A Cross-Layer Optimization Framework for Multicast in Multi-hop Wireless Networks

Jun Yuan, Zongpeng Li, Wei Yu, Baochun Li
Department of Electrical and Computer Engineering
University of Toronto
{steveyuan@comm, arcane@eecg, weiyu@comm, bli@eecg}.toronto.edu

***Abstract*— Achieving optimal transmission throughput in data networks is known as a fundamental but hard problem. The situation is exacerbated in multi-hop wireless networks due to the interference among local wireless transmissions. In this paper, we propose a general modeling and solution framework for the throughput optimization problem in wireless networks. In our framework, data routing, wireless medium contention and network coding are jointly considered to achieve the optimal network performance. The primal-dual solution method in the framework represents a cross-layer optimization approach. It decomposes the original problem into data routing sub-problems at the network layer, and power allocation sub-problems at the physical layer. Various effective solutions are discussed for each sub-problem, verifying that our framework may handle the throughput optimization problem in an efficient and distributed fashion for a broad range of wireless network scenarios.**

I. Introduction

Multi-hop wireless networks consist of wireless nodes that communicate with each other by relaying data flows over multiple hops, without infrastructure support. Due to the broadcast nature of omni-directional antennas, geographically nearby transmissions interfere with each other.

In this paper, we study the problem of achieving optimal transmission throughput for multiple concurrent data sessions in multi-hop wireless networks. As unicast, broadcast and group communication sessions may all be viewed as (or transformed into) multicast sessions [1], we assume that a data session is a multicast session without loss of generality. A desirable solution to the problem of achieving optimal transmission throughput includes a routing strategy of data flows at the network layer, as well as a power allocation scheme that leads to high capacity at the physical layer.

As the main original contribution of this paper, we propose a general framework to model and solve the optimal throughput problem in wireless networks. At a high level, our framework maximizes the overall throughput subject to three groups of constraints: (1) dependence of overall throughput on per-link data flow rates, (2) dependence of per-link flow rates on link capacities, and (3) dependence of link capacities on per-node radio power levels. We present a general primal-dual method that iteratively solves two disjoint sub-problems and converges to the optimal solution of the original problem. The first sub-problem is a multi-hop flow routing problem at the network layer, and the second sub-problem is a power allocation problem at the physical layer. We further illustrate how each sub-problem can be solved efficiently under different wireless network scenarios. We also take different algorithmic perspectives: at the network layer, we discuss the problem with or without network coding; at the physical layer, we consider the dual optimization method, geometric programming, as well as game theoretic designs.

The general framework proposed in this paper represents a cross-layer optimization strategy. It balances the *demand* and *supply* of link bandwidth at the network and the physical layers, respectively. It also provides an optimal flow routing scheme and a corresponding optimal power allocation scheme in such a balanced state. Important properties and highlights of our solution framework are as follows.

- **Comprehensiveness:** It provides a comprehensive approach for system designers to incorporate networking factors from different layers into a joint optimization problem. Such joint consideration of network coding, data routing, and wireless interference is indeed necessary to approach optimal performance in multi-hop wireless networks.
- **"Divide and conquer" strategies:** It explores the underlying modular structure of the joint optimization problem from the perspective of primal-dual solutions. It applies the classic technique, Lagrange relaxation combined with subgradient optimization, to decompose the original complex problem into smaller sub-problems that are easier to solve. Each sub-problem resides in only one networking layer, it therefore provides a clear justification for layered protocol design.
- **Flexibility:** How each sub-problem is solved depends on the specific model and the available solution techniques, and is independent of the general framework that we propose. As new advances occur in networking technologies or in optimization algorithm design, the module for solving the corresponding sub-problem in our framework can be easily replaced to reflect such advances and to achieve better performance. As examples, the module for flow routing may be implemented as multicommodity flow routing, tree packing and network coding; and the module for power allocation can be computed with the dual optimization method, geometrical programming, and a *cooperative selfish* game.
- **Effectiveness:** Our solution framework transforms the overall throughput optimization into a sequence of simpler sub-problems. As long as the computation of each

individual sub-problem is optimal, efficient and/or distributed, the overall solution is guaranteed to be optimal, efficient, and/or distributed, respectively. We shall illustrate how optimality can be achieved or closely approached for each sub-problem, in an efficient and distributed fashion.

The remainder of this paper is organized as follows. We first discuss related work in Section II. In Section III, we propose the joint optimization framework and the layering approach, together with an efficient primal-dual algorithm to solve the problem. In Section IV, we discuss the module structure of sub-problems, point out several new techniques in network and physical layers, and show how the sub-problems are incorporated in the overall framework. We then present an example to illustrate the main concept in Section V. Finally, we discuss the limitations and extensions of this paper in Section VI, and conclude in Section VII.

II. Related work

In the general model of data networks, recent research in information theory discovers that routing alone is not sufficient to achieve maximum information transmission rates [2], [3]. Rather, encoding and decoding operations at relay nodes in addition to the sender and receivers are in general necessary in an optimal transmission strategy. Such coding operations are referred to as *network coding*. The pioneering work by Ahlswede, Cai, Li and Yeung [2] and Koetter and Medard [3] proves that, in a directed network with network coding, a multicast rate is feasible if and only if it is feasible for a unicast from the sender to each receiver. Li, Yeung and Cai [4] then shows that linear coding usually suffices in achieving the maximum rate.

With the assistance of network coding, the problem of achieving optimal throughput has been studied in undirected networks where each link has a known capacity, shared by flows in both directions. As shown by Li *et al* [1], [5], the problem of computing optimal throughput can be formulated as a linear optimization problem, and distributed algorithms can be designed to efficiently solve the problem by applying Lagrange relaxation and subgradient optimization. However, the assumption of fixed link capacities is not realistic in multi-hop wireless networks, where link capacities are subject to interference from other links in the neighborhood. If we assume that two transmissions can be simultaneously active if and only if neither node in one transmission is within the communication range of the other transmission [6] (also referred to as the *logical* interference model), then even optimizing link scheduling for a set of fixed multi-hop flows is NP-hard, since it essentially corresponds to the graph coloring problem.

A main contribution of this paper is to take the physical layer interference into account when solving the optimal throughput problem for multi-hop wireless networks. Towards this objective, we draw from previous studies related to power control in wireless networks and show how physical layer models can be incorporated in the overall problem of optimizing throughput, with the assistance of network coding. In particular, we take advantage of recent progress in solving non-convex problems using a dual optimization method as proposed by Yu *et al* [7] for Orthogonal Frequency Division Multiplex (OFDM) systems in which the non-convex problem is solved efficiently and globally in the dual domain, and the geometric programming approach proposed by Chiang [8] for Code Division Multiple Access (CDMA) systems in which the non-convex problem is transformed into a convex one under the assumption of a high signal to interference and noise ratio (SINR). Finally, we also propose an approximate but near-optimal solution for the interference channel based on game theory.

The main technique used in this paper is the method of dual decomposition for convex optimization problems. The dual decomposition technique is related to the duality analysis of TCP as a flow control protocol by Low [9] and Wang *et al* [10], in which network congestion parameters are interpreted as primal and dual optimization variables and the TCP protocol is interpreted as a distributed primal-dual algorithm. Our work is also related to the extension of the above work to multi-hop wireless networks by Chiang [8], in which power levels and TCP window sizes are jointly optimized, and a dual variable is used as a means for cross-layer optimization. In a related work, Johansson, Xiao, and Boyd have also carried out a similar convex optimization approach to jointly perform routing and resource allocation in wireless CDMA networks [11], [12], where a high SINR is assumed in order to guarantee convexity. In our previous work [13], we have also studied a dual method for the joint source coding, routing, and power allocation problem for sensor networks, where the focus is a lossy source coding problem in the application layer. All of the above work treat the multi-session unicast problem only. The main idea of the present work is to propose a similar framework for multicast problems in a network coding context.

For wireless multicast in ad hoc networks, Wu *et al* [14] studied the issue of network planning and solved an energy minimization problem with centralized control. They also adopt SINR to model contention at the physical layer, and their transmission plan involves a time sharing scheme among a set of selected power allocation states. Both the focus and solution approaches of our paper are different as compared to [14]. We target maximum throughput instead of minimum cost, and present a general solution framework that decomposes the optimization into different layers. We also target distributed solutions and employ physical layer models, which directly connects capacity rates to power allocation.

III. A Joint Optimization Framework

We now present a general framework to model and solve the problem of optimizing throughput in multi-hop wireless networks. We first give a high-level formulation of the optimization problem, which involves variables from both the network layer and the physical layer. We then show that Lagrange relaxation and subgradient optimization can be applied to decompose such an overall optimization problem into a sequence of smaller sub-problems, each only involving variables from either the network layer or the physical layer. Interactions between the two sub-problems are then discussed.

A. General Framework of Joint Optimization

The formulation of the throughput optimization problem is based on the following facts. First, throughput is realized by routing data flows from senders to receivers, therefore the achievable throughput for each data session is decided by the corresponding data flows of that session. Second, at each transmission link, the aggregated data flow rate can't exceed the effective capacity of that link. Third, the achievable link capacity is decided by SINR, which in turn is decided by the power level at all the senders.

Let $G = (V, E)$ be the network topology. Let S be the set of multiple data sessions supported in the network. Let $\mathbf{r} = \{r^i\}$ be the set of multicast rates for each session $i \in S$. Fix $\mathbf{r}$, let $\mathcal{N}(\mathbf{r})$ be the set of network flow rates that are needed to support $\mathbf{r}$. The flow rates $\mathbf{f}$ is a vector $\{f_l^i\}$, where i denotes the session index $i \in S$ and l denotes the link index $l \in E$. Let $\mathbf{p}_{\max}$ be the set of power constraints on each node in V. Let $\mathcal{C}(\mathbf{p}_{\max})$ be the set of achievable rates $\mathbf{c} = \{c_l\}$ that the physical layer can support on each link $l \in E$.

The throughput optimization problem can now be formulated as:

$$\begin{aligned} \max \quad & U(\mathbf{r}) \\ \text{s.t.} \quad & \mathbf{f} \in \mathcal{N}(\mathbf{r}) \\ & \mathbf{c} \in \mathcal{C}(\mathbf{p}_{\max}) \\ & \sum_{i \in S} f_l^i \le c_l, \quad \forall l \in E \end{aligned} \tag{1}$$

where we seek to maximize some concave utility function of the throughput vector $U(\mathbf{r})$. For example, a utility that leads to proportional fairness is $U(\mathbf{r}) = \sum_i U_i(r^i) = \sum_i \log(r^i)$. The constraint $\mathbf{f} \in \mathcal{N}(\mathbf{r})$ models the inter-dependence of achievable throughput $\mathbf{r}$ and the data flow routing scheme $\mathbf{f}$. The constraint $\mathbf{c} \in \mathcal{C}(\mathbf{p}_{\max})$ models the inter-dependence of link capacity vector $\mathbf{c}$ on the node power constraint $\mathbf{p}_{\max}$. The constraint $\sum_i f_l^i \le c_l$ reflects the fact that the aggregated flow rate at each link is bounded by link capacity. Here, i is the index of data sessions, and l is the index of links.

The detailed characterization of the regions $\mathcal{N}(\mathbf{r})$ and $\mathcal{C}(\mathbf{p}_{\max})$ are independent of our general formulation and will be discussed in the next section. We now proceed to investigate solution techniques that are applicable to our general problem formulation.

B. Decomposing the Problem

When both $\mathcal{N}(\mathbf{r})$ and $\mathcal{C}(\mathbf{p}_{\max})$ are convex regions, general convex optimization methods can be used to solve the overall optimization problem (1). However, such a solution does not take advantage of the special problem structure and it in general requires global information to be collected at a central point of computation. In this paper, we instead propose a general solution framework, within which the original problem is decomposed into smaller sub-problems, each of which can be solved efficiently and distributively. We start by relaxing the link capacity constraints $\sum_i f_l^i \le c_l$ and introduce prices into the objective function:

$$L = U(\mathbf{r}) + \sum_l \lambda_l \left[c_l - \sum_i f_l^i \right]. \tag{2}$$

Observe that the maximization of the Lagrangian above now consists of two sets of variables: network layer variables $(\mathbf{f}, \mathbf{r})$, and physical layer variables $(\mathbf{p}, \mathbf{c})$, where $\mathbf{p}$ is a set of link power constrained by node power budget $\mathbf{p}_{\max}$. More specifically, the Lagrangian optimization problem now decouples into two disjoint parts. The network layer part is a data flow routing problem:

$$\begin{aligned} \max \quad & U(\mathbf{r}) - \sum_l \lambda_l \sum_i f_l^i \\ \text{s.t.} \quad & \mathbf{f} \in \mathcal{N}(\mathbf{r}) \end{aligned} \tag{3}$$

The physical layer part is a power allocation problem:

$$\begin{aligned} \max \quad & \sum_l \lambda_l c_l \\ \text{s.t.} \quad & \mathbf{c} \in \mathcal{C}(\mathbf{p}_{\max}) \end{aligned} \tag{4}$$

Thus, the optimization framework naturally provides a layering approach to the wireless throughput optimization problem. The global maximization problem decomposes into two parts: routing at the network layer and power allocation at the physical layer. The power allocation problem ensures that the maximal capacity is provided in individual network links, while the routing problem ensures that underlying link support is efficiently utilized to maximize the multicast rates.

The decoupling of the network optimization problem also reveals that cross-layer design can be achieved in a theoretically optimal way. The dual variable λ plays a key role in coordinating the network layer *"demand"* and physical layer *"supply"*. In particular, the lth component of λ (λ_l) can be interpreted as the rate cost in link l. A higher value of λ_l signals to the underlying physical layer that more resources should be devoted to transporting the traffic in link l. At the same time, it signals to the upper network layer that transporting bits in link l is expensive and it provides incentive for the network layer to find alternative routes for traffic.

The key requirement that allows the decoupling of the network optimization problem into routing and resource allocation is the underlying convexity structure of the problem. Below, we first provide a justification for convexity, then propose a primal-dual algorithm that can be used to solve the joint network optimization problem efficiently.

C. The Role of Convexity

We start by observing that at the physical layer, $\mathcal{C}(\mathbf{p}_{\max})$ can always be made a convex region and $\mathcal{C}(\mathbf{p}_{\max})$ can always be made a convex function of $\mathbf{p}_{\max}$ via time sharing. If two sets of rates are both achievable under the same power constraint, then their linear combination must also be achievable by simply dividing the frequency (or time) into two sub-channels and transmitting using the two different strategies in the two sub-channels.

From a network routing point of view, the technique of network coding resolves the competition among flows by introducing *conceptual flows* [1]. The network coded routing region specifies three kinds of linear constraints for conceptual flows: (a) the maximum flow rate must be upper bounded by the link capacity; (b) the law of *flow conservation*, *i.e.*, the incoming conceptual flow rate is equal to the outgoing conceptual flow rate at a relay node for every given source-sink pair; and (c) the multicast rate must be less than or equal to the rate for each source-sink pair in every session. Hence, it is obvious that the constraint set of network coded routing region is convex.

D. The Primal-Dual Solution Framework

A direct implication of the convexity of the capacity region $C(\mathbf{p}_{\max})$ and the data routing region $\mathcal{N}(\mathbf{r})$ is that the joint network optimization problem (1) can be solved efficiently. Further, as strong duality holds, the optimization problem (1) can be solved via its dual. As the dual problem has a natural decomposition, it may be solved by solving its two sub-problems (3) and (4) individually and by updating the shadow price λ_l in succession.

More specifically, we now propose the following primal-dual algorithm that solves the entire network optimization problem:

Algorithm 1: Primal-Dual Algorithm:

1) Initialize $\lambda_l^{(0)}$
2) In primal domain, given $\lambda_l^{(t)}$, solve the following sub-problems:

$$\max_{\mathbf{f},\ \mathbf{r}} \quad \left\{ U(\mathbf{r}) - \sum_l \lambda_l \sum_i f_l^i \mid \mathbf{f} \in \mathcal{N}(\mathbf{r}) \right\} \tag{5}$$

$$\max_{\mathbf{p},\ \mathbf{c}} \quad \left\{ \sum_l \lambda_l c_l \mid \mathbf{c} \in \mathcal{C}(\mathbf{p}_{\max}) \right\} \tag{6}$$

3) In dual domain, update λ using the following rule:

$$\lambda_l^{(t+1)} = \lambda_l^{(t)} + \nu_l^{(t)} (c_l - \sum_i f_l^i) \tag{7}$$

4) Return to step 2 until convergence.

Theorem 1: Algorithm 1 always converges to the global optimum of the overall network optimization problem (1), provided that the step sizes $\nu^{(t)}$ is chosen to be sufficiently small.

Proof: We outline the proof here. The crucial ingredient that validates Algorithm 1 is convexity. The convexity of $\mathcal{N}(\mathbf{r})$ and $C(\mathbf{p}_{\max})$ ensures an efficient numerical solution for the network optimization problem (1). Define the dual objective function

$$g(\lambda) = \max_{\mathbf{r},\mathbf{f},\mathbf{p},\mathbf{c}} L(\mathbf{r}, \mathbf{f}, \mathbf{p}, \mathbf{c}, \lambda). \tag{8}$$

The maximization above can be decomposed into two sub-problems (5) (6). The solutions of the two sub-problems allow $g(\lambda)$ to be evaluated.

Now, by strong duality, the overall network optimization problem (1) is solved by the following dual minimization problem:

$$\min_{\lambda} g(\lambda) \tag{9}$$

It remains to show that the update steps (7) solve the dual minimization problem. This is due to the fact that the update steps are subgradient updates for λ. It is not difficult to show that $(c_l - \sum_i f_l^i)$ are subgradients for λ_l. Thus, as long as step sizes $\nu_l^{(t)}$ are chosen to be sufficiently small, the subgradient update eventually converges, and it converges to the global optimum of the overall network optimization problem. ■

Note that the primal-dual algorithm can be implemented in a distributed fashion, if the sub-problems have distributed solutions. This is true because in (7), the update of dual variable λ_l in the lth link only requires the local capacity c_l and the rates of local flows $\sum_i f_l^i$.

IV. Sub-problem Modules

We have so far proposed a primal-dual solution framework to solve the problem of achieving optimal throughput in multi-hop wireless networks using joint optimization across the network and the physical layers. It remains to show how routing sub-problem at the network layer and the power allocation sub-program at the physical layer are effectively solved. We investigate different alternative solutions in different network scenarios. At the network layer, we examine solutions with or without the assumption of network coding. At the physical layer, we discuss different algorithmic perspectives, including dual optimization, geometric programming, and cooperative selfish games. These alternatives consist of *modules* to be readily plugged into the general framework that we have proposed.

A. Network Layer Modules

1) Routing based on Multicommodity Flows: When network coding is not considered and data sessions are unicast sessions, the network layer is naturally modeled as a multicommodity flow problem. The first group of constraints in our general framework, which characterizes the dependence of overall throughput on link flow rates, is then a standard multicommodity flow polytope. The corresponding data routing sub-problem at the network layer can be solved using available solution techniques for multicommodity flow problems.

2) Routing based on Tree Packing: When network coding is not considered and data sessions are multicast or broadcast sessions, routing is achieved by data forwarding and replication at each wireless node. With these assumptions, each atomic data flow propagates along a tree that spans every node in the data session. The maximum achievable throughput can be computed by finding the maximum number of pairwise capacity-disjoint trees, in each of which the multicast group remains connected. Such optimization has a straightforward linear programming formulation with an exponential number of tree capacity variables. In the case of broadcast sessions, such a problem corresponds to the *spanning tree packing* problem, in which we can work on the dual and employ minimum spanning tree algorithms as the separation oracle. In the case of multicast sessions, the problem corresponds to the *Steiner tree packing* problem, where this approach does not work effectively. This is due to the fact that we need to solve the minimum Steiner tree problem in the dual, which is exactly as hard as the Steiner tree packing problem itself [15].

3) Multicast Routing with Network Coding: With the advantages of network coding [2], we can model the routing problem at the network layer as follows:

$$
\begin{aligned}
\max_{\mathbf{r},\mathbf{f},\mathbf{e}} \quad & U(\mathbf{r}) - \sum_l \lambda_l \sum_i f_l^i && (10)\\
\text{s.t.} \quad & r^i \leq \sum_{l \in \mathcal{I}(T_j^i)} e_l^{i,j}, && \forall i, \forall j, \forall\, T_j^i \in V\\
& e_l^{i,j} \leq f_l^i, && \forall i, \forall j, \forall\, l \in E\\
& \sum_{l \in \mathcal{O}(n)} e_l^{i,j} = \sum_{l' \in \mathcal{I}(n)} e_{l'}^{i,j}, && \forall i, \forall j, \forall\, n \in V \backslash \{s^i, T_j^i\}\\
& f_l^i \geq 0,\ e_l^{i,j} \geq 0,\ r^i \geq 0
\end{aligned}
$$

The first inequality represents the constraint that the ith session multicast rate r^i is less than or equal to the sum of all the conceptual flow rates from source s^i to each of its jth destination T_j^i. Here, $e_l^{i,j}$ is the conceptual flow rate on link l in the ith multicast session to its jth destination T_j^i. The second inequality means that the actual flow rate f_l^i of session i on link l is the maximum of all the conceptual flows from source to destinations in that session. It is the advantage of network coding that allows us to consider the maximum, rather than the summation, of the conceptual flow rates. The third equality constraint represents the law of flow conservation for conceptual flows, where $\mathcal{I}(n)$ is defined as the set of links that are incoming to node n; and $\mathcal{O}(n)$ is the set of links that are outgoing from node n.

Theorem 2: For a data network with multiple multicast sessions, the maximum utility and its corresponding optimal routing strategy can be computed efficiently and in a distributed fashion.

Proof: First, the utility function is a concave function. Second, it is easy to prove that the network coding region constraint (linear constraint) is a convex set. Therefore, solving the sub-problem (10) is a convex optimization problem, which can be solved efficiently. A distributed solution is also possible by further relaxation. For more detailed discussions, refer to the conceptual flow approach presented in [1] [5]. ■

B. Physical Layer Modules

Interference management is one of the main challenges in physical layer design of wireless networks. A key concept in physical layer is the capacity region (rigorously speaking, the *achievable rate region*), which characterizes a tradeoff between achievable rates at different links. The capacity region optimization problem in the physical layer may be formulated as follows

$$
\begin{aligned}
\max_{\mathbf{p}} \quad & \sum_l \lambda_l c_l && (11)\\
\text{s.t.} \quad & c_l = \log\left(1 + \text{SINR}_l\right) && \forall l\\
& \text{SINR}_l = \frac{G_{ll} p_l}{\sum_{j \neq l} G_{lj} p_j + \sigma_l^2} && \forall l\\
& \sum_{l \in \mathcal{O}(n)} p_l \leq p_{n,\max} && \forall n
\end{aligned}
$$

where c_l is the capacity of link l, SINR_l is the signal to interference and noise ratio of link l, G_{ll}, p_l, and σ_l^2 are the link gain, power, and noise, respectively. G_{lj} is the interference gain from link j to link l. Each node has a power budget $p_{n,\max}$.

Because of the interference, the power allocation problem (11) is a non-convex optimization problem which is inherently difficult to solve. In this subsection, we discuss three recent techniques to ease the way of characterizing the feasible capacity region: the dual optimization method, geometric programming, and the game theoretic approach.

1) Dual Optimization Method: The main idea of the *dual optimization method* [7] is to recognize that although the constraints of (11) are not convex by itself, the time-sharing version of these constraints always is. Further, the time-sharing version of the problem can sometimes be more efficiently solved in the dual domain by solving the Lagrangian dual problem

$$
g(\mu) = \max_{\mathbf{p}} \sum_l \lambda_l c_l + \sum_n \mu_n \left(p_{n,\max} - \sum_{l \in \mathcal{O}(n)} p_l \right),
$$

for each fixed $\boldsymbol{\mu}$ and by adjusting $\boldsymbol{\mu}$ via a subgradient update. Optimizing for $g(\mu)$, although still not trivial, is often easier than solving the original problem, as $g(\mu)$ is always convex. (The exact evaluation of $g(\mu)$ may still take exponential complexity. However, efficient and sub-optimal algorithms can often be used to evaluate $g(\mu)$ approximately.) Under the time (frequency) sharing condition, the minimal value of $g(\mu)$ over all positive μ's is equal to the optimal solution of (11)

$$
\min_{\boldsymbol{\mu}} g(\mu) = \max_{\mathbf{p}} \sum_l \lambda_l c_l
$$

As in practical physical-layer system design, time-sharing can often be implemented either directly or via frequency sharing using (for example with OFDM modulation), the above method gives an efficient way to solve the capacity region maximization problem (11).

2) Geometric Programming: Recent developments in geometric programming show that, in high SINR scenarios, solving the problem (11) can be efficiently accomplished by convex programming [8].

The idea is to first approximate the link capacity rate $c_l = \log(1+\text{SINR}_l) \approx \log(\text{SINR}_l)$ if the SINR is much larger than 1. Then by logarithmic transformation of power vector $\tilde{p}_l = \log(p_l)$, the transformed problem (12) is a convex function over variables $\tilde{p}$:

$$
\begin{aligned}
\max_{\tilde{p}} : \quad & \sum_l \lambda_l \tilde{c}_l && (12)\\
\text{s.t.} : \quad & \tilde{c}_l = \log\left(\widetilde{\text{SINR}}_l\right) && \forall l\\
& \widetilde{\text{SINR}}_l = \frac{G_{ll} e^{\tilde{p}_l}}{\sum_{j \neq l} G_{lj} e^{\tilde{p}_j} + \sigma_l^2} && \forall l\\
& \sum_{l \in \mathcal{O}(n)} e^{\tilde{p}_l} \leq \tilde{p}_{n,\max}, && \forall l, \forall n
\end{aligned}
$$

Furthermore, Chiang [8] has proposed a distributed power allocation algorithm with gradient step size κ as follows:

$$p_l^{(t+1)} = p_l^{(t)} + \kappa \left(\frac{\lambda_l}{p_l^{(t)}} - \sum_{s \neq l} \frac{\lambda_s G_{ls}}{G_{ss} p_s^{(t)}} \mathrm{SINR}_s^{(t)} \right)$$

3) The Game Theoretic Approach: In this section, we explore ways to approximate the solutions of the non-convex achievable rate maximization problem by game theory. In a power control game, each link is modeled as a player with an aim of maximizing its utility function. The main idea here is to design a set of utility functions so that the competitive equilibrium of the game is approximately the global optimum. As reaching the competitive equilibrium of a game is typically computationally efficient and amenable to distributed implementation, this gives us an effective means of approximately solving the physical layer power control problem. This type of games is inherently different from traditional selfish games in that by *pretending* to be selfish, nodes actually help achieve the joint social welfare. Such a game is referred to as a cooperative selfish game.

In conventional game theoretical approaches for power control [16], each link uses its own achievable rate as its utility function (while treating all other users as interference). Competitive equilibria in such a game may not correspond to desirable operating points, especially when the interference level is high. This is typified by the well-known prisoner's dilemma. The main idea of this section is to modify the utility function of each link to include not only its own achievable rate but also the detrimental effect of the interference each link causes on other links. Under these modified fictitious utility functions, each link then has an incentive to settle on a power level that strikes a balance between maximizing its own rate and minimizing the interference caused to others.

In a distributed implementation, the amount of interference caused by each link has to be estimated by its neighbors. This motivates us to propose a message-passing mechanism with which the interference information can be communicated between the links via a side channel. Mathematically, we propose the following utility function for each link l:

$$Q_l = \lambda_l \log\left(1 + \frac{G_{ll} p_l}{\sum_{j \neq k} G_{lj} p_j + \sigma_l^2}\right) - m_l p_l - \mu_n p_l \quad (13)$$

where m_l is the dual variable summarizing the effect of interference from all other links and μ_n is the dual variable that indicates the price of transmitter power at node n. A sensible choice for m_l is the derivative

$$m_l = -\frac{\partial \sum_{s \neq l} c_s}{\partial p_l}. \quad (14)$$

In other words, m_l is the rate at which other users' achievable data rates decrease with an additional amount of power. The power price μ_n reflects how tight the resource at node n is being utilized by its outgoing links under the constraint $\sum_{l \in \mathcal{O}(n)} p_l \leq p_{n,\max}$.

The following algorithm implements this game theoretical power control scheme where each link player selfishly maximizes Q_l, while continuously updating the messages.

Algorithm 2: Message Passing Game Algorithm

1) Initialize $p^{(0)}$, $m^{(0)}$, $\mu^{(0)}$. Set $t = 0$.
2) Set $p^{(\tau_0)} = p^{(t)}$. Set $i = 0$, iteratively update $p^{(\tau_i)}$ as follows:

$$p_l^{(\tau_{i+1})} = \frac{\lambda_l}{m_l^{(t)} + \mu_n^{(t)}} - \sum_{j \neq l} \frac{G_{lj}}{G_{ll}} p_j^{(\tau_i)} - \frac{\sigma_l^2}{G_{ll}}$$

Repeat until $p^{(\tau_i)}$ converges. Set $p^{(t+1)} = p^{(\tau_i)}$.

3) Update price μ_n via subgradient

$$\mu_n^{(t+1)} = \mu_n^{(t)} + \gamma_n^{(t)} \left(\sum_{l \in \mathcal{O}(n)} p_l^{(t)} - p_{n,\max} \right)$$

4) Update the message m_l using (14)

$$m_l^{(t+1)} = \sum_{s \neq l} G_{ls} \lambda_s \frac{\mathrm{SINR}_s^{(t+1)}}{G_{ss} p_s^{(t+1)}} \frac{\mathrm{SINR}_s^{(t+1)}}{1 + \mathrm{SINR}_s^{(t+1)}}$$

5) Repeat (2-4) until convergence.

The power update in steps (2) is based on the following. At each step, each player tries to maximize its own utility Q_l while assuming the power levels for all other players and the messages are fixed. The expression for optimal p_l is obtained by setting the derivative Q_l with respect to p_l to zero. Such a locally optimal p_l strikes a balance between maximizing its own rate and minimizing its effect for other users (which is taken into account via m_l). For example, a large value for m_l indicates that link l is producing severe interference to other links. This is reflected in the power update as a large m_l leads to a lower p_l. Similarly, the value of the pricing variable μ_n indicates the tightness of the per-node power constraint. A high value for μ_n signals that the supply for power is tight and it entices link l to reduce its power. Finally, the demand for link capacity from the upper layer is reflected in the network layer shadow price λ_l. A large value for λ_l indicates that a higher capacity rate in the lth link is needed to support upper layer traffic, and it prompts the physical layer to increase its power.

Although each player appears to be selfish in maximizing its own utility only, because the utility function incorporates social welfare, the Nash equilibrium of this game is in fact a cooperative social optimum.

The message-passing algorithm can be implemented in a distributed fashion. This is because the messages can be locally collected and broadcast to the neighbors of each link. Although this power control game does not necessarily converges to the global optimum, experimental evidence suggests that it performs very well in practice. Finally, we have the following result on the conditions for convergence.

Theorem 3: Algorithm 2 always converges, and it converges to a stable Nash equilibrium of the message passing game, if the absolute value of eigenvalues of dynamics stability matrix are less than one.

Due to space constraints, a detailed analysis on the existence, uniqueness, stability and efficiency loss for the Nash equilibrium in this message passing game is omitted.

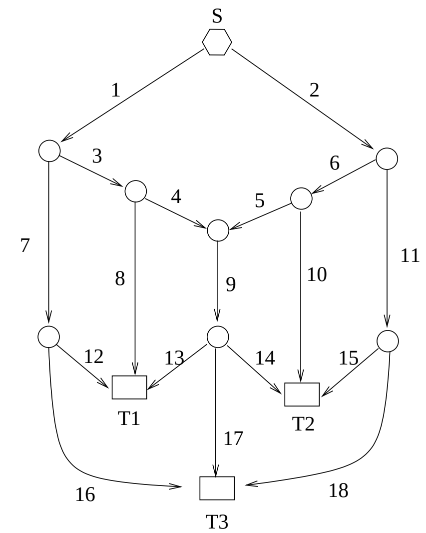


Fig. 1. The network topology.

V. Illustrative Examples

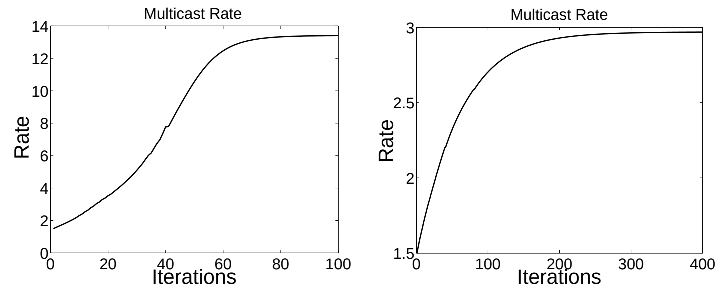
We illustrate an example of a single session multicast in Fig. 1. The source S attempts to establish a session with maximum multicast throughput to the three sinks T_1, T_2, T_3, with network coding. The model for the physical layer is an interference channel. An OFDM transmission scheme is simulated. The channel gain and interference are randomly generated. We define interference degree (ID) as the average ratio between the desired link gain and the sum of all interference gains. Both the low interference case (ID = 12dB) and the high interference case (ID = -2dB) are investigated.

We use the proposed primal-dual algorithm to achieve the optimal solution for the joint routing and power allocation problem (1). Specifically, a distributed message passing game algorithm is employed in order to find the achievable rate region.

Fig. 2 illustrates the multicast rate maximization process. In both high and low interference cases, the multicast rate continuously converges to the optimal solution. However, the convergence speed is different. Convergence is much faster in the low interference scenario (*i.e.*, 60 iterations) than in high interference scenario (*i.e.*, 200 iterations). This is because that in the low interference case, it is not very important for the links to exchange messages in order to arrive at a good tradeoff. Instead, each link simply maximizes its own link capacity to support the network traffic.

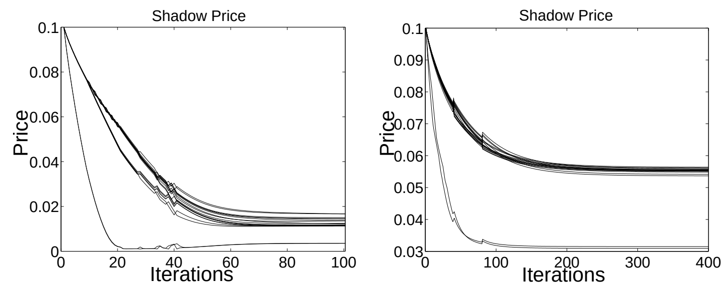
The convergence process for the cross-layer dual variables is illustrated in Fig. 3. The dual variables (shadow price) control the inter-layer interface so that both routing in the network layer and power allocation in the physical layer can reach an optimal matching point. As the shadow prices converge, the entire system reaches an optimal solution.

Fig. 4 shows the matching process between the network layer flows and the physical layer rates. At the beginning, the network flows oscillate in order to find a good routing strategy for each set of physical-layer rates. At the same time, physical layer rates increase and decrease among themselves in order to support network layer traffic. This process is coordinated by shadow prices as shown in Fig. 3. Eventually, the network flows and capacity rates agree. This solution is optimal, in the sense that the physical layer comes up with the best resource allocation while the network layer routes the best paths from the source to multiple sinks. Together, the multicast rate utility function is maximized.



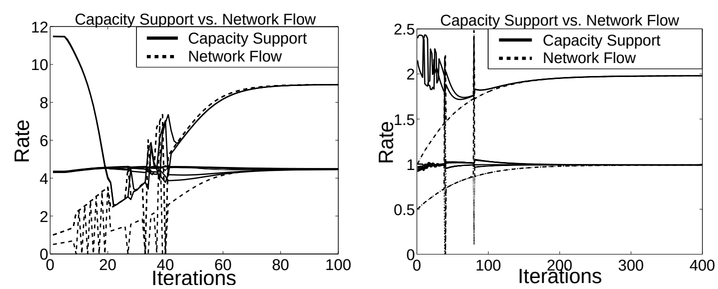
(a) Low Interference (b) High Interference

Fig. 2. Convergence of the multicast rate.



(a) Low Interference (b) High Interference

Fig. 3. Convergence of cross-layer dual variables.



(a) Low Interference (b) High Interference

Fig. 4. Convergence between network flow rates and capacity support.

For example, for the network in Fig. 1, the final solution in the high interference case, network flows in each link are $f_1 = f_2 = 1.98,\ f_3 = ... = f_{18} = 0.99$; the capacity rates are $c_1 = c_2 = 1.98,\ c_3 = ... = c_{18} = 0.99$; and the multicast rate is $r = 2.97$. Finally, it is interesting to point out that our algorithm is energy efficient because there is no wasted capacity rates in the system. All capacity rates exactly support the network flow ($f_l = c_l$). Consequently, all the shadow prices are non-zero, which means the capacity constraints are all active in the original problem (1).

This solution has a max-flow min-cut interpretation. If we normalize the throughput, the optimal flow and capacity rates will all be one unit except for link 1 and 2, where it is

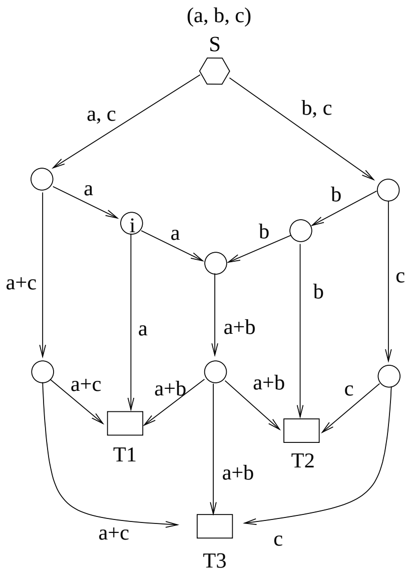


Fig. 5. Transmission scheme with network coding.

two units. As we can see from the optimization solution, the source can send three units of information in total to each sink, where the max-flow rate is exactly equal to the min-cut bound as shown in Fig. 5. We further show a network coding scheme to achieve this. In our example, source S has three units (a, b, c) to send, and each of the sinks (T_1, T_2, T_3) can exactly receive them, by using the flow solution and coding scheme as illustrated in Fig. 5.

VI. Limitations and Extensions

This paper proposes a general modeling and solution framework for the throughput optimization problem for multi-hop wireless networks. In this section, we point out several limitations and possible extensions of the current framework.

- Inter-session network coding for multiple data sessions is not considered in our framework. However, inter-session coding provides only marginal throughput gains [1], while it renders the data routing sub-problem NP-hard. Thus, ignoring such possibilities seems justifiable.
- One of the main physical layer assumptions in this paper is that each link transmits and receives signals independently. Possibilities of multi-access, broadcast or relay communications are not considered. The model presented in this paper is realistic in an ad-hoc network where no synchronization between the nodes is possible and interference is always regarded as noise. However, when a moderate amount of node cooperation can be implemented (*e.g.*, as in [17]), the utilization of multiuser techniques in the physical layer is expected to provide further gains.

VII. Conclusions

In this paper, we have proposed a general framework for both modeling and solving the throughput optimization problem in multi-hop wireless networks. In our framework, network coding, data routing, and wireless interference can be jointly considered to achieve the overall optimal performance. Our solution framework decomposes the optimization problem into smaller sub-problems: data routing at the network layer and power allocation at the physical layer. Modeling and solution algorithms for each sub-problem can be easily tuned according to a specific networking technology, as well as available optimization techniques.

References

[1] Z. Li, B. Li, D. Jiang, and L. C. Lau, "On Achieving Optimal Throughput with Network Coding," *Proc. of IEEE INFOCOM*, 2005.
[2] R. Ahlswede, N. Cai, S. R. Li, and R. W. Yeung, "Network Information Flow," *IEEE Trans. on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.
[3] R. Koetter and M. Medard, "An Algebraic Approach to Network Coding," *IEEE Trans. on Networking*, vol. 11, no. 5, pp. 782–795, October 2003.
[4] S. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Trans. on Information Theory*, vol. 49, 2003.
[5] Z. Li and B. Li, "Efficient and Distributed Computation of Maximum Multicast Rates," *Proc. of IEEE INFOCOM*, 2005.
[6] H. Luo, S. Lu, and V. Bharghavan, "A New Model for Packet Scheduling in Multihop Wireless Networks," in *Proceedings of the 6th Annual ACM International Conference on Mobile Computing and Networking (MobiCom)*, 2000.
[7] W. Yu, R. Liu, and R. Cendrillon, "Dual Optimization Methods for Multiuser Orthogonal Frequency Division Multiplex Systems," *IEEE Globecom*, 2004.
[8] M. Chiang, "Balancing Transport and Physical Layers in Wireless Multihop Networks: Jointly Optimal Congestion Control and Power Control," *IEEE J. Sel. Areas Comm*, vol. 23, no. 1, January 2005.
[9] S. H. Low, "A Duality Model of TCP and Queue Management Algorithms," *IEEE/ACM Trans. on Networking*, vol. 11, no. 4, pp. 525–526, August 2003.
[10] J. Wang, L. Li, S. H. Low, and J. C. Doyle, "Cross-layer Optimization in TCP/IP Networks," *IEEE/ACM Trans. on Networking*, 2005.
[11] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous Routing and Resource Allocation for Wireless Networks," *Proc. of the Fourth Asian Control Conference*, 2002.
[12] M. Johansson and L. Xiao, "Cross-layer Optimization of Wireless Networks Using Nonlinear Column Generation," *Technical report IR-S3-REG-0302, Royal Institute of Technology, Sweden*, 2004.
[13] W. Yu and J. Yuan, "Joint Source Coding, Routing and Resource Allocation for Wireless Sensor Networks," *IEEE International Conference on Communications (ICC)*, May 2005.
[14] Y. Wu, P. A. Chou, Q. Zhang, K. Jain, W. Zhu, and S. Y. Kung, "Network Planning in Wireless Ad Hoc Networks: A Cross-Layer Approach," *IEEE J. Sel. Areas Comm*, vol. 23, no. 1, January 2005.
[15] K. Jain, M. Mahdian, and M. R. Salavatipour, "Packing Steiner Trees," in *Proceedings of the 10th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2003.
[16] W. Yu, G. Ginis, and J. Cioffi, "Distributed Multiuser Power Control for Digital Subscriber Line," *IEEE J. Sel. Areas Comm*, vol. 20, no. 5, pp. 1105–1115, 2002.
[17] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "On the Construction of Energy-Efficient Broadcast and Multicast Trees in Wireless Networks," *Proc. IEEE INFOCOM*, 2000.