NSF Award CNS-0626964. S. Parthasarathy's research was supported in part by NSF Award CCR-0208005 and NSF ITR Award CNS-0426683. A. Srinivasan's research was supported in part by NSF Award CCR-0208005, NSF ITR Award CNS-0426683, and NSF Award CNS-0626636. We also thank the referees for their valuable suggestions.

6. REFERENCES

[1] Mansoor Alicherry and Randeep Bhatia. Joint Channel Assignment and Routing for Throughput Optimization in Multi-radio Wireless Mesh Networks *Proc. 11th ACM MOBICOM*, 2005.

[2] D. J. Baker, J. E. Wieselthier, and A. Ephremides. A distributed algorithm for scheduling the activation of links in self-organizing mobile radio networks. In *IEEE Int. Conference Communications*, pages 2F6.1–2F6.5, 1982.

[3] C. L. Barrett, M. Drozda, A. Marathe, and M. V. Marathe. Characterizing the interaction between routing and MAC protocols in ad-hoc networks. In *Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing*, pages 92–103, 2002.

[4] B. N. Clark, C. J. Colbourn, and D. S. Johnson. Unit disk graphs. *Discrete Math*, 86:165–177, 1990.

[5] L. Chen, S. H. Low, M. Chiang, and J. C. Doyle. Jointly optimal congestion control, routing, and scheduling for wireless ad hoc networks. Manuscript, July 2005.

[6] M. Chiang. Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control. *IEEE J. Sel. Areas Comm.*, pages 104–116, Jan 2005.

[7] S. Cho and A. Goel. Bandwidth allocation in networks: a single dual-update subroutine for multiple objectives. In *Workshop on Combinatorial and Algorithmic Aspects of Networking (CAAN)*, 2004.

[8] Naveen Garg and Jochen Koenemann. Faster and simpler algorithms for multicommodity flow and other fractional packing problems. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, page 300. IEEE Computer Society, 1998.

[9] A. Goel and A. Meyerson. Simultaneous optimization via approximate majorization for concave profits or convex costs. Tech. Report CMU-CS-02-203, December 2002, Carnegie-Mellon University.

[10] A. Gräf, M. Stumpf, and G. Weißenfels. On coloring unit disk graphs. *Algorithmica*, 20:277–293, 1998.

[11] Piyush Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, 46(2):388–404, March 2000.

[12] B. Hajek and G. Sasaki. Link scheduling in polynomial time. *IEEE Transactions on Information Theory*, 34:910–917, 1988.

[13] Kamal Jain, Jitendra Padhye, Venkata N. Padmanabhan, and Lili Qiu. Impact of interference on multi-hop wireless network performance. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 66–80. ACM Press, 2003.

[14] M. Johansson and L. Xiao. Cross-layer optimization of wireless networks using nonlinear column generation. To appear in *IEEE Transactions on Wireless Communications.*

[15] K. Kar, M. Kodialam, T. V. Lakshman, and L. Tassiulas. Routing for network capacity maximization in energy-constrained ad-hoc networks. In *IEEE INFOCOM*, 2003.

[16] Murali Kodialam and Thyaga Nandagopal. Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problem. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 42–54. ACM Press, 2003.

[17] V. S. Anil Kumar, Madhav V. Marathe, Srinivasan Parthasarathy, and Aravind Srinivasan. End-to-end packet scheduling in ad hoc networks. In *ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1014–1023, 2004.

[18] V. S. Anil Kumar, Madhav V. Marathe, Srinivasan Parthasarathy, and Aravind Srinivasan. Algorithmic aspects of capacity in wireless networks. In *ACM SIGMETRICS*, pages 133–144, 2005.

[19] S. H. Low, L. Chen and J. C. Doyle. Joint congestion control and media access control design for wireless ad hoc networks. In *IEEE Infocom*, 2005.

[20] S. H. Low and M. Chiang. Nonlinear Optimization of Communication Systems. Tutorial presented at *IEEE Infocom* 2005.

[21] Frank Kelly, Aman Maulloo and David Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society,* 49 (1998) 237-252.

[22] S. Krumke, M.V. Marathe and S.S. Ravi. Models and Approximation Algorithms for Channel Assignment in Radio Networks, *Wireless Networks*, 7, pp 575-584, 2001.

[23] F. Kuhn, R. Wattenhofer and A. Zollinger. Ad-Hoc Networks Beyond Unit Disk Graphs, *ACM Joint Workshop on Foundations of Mobile Computing (DIALM-POMC),* 2003.

[24] F. Kuhn, T. Moscibroda, and R. Wattenhofer. Unit Disk Graph Approximation, ACM Joint Workshop on Foundations of Mobile Computing (DIALM-POMC), Philadelphia, Pennsylvania, USA, October 2004.

[25] X. Lin and N. B. Shroff. The impact of imperfect scheduling on cross-layer rate control in multihop wireless networks. In *IEEE INFOCOM*, 2005.

[26] X. Lin and N. B. Shroff. Joint Rate Control and Scheduling in Multihop Wireless Networks. *43rd IEEE Conference on Decision and Control, Paradise Island,* Bahamas, December 2004.

[27] S. Plotkin, D. Shmoys, and É. Tardos. Fast approximation algorithms for fractional packing and covering problems. *Math. Oper. Res.*, pages 257–301, 1994.

[28] S. Ramanathan. A unified framework and algorithm for channel assignment in wireless networks. *Wireless*

Networks, 5(2):81–94, 1999.

[29] A. Sen and M. L. Huson. A new model for scheduling packet radio networks. *Wireless Networks*, 3:71–82, 1997.

[30] P.-J. Wan, K. M. Alzoubi, O. Frieder, Distributed Construction of Connected Dominating Set in Wireless Ad Hoc Networks, INFOCOM 2002.

[31] A. Srinivas and E. Modiano. Minimum energy disjoint path routing in wireless ad-hoc networks. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, pages 122–133, 2003.

[32] S. Toumpis and A. J. Goldsmith. Capacity regions for wireless ad hoc networks. *IEEE Trans. on Wireless Communications*, pages 736–748, July 2003.

[33] J. Wang, L. Li, S. H. Low and J. C. Doyle. Cross-layer Optimization in TCP/IP Networks *IEEE/ACM Trans. on Networking,* 13(3):582-568, June 2005.

[34] L. Xiao, M. Johansson, and S. Boyd. Simultaneous routing and resource allocation via dual decomposition. *IEEE Transactions on Communications*, 52(7):1136-1144, July 2004.

[35] Y. Yi and S. Shakkottai. Hop-by-hop congestion control over a wireless multi-hop network. In *IEEE Infocom*, 2004.