QoS and Fairness Constrained Convex Optimization of Resource Allocation for Wireless Cellular and Ad Hoc Networks

David Julian, Mung Chiang, Daniel O'Neill and Stephen Boyd.

Abstract—

For wireless cellular and ad hoc networks with QoS constraints, we propose a suite of problem formulations that allocate network resources to optimize SIR, maximize throughput and minimize delay. The distinguishing characteristics of these resource allocation formulations is that, by using convex optimization, they accommodate a variety of realistic QoS and fairness constraints. Their globally optimal solutions can be computed efficiently through polynomial time interior point methods, even though they use nonlinear objectives and constraints.

Through power control in wireless cellular networks, we optimize SIR and delay for a particular QoS class, subject to QoS constraints for all other QoS classes. For wireless ad hoc networks with multihop transmissions and Rayleigh fading, we optimize various objectives, such as the overall system throughput, subject to constraints on power, probability of outage, and data rates. These formulations can also be used for admission control and relative pricing. Both proportional and minmax fairness can be implemented under the convex optimization framework, where fairness parameters can be jointly optimized with QoS criteria. Simple heuristics are also shown and tested using the convex optimization tools.

Index Terms— **Ad hoc Networks, Cellular Networks, Convex Optimization, QoS Constrained Resource Allocation, Fairness**

I. INTRODUCTION

As users of communication networks become less satisfied with best efforts transmission, Quality of Service (QoS) has become an important research and commercial issue. QoS covers a wide array of network attributes, including bandwidth, delay, and packet delivery guarantee. Voice, data, image, and video have different bandwidth requirements. Some classes of traffic, such as voice, are also much more sensitive to delays than other classes, such as data.

QoS provisioning in a wireless network is a particularly difficult problem due to the time varying and unreliable physical channel. We present a new framework of convex optimization as a computationally efficient tool for resource allocation, including power control and admission control, under QoS and fairness constraints.

In wireless cellular networks, power control can be used to control interference, and in doing so, indirectly control the QoS seen by users on the network. In the downlink, a mobile user can receive interfering transmissions form base stations in adjacent cells resulting in adjacent channel interference. In the uplink, a base station experiences adjacent channel interference from users in adjacent cells, and also co-channel interference from mobile users in the same cell interfering with one another. The Signal to Interference Ratio (SIR) is often used to capture the effect of both co-channel and adjacent channel interference, and is routinely used in this paper to characterize the QoS parameter of throughput of a particular link. Extending our work in [10], sections IV and V formulate the following problems $P1$ to $P3$ for wireless cellular networks and show that they can be solved using convex optimization techniques.

- $P1$ Determining feasibility of a set of SIR requirements.
- $P2$ Maximizing SIR for a particular class of users with lower bounds on the QoS of all other users.
- $P3$ Satisfying queuing delay requirements for users in various QoS classes.

Ad hoc wireless networks pose additional technical challenges for QoS support. Unlike cellular wireless networks, ad hoc networks have no cells or base stations, but are composed of a set of nodes that transmit, receive and relay information among each other. Packets traverse the network by multihop transmissions from node to node until arriving at the destination. Consequently user QoS requirements are transformed into a set of QoS link requirements for the hops taken from source to destination.

Extending the results for cellular networks to ad hoc networks and our work in [7], the following problems $P4$ to $P7$ are solved for multi-hop networks in sections VI and VII:

- $P4$ Finding the optimum power control to maximize overall system throughput consistent with QoS guarantees in a fading environment.
- $P5$ Determining feasibility of a set of service level agreements (SLA) under network resource constraints.
- $P6$ Solving for the minimum total transmission delay of the most time sensitive class of traffic by optimizing over powers, capacities, and SLA terms.
- $P7$ Maximizing the unused capacity of the network.

Apart from performance optimization, fairness is another important issue in QoS provisioning. We show that both propor-

Information Systems Laboratory,Electrical Engineering Department, Stanford University, CA 94305, USA. E-mail:{djulian, chiangm, dconeill, boyd} @Stanford.edu

This work was supported by NSF Grant CCR-9973134, Stanford Network Research Center, Hertz Foundation Fellowship, and Stanford Graduate Fellowship.

tional fairness and minmax fairness can be formulated within the convex optimization framework. A joint optimization of the fairness parameters and the QoS criteria can also be implemented under the proposed framework.

This paper unifies and extends our work in cellular networks [10] and ad hoc networks [7]. In addition to unifying results into one framework, the following new extensions are made. Admission control and pricing in the ad hoc network is described in detail with a simulation of the concepts, and delay constraints in the geometric programming framework are addressed. Also, the fairness weights are extended to the non-integer case, and it is shown that the weights can be used as optimization variables. Finally, two simple heuristics are described, and illustrate that the convex optimization formulations can be used as a benchmark to compare the performance of heuristics against.

II. Related Work

As an important special case of resource allocation in wireless networks, power control in cellular networks has been studied extensively, and various schemes have been proposed or adopted in 2G and 3G networks. For example, the classical Qualcomm power control is used to solve the near far problem in CDMA networks. We show in section IV that this particular power control scheme is a special case of the proposed general convex optimization framework.

Various iterative methods have been proposed to optimally maximize the minimum SIR, to minimize total or individual power, or to maximize throughput in [1], [3], [4], [6], [8], [9], [13], [15], [18], [19], but these methods are not general enough to allow a diverse set of QoS constraints and other objective functions. We present a new framework of resource allocation based on the computational efficiency and versatility of convex optimization, which strikes a balance between efficiently achieving global optimality and flexibly allowing different constraints, objectives and variables.

A similar idea has been used to minimize outage probability in cellular networks under Rayleigh fading without QoS constraints in [11]. We show that the convex optimization framework can be used to incorporate a variety of QoS constraints and objectives, not just for cellular networks, but for ad hoc networks as well.

III. Convex Optimization and Geometric Programming

We need efficient algorithms to find the optimal solution to nonlinear problems $P1$ to $P7$. Fortunately, these problems can be turned into convex optimization formulations, which have efficient polynomial time algorithms such as the primal dual interior point method.

Convex optimization refers to minimizing a convex objective function over convex constraint sets. The particular type of convex optimization we use is in the form of geometric program. [5]. Geometric programming focuses on monomial and posynomial functions.

Definition 1: A *monomial* is a function $f : \mathcal{R}^n \rightarrow \mathcal{R}$, where the domain contains all real vectors with non-negative components:

$$h(x) = cx_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}, \;\; c \geq 0 \; and \; a_i \in \mathcal{R} \tag{1}$$

Definition 2: A *posynomial* is a sum of monomials $f(x) = \sum_k c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}$.

Geometric program is an optimization problem with the following form:

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) & \leq \;\; 1 \\ & h_j(x) & = \;\; 1 \end{array} \tag{2}$$

where f_0 and f_i are posynomials and h_j are monomials. Geometric programming in the above form is not a convex optimization problem. However, with a change of variables: $y_i = \log x_i$ and $b_{ik} = \log c_{ik}$, we can put it into convex form:

$$\begin{array}{llllll} \text{minimize} & p_0(y) & = & \log \sum_k \exp(a_{0k}^T y + b_{0k}) & & \\ \text{subject to} & p_i(y) & = & \log \sum_k \exp(a_{ik}^T y + b_{ik}) & \leq & 0 \\ & q_j(y) & = & a_j^T y + b_j & = & 0 \end{array} \tag{3}$$

It can be verified that the logarithm of a sum of exponentials is a convex function. Therefore p_i are convex functions and q_j are affine functions, and we have a convex optimization problem. Note that if all posynomials are in fact monomials, geometric programming becomes linear programming.

Convex optimization problems can be solved globally and efficiently through the interior point primal dual method [14], with polynomial running times that are often $\mathcal{O}(\sqrt{N})$ where N is the size of the problem. Apart from computational efficiency, convex optimization also offer duality interpretations, stability analysis and accommodate a variety of constraints. Solution algorithms also unambiguously and efficiently determine feasibility. This paper shows how geometric programming can solve many versions of QoS provisioning and resource allocation problems in wireless cellular and ad hoc networks.

IV. Throughput Optimization for Wireless Cellular Networks

A. Problem formulations

We first consider power control in a wireless cellular network in this section. For notational simplicity, this section considers a single base station and N links. Extensions to multiple base stations and the associated links are straight forward. Each link is a unidirectional path from the transmitter to the receiver. The propagation model used in this section is as follows:

$$P_r = PK \left(\frac{d_0}{d} \right)^{\gamma} \tag{4}$$

where P_r is the received power, P is the transmitted power, d is the propagation path length, and d_0 is a reference distance for the antenna far-field, usually taken so that the normalization constant K equals 1. The path loss exponent γ is usually between 2 and 6 for most indoor and outdoor environments. The interfering users' powers are decreased by the inverse of K_s, which can be the spreading gain for a CDMA system or the power falloff with frequency for an FDMA system. Accordingly, SIR_i for the i^{th} link is defined as

$$SIR_i = \frac{P_i d_i^{-\gamma_i} \alpha_i}{\sum_{j \neq i}^{N} P_j K_s^{-1} d_j^{-\gamma_j} \alpha_j + n_i}, \quad (5)$$

where the factors α_j are introduced to accommodate normalization constants and other factors, such as the effects of beamforming in multiantenna systems. SIR is well justified to be used as a throughput QoS parameter. For example, channel capacity scales with $\log SIR$ for high SIR, and the probability of symbol decoding error for coherent MPSK and MQAM is approximately $\alpha_M Q(\sqrt{\beta_M \cdot SIR})$, where α_M and β_M depend on the modulation type and the Q function is the complementary Gaussian CDF.

The problem of SIR maximization can be formulated as a geometric program. In the following basic formulation, the SIR is maximized for a particular mobile i. At the same time QoS for the other mobiles should also satisfy certain requirements or constraints. The following four kinds of constraints are reflected in Formulation 1 below.

1) Interference due to users, including base stations and mobiles, in index set $I_{1,k}$ must be smaller than a positive constant c_k because their assigned QoS values are relatively low.
2) Interference due to users in index set $I_{2,k}$ has to be smaller than the received signal power for some mobile k so as to achieve a required SIR, β_k.
3) The received signal power for some mobile k needs to be exactly equal to a positive constant C_k.
4) As in the special case of the classical power control scheme to solve the near-far problem in CDMA, the received signal power for one mobile k_1 needs to be equal to that of another mobile k_2.

With the objective and constraints thus formulated and upper bounds $P_{i,UB}$ on all transmitted powers P_i included, we obtain the following non-linear optimization formulation:

Formulation 1: (SIR constrained optimization of power control) The following nonlinear problem of optimizing node powers to maximize SIR for a particular user under QoS constraints for other users in a cellular network is a convex optimization problem.

$$\begin{array}{lcl} \text{maximize} \quad SIR_i & = & \frac{P_i d_i^{-\gamma_i}\alpha_i}{\sum_{j\neq i}^{N} P_j K_s^{-1} d_j^{-\gamma_j}\alpha_j + n_i} \\ \text{subject to} & & \\ \sum_{j \in I_{1,k}} P_j d_j^{-\gamma_j}\alpha_j & < & c_k \\ \beta_k \sum_{j\in I_{2,k}} P_j K_s^{-1} d_j^{-\gamma_j}\alpha_j + \beta_k n_k & < & P_k d_k^{-\gamma_k}\alpha_k \\ P_k d_k^{-\gamma_k}\alpha_k & = & C_k \\ P_{k1} d_{k1}^{-\gamma_{k1}}\alpha_{k1} & = & P_{k2} d_{k2}^{-\gamma_{k2}}\alpha_{k2} \\ P_j & \leq & P_{j,UB} \quad \forall j \\ P_j & \geq & 0 \quad \forall j \end{array} \quad (6)$$

where the first four constraints are for all k in the appropriate index sets. While the objective function is not a posynomial, its inverse $ISR = \frac{1}{SIR}$ is, and minimizing ISR over the same constraints is equivalent to the original problem. The inequality constraints above are posynomials, since posynomials when divided by monomials are necessarily posynomials in the parameters P_i, d_i and α_i. The equality constraints are monomials in the same parameters. The variables are the transmitted powers P_i. Therefore, this is indeed a geometric program, and therefore a convex optimization problem with efficient algorithms that obtain the global optimality.

This general formulation can be applied to different power control situations. For example, if there is no objective function, the above formulation reduces to a SIR requirement feasibility problem. Also, the objective function can be replaced by "minimize $\sum_i P_i$" as in the following formulation, and then the minimum power vector under the QoS constraints can be determined.

Formulation 2: (SIR constrained optimization for minimum power)

The following nonlinear problem of minimum power allocation in a cellular network is a convex optimization problem.

$$\begin{array}{ll} \text{minimize} & \sum_i P_i \\ \text{subject to} & \text{Same constraints as in Formulation 1.} \end{array} \quad (7)$$

Additionally, a weighted sum or powers, or the maximum user power can be minimized. The d_i can also be treated as optimization variables for optimization of antenna sectoring, which is a popular technique for interference mitigation.

B. Interpretations of the QoS Constrained Power Control

The log-sum-exp function can be interpreted as a smooth approximation of the maximum function [5]:

$$\max(x_i) \leq \log \sum_i^n e^{x_i} \leq \log(n) + \max(x_i) \quad (8)$$

Therefore, the above convex optimization of power control is minimizing a smooth approximation of the maximum of

$$\log(\frac{P_j}{P_i}) + \log(\frac{d_i^{\gamma_i}}{d_j^{\gamma_j}}). \quad (9)$$

When $\gamma_i = \gamma_j$, this is a weighted sum of the difference in powers (measured in dB) and the difference in distance (also measured in dB) for users j and i.

The dual problem is a generalized entropy maximization [5]. By duality analysis, it can be shown that solving the QoS constrained power control problem is equivalent to finding the linear combiners of the Lagrangian function (the augmented objective function) with the maximum weighted sum of entropy, where the weights are induced by the constraints of the dual problem.

C. Proportional and Minmax Fairness Extensions

Fairness is another major issue in a QoS policy. Both proportional fairness and minmax fairness can be accommodated in the framework of geometric programming.

Formulation 3: (SIR constrained optimization with proportional fairness)

The following nonlinear problem of weighted fair power allocation in a cellular network is a convex optimization problem.

$$\begin{array}{ll} \text{maximize} & \sum_i w_i \log SIR_i \\ \text{subject to} & \text{Same constraints as in Formulation 1.} \end{array} \quad (10)$$

This extended version of power control for general $w_i \in \mathcal{R}$ is still a convex optimization problem in geometric program form because maximizing $\sum_i w_i \log SIR_i$ is equivalent to maximizing $\log \Pi_i SIR_i^{w_i}$, which is in turn equivalent to minimizing $\Pi_i ISR_i^{w_i}$. Since a product of posynomials is also a posynomial, both the objective function and constraints are posynomials if the weights w_i are integers.

While in general posynomials to noninteger powers can not be handled in the geometric programming framework, the structure of this problem allows for noninteger weights. Auxiliary variables t_i can be introduced, the objective function changed to $\min \prod_i t_i^{w_i}$, and the constraints $ISR_i \leq t_i$ added to the existing set of constraints for all i. The optimization variables are now t_i and P_i, and the objective function and constraints are posynomials in the optimization variables. Further, the value of the objective function and the optimizing powers are the same as in the original formulation.

The minimum SIR_i can also be maximized subject to QoS constraints for other users through convex optimization. This minimax algorithm is useful in situations where the worst case is of concern.

Formulation 4: (SIR constrained optimization with minmax fairness)

The following nonlinear problem of minmax fair power allocation in a cellular network is a convex optimization problem.

$$\begin{array}{ll} \text{maximize} & \min_i SIR_i \\ \text{subject to} & \text{Same constraints as in Formulation 1.} \end{array} \quad (11)$$

This is a geometric programming problem because minimizing $\left[\max_i \frac{1}{SIR_i}\right]$ is equivalent to minimizing over an auxiliary scalar variable t such that $\max_i ISR_i \leq t$, which is in turn equivalent to minimizing t such that $ISR_i \leq t \ \forall i$. So the auxiliary variable t acts as an upper bound on all ISR's. When minimized over all feasible P, the value of t is reduced until it achieves the minimax value.

D. SIR optimization simulation

A simple system comprised of five users is used for a simulation of Formulation 1. The setup is as follows. First, the five users are spaced at distances d of $1, 5, 10, 15$, and 20 units from the base station. The power drop off factor $\gamma = 4$, and $\alpha = 1$. Each user has a maximum power constraint of $P_{UB} = 0.5W$. The noise power is 0.5μW for all users. CDMA is used with a spreading gain of $K_s = 10$. The SIR of all users, other than the user we are optimizing for, must be greater than a common threshold SIR level β. β is varied to observe the effect on the optimized user's SIR. This is done independently for the near user at $d = 1$, a medium distance user at $d = 15$, and the far user at $d = 20$. The results are plotted in figure 1.

Several interesting effects are illustrated by this simulation. First, when the required threshold SIR for the non-optimized users is high there are no feasible power control solutions. At moderate threshold SIR, as β is decreased, the optimized SIR initially increases rapidly. This is because it is allowed to increase its own power by the sum of the power decrease in the four other users, and the noise is relatively insignificant. At low threshold SIR, the noise becomes more significant and the power trade-off from the other users less significant, so the curve starts to bend over. Eventually, the optimized user reaches its upper bound on power and cannot utilize the excess power allowed by the lower threshold SIR. Therefore, during that stage, the only gain in the optimized SIR is the lower power transmitted by the other users. This is exhibited by the sharp bend in the curve to a much shallower sloped curve. We also note that the most distant user in the constraint set dictates feasibility.

E. Admission Control and Pricing

As shown by the interpretation of the above simulation, convex optimization can also be used for admission control in wireless communication networks. A new user is only allowed admission into the network when a feasible solution of this geometric program exists.

The effect of adding a new user to the system could be used to establish pricing for that user. The concept is to charge more to users who consume more of the total system user capacity. Unlike fixed QoS systems, the effect of admitting a new user depends heavily on the QoS requirements of the new user. A user with a high QoS requirement will reduce system user capacity more than a user with easily supported QoS requirements and should be charged commensurately more. For example, a user seeking high data rates close to a group of other users will more adversely affect user capacity, through interference and power limitations, than a user seeking a lower data rate in a low interference region.

The effect on capacity can be modeled using geometric programming by determining the number of standardized users that could be added to the system both before and after the new user is admitted. The difference between these two numbers can be taken as the reduction in the system (standard) user capacity that results from admitting the new user to the system. The price could then be set as an linear function of this difference.

Under this approach, a user could experience different "spot" pricing at different times depending on the existing load on the system when the user sought to access the network. Spot pricing is different from the current rate based pricing approaches, and more realistically models the effect a user has on the network from a revenue potential point of view.

V. Queuing Delay Optimization for Wireless Cellular Networks

Delay can be an important part of QoS for a wireless cellular network. There are three main component of the overall delay: propagation delay, transmission delay and queuing delay. Queuing delay is particularly important for bursty digital data, where the short term data rate of the information to be transmitted may exceed the data rate supported by the wireless link. Buffering is used to address this short term imbalance. This

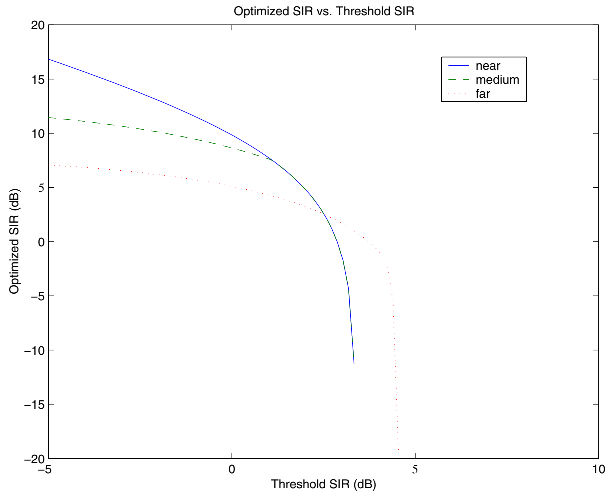

Fig. 1. Simulation results for constrained optimization of power control in a cellular network, with implications for admission control and pricing scheme.

queuing delay can dominate the propagation delay for reasonable link data rates.

By assuming that packets arrive according to a Poisson distribution and that packets are of variable length, the system can be modeled as an M/M/1 queue. The average queueing delay, D, can then be expressed as

$$D = \frac{1}{\mu(P) - \lambda} \tag{12}$$

where $\mu(P)$ is link transmission rate, or service rate, and λ is the arrival rate.

If a QoS agreement specifies an average delay bound D_b and an average maximum arrival rate, then this bound can be met by constraining the SIR on this link to exceed a minimum threshold, so that link transmission rate, as determined by the modulation type and the SIR, is larger than $\frac{1}{D_b} + \lambda$.

VI. Power control for throughput optimization in Wireless Ad Hoc Networks

In this section, we turn to power control in wireless ad hoc networks with multihop transmission. We will further expand the suite of formulations to more general resource allocation settings in the next section.

The formulation used in this section explicitly takes into account the statistical variation of the received signal and the interference power over a multi-hop network.

A. Multi-hop network model and Rayleigh fading

Consider a wireless ad hoc network with n transmitter/receiver pairs, labelled $1, \ldots, n$, which transmit at powers $P_1 \ldots, P_n$. The power received from transmitter j, at receiver i is given by

$$G_{ij} F_{ij} P_j \tag{13}$$

The nonnegative number G_{ij} represents the path gain in the absence of fading from the j^{th} transmitter to the i^{th} receiver. G_{ij} can encompass path loss, shadowing, antenna gain, coding gain, and other factors.

The Rayleigh fading between each transmitter j and receiver i is given by F_{ij}. The F_{ij}'s are assumed to be independent and have unit mean. The G_{ij}'s are appropriately scaled to reflect variations from this assumption. The distribution of the received power between any pair of transmitters j and receivers i is exponential with mean value,

$$E\left[G_{ij} F_{ij} P_j\right] = G_{ij} P_j \tag{14}$$

The signal to interference ratio (SIR) for user i now becomes

$$SIR_i = \frac{P_i G_{ii} F_{ii}}{\sum_{j \neq i}^{N} P_j G_{ij} F_{ij} + n_i} \tag{15}$$

B. Outage probability and system throughput

Due to multihop transmission over unreliable links, outage probability is an important QoS parameter in wireless ad hoc networks. An outage is declared when the received SIR falls below a given threshold defined as SIR_{th}, often computed from a BER requirement. Neglecting the noise in the high power interference limited case, the outage probability associated with the i^{th} hop is given by

$$\begin{aligned} O_i &= Pr(SIR_i \leq SIR_{th}) \\ &= Pr(G_{ii} F_{ii} P_i \leq SIR_{th} \textstyle\sum_{k \neq i} G_{ik} F_{ik} P_k) \end{aligned} \tag{16}$$

The outage probability can be expressed as [11]

$$O_i = 1 - \textstyle\prod_{k \neq i} \frac{1}{1 + \frac{SIR_{th} G_{ik} P_k}{G_{ii} P_i}} \tag{17}$$

Outage probability over a hop induces an outage probability over a path S

$$\begin{aligned} O_{pathS} &= \text{Prob(outage along the path S)} \\ &= 1 - \textstyle\prod_{s \in S} (1 - O_i) \\ &= 1 - \textstyle\prod_{s \in S} \prod_{k \neq s} \frac{1}{(1 + \frac{SIR_{th} G_{ik} P_k}{G_{ii} P_i})}. \end{aligned} \tag{18}$$

The constellation size M used by a hop can be closely approximated for MQAM modulation as follows

$$M = 1 + \frac{-1.5}{\ln(5BER)} SIR \tag{19}$$

where BER is the bit error rate. Defining $K = \frac{-1.5}{\ln(5BER)}$ leads to a monotonic expression for the data rate of the i^{th} hop as a function of the received SIR:

$$R_i = (1/T) \log_2(1 + KSIR_i) \tag{20}$$

The aggregate data rate for the system can then be written as a sum of terms of this form.

$$R_{system} = \textstyle\sum_i R_i = (1/T) \log_2 \prod_i (1 + KSIR_i) \tag{21}$$

So throughput maximization is equivalent to maximizing the product of SIR. This was also observed by Qiu and Chawla in [6], [15] where they used it for optimizing throughput in cellular networks. Overall system throughput is now defined as the maximum aggregate data rate supportable by the system given a set of users with defined QoS.

C. Throughput optimization

Formulation 5: (Optimize power for throughput maximization) The following nonlinear problem of optimizing user node powers to maximize total network throughput is a convex optimization problem.

maximize R_{system}
subject to

$$\begin{array}{lcl} R_i & \geq & R_{i,LB}, \ \forall i \\ 1 - \prod_{k \neq i} \frac{1}{1+\frac{SIR_{th} G_{ik} P_k}{G_{ii} P_i}} & \leq & Pr_{out_i} \\ 1 - \prod_{s \in S} \prod_{k \neq s} \frac{1}{(1+\frac{SIR_{th} G_{ik} P_k}{G_{ii} P_i})} & \leq & Pr_{out\_path\_s} \\ P_i & \leq & P_{i,UB} \end{array} \quad (22)$$

The objective function is the overall system throughput. In the actual optimization the posynomial objective function $\prod_i ISR_i$ is minimized; which, as shown previously, is equivalent to maximizing the system throughput. The objective function is now optimized over the set of all feasible powers P_i.

The first constraint is the data rates demanded by existing system users. The second constraint is the outage probability limitations demanded by users using single hops. The third constraint is the outage probability limitations for users using a multi-hop path. Lastly, the forth constraint is regulatory or system limitations on transmitted powers.

D. Throughput maximization simulation

A simple four node multi-hop network is considered in the following simulation. As shown in figure 2, the network consists of 4 nodes A, B, C, and D, and 4 links $1, 2, 3$, and 4. On link 1 node A is the transmitter and node B is the receiver; similarly, the transmitter and receiver nodes for each link are shown in the figure. Note that node A is the transmitter on both links 1 and 3, illustrating that a node can be a transmitter and/or receiver on many links. Nodes A and D as well as B and C are separated by a distance of 20m. By geometry the distance of each transmit path is $10\sqrt{2}$m.

For our simulation each link has a maximum transmit power of 1W. Alternatively, we could also have placed the power constraint on each node instead of each link by adding a constraint that $P_1 + P_3 \leq 1$W. All nodes are using MQAM modulation. The baseband bandwidth for each link is 10kHz, the minimum data rate for each link is 100bps for maintenance data, and the target BER is 10^{-3}. For the Rayleigh fading we require a probability of outage of $P_{out} = 0.1$ for an SIR threshold of 10dB. The gains for each link are computed as $G_{ij} = \frac{1}{200}\left[\frac{1}{d}\right]^4$ for $i \neq j$, and $G_{ii} = \left[\frac{1}{d}\right]^4$, with the exception of G_{12} and G_{34} which we set equal to 0 since we assume that a node does not transmit and receive at the same time. The factor of $\frac{1}{200}$ can be viewed as the spreading gain in a CDMA system, or power falloff with frequency in a FDMA system. This gives the following gain matrix:

$$G = 10^{-4} \cdot \begin{bmatrix} 0.2500 & 0.0003 & 0.0012 & 0.0003 \\ 0 & 0.2500 & 0.0003 & 0.0012 \\ 0.0012 & 0.0003 & 0.2500 & 0.0003 \\ 0.0003 & 0.0012 & 0 & 0.2500 \end{bmatrix} \quad (23)$$

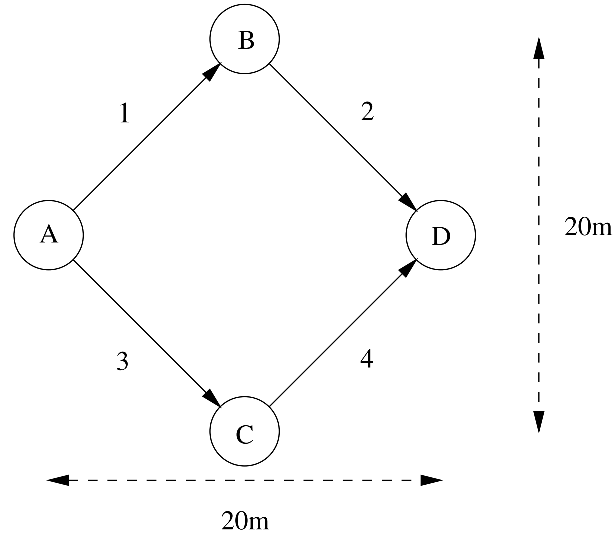

Fig. 2. Network Topology for Simulation

Using the geometric programming optimization method we find the maximum aggregate data rate is $R = 216.8$kbps, with $M = 42.8$QAM modulation for each link, $R_i = 54.2$kbps for each link, and $P_1 = P_3 = 0.709$W and $P_2 = P_4 = 1$W link transmit powers. The resulting $SIR = 21.7$dB on each link. The symmetry in modulation levels and SIR is due to the symmetries in the network topology, and not due to any explicit optimization constraint.

E. Admission control and pricing

In this subsection admission control and a possible approach to pricing are considered. As discussed in section IV E, a new user is admissible if his QoS requirements can be supported by the system without disturbing current users. In this model a user is admissible if a feasible solution of the problem in formulation 5 exists after the new user's QoS constraints have been added.

In the pricing discussion of section IV, a new user is charged according to the equivalent number of standard users the new user costs the system in lost future user capacity. In economics this approach is termed the opportunity cost associated with serving a new user.

The concept used for multi-hop networks is similar, but a different measure of opportunity cost is used. In a multi-hop network, the route taken by a stream of packets, not just its QoS, jointly determine its effect on the network. Therefore, the opportunity cost is taken as the data transport capacity lost by the entire network in supporting a new user. When a new user is added to the system the ability of the system to support additional data traffic is reduced. The opportunity cost can be estimated for the multi-hop network by subtracting the maximum data transport capacity of the system, in bits per second, after the user is added to the system, from the capacity before the user is added to the system. The value of the objective function used in the multi-hop formulation is precisely the maximum aggregate data rate for the network.

F. Pricing Simulation

Consider admission control and pricing for the simulation in Section VI-D above. Initially the system has no users with QoS constraints beyond the basic setup given previously. So current user data is admitted and priced based on a best effort transmission. Three new users U_1, U_2, and U_3 are going to arrive to the system in order. U_1 and U_2 require 30kbps sent along the upper path $A \rightarrow B \rightarrow D$, while U_3 requires 10kbps sent from $A \rightarrow B$. All three users require the outage probability to be less than 0.1. When U_1 arrives to the system the optimization with his QoS demands has the same solution as without the demands, so his price is the baseline price. Next, U_2 arrives, and his QoS demands decrease the maximum system throughput from 216.82 kbps to 216.63 kbps, so his price is the baseline price plus an amount proportional to the change in system throughput. Finally, U_3 arrives, and his QoS demands have no feasible solution, so he is not admitted to the system.

Note that the prices charged are a function of system demands when the user arrives. If U_2 had arrived before U_1, U_2 would have paid less and U_1 more. Similarly, U_3 would have been admitted to the system for the baseline price if he had arrived before U_2, where as his price was effectively infinite when he arrived after U_2.

VII. Resource allocation for delay and efficiency optimization in wireless ad hoc networks

In this section multi-hop networks are treated from a general perspective of resource allocation. Resources include power, the number of flows in each category of service, bandwidth and capacity of each link. These resources are allocated according to the optimization criteria of transmission delay, unused capacity and overall system throughput.

A. Problem formulations

Consider a network with J links with capacity of C_j packets per second for each link j. There are K classes of traffic with different QoS requirements to be transported over the network. For each QoS class k, the bandwidth required is b_k Hz, and the delay guarantee in the service level agreement (SLA) is $d_{k,UB}$ seconds. End to end total delay consists of propagation delay, transmission delay and queuing delay. Complementary to the discussion on queuing delay in section V, in this section we assume that transmission delay is the dominant term for ad hoc networks. The minimum acceptable probability of delivering the packet across the unreliable network in the SLA is denoted by $p_{k,LB}$.

Similar to the last section, each stream of traffic from source s to destination d will traverse certain specific links as dictated by the particular routing protocol used for the network. Denote by K_j the set of traffic using link j and by J_k the set of links traversed by QoS class k. Denote by n_k the number of packets dynamically admitted in the k^{th} class of traffic.

In an ad hoc network each link may fail due to a node leaving the network or due to an outage. p_j is defined as the probability that this link will be maintained during the transmission. Note that by increasing transmitter power over a link j while keeping other parameters of the network constant, the SIR of link j can be increased. Consequently, the outage probability of link j will decrease and p_j will increase. Therefore, power control is reflected through the optimization variable p_j. The sixth formulation is the following:

Formulation 6: (SLA feasibility under network constraints) The following nonlinear problem of testing SLA feasibility is a convex optimization problem.

$$\begin{array}{lllll} \text{minimize} & \text{No Objective Function} & & \\ \text{subject to} & \sum_{k \in K_j} b_k n_k & \leq & C_j, \ \forall j \\ & \sum_{j \in J_k} \left(\frac{\sum_{i \in K_j} n_i}{C_j} \right) & \leq & d_{k,UB}, \ \forall k \\ & \prod_{j \in J_k} p_j & \geq & p_{k,LB}, \ \forall k \\ & b_k n_k & \geq & R_k, \ \forall k \\ & b_{k^*} n_{k^*} & = & C_{j^*} \\ & \frac{n_{k^*}}{C_{j^*}} & = & d^*_{k,j} \\ & p_j & \leq & p_{j,UB} \\ & b_k, C_j, p_j, d_{k,UB}, p_{k,LB} & \geq & 0 \end{array} \tag{24}$$

No objective function is necessary to test feasibility of the SLA terms $p_j, d_{k,UB}$ and $p_{k,LB}$. Note that the first constraint is the link capacity constraint, the second one is the delay guarantee constraint and the third one the delivery probability constraint. The fourth constraint delivers a guaranteed data rate to each class of traffic. The fifth constraint makes room for SLA terms that give a class of traffic the sole right to traverse a link j^*. This could be for bandwidth requirements or for security reasons as in virtual private networks. The sixth constraint allows for SLA terms that specify not just an end to end total delay guarantee, but also an exact delay requirement for a particular traffic class k^* on a link j^*. The other constraints are positivity constraints on the variables, and upper bound constraints on p_j.

The following parameters can all become variables in the optimization: $b_k, n_k, p_j, C_j, d_{k,UB}$ and $p_{k,LB}$. Variables $b_k, d_{k,UB}$ and $p_{k,LB}$ are terms in the SLA. The link capacities C_j and probability of maintaining a link p_j are network resources to be optimized over. Admission control is reflected in n_k.

In the seventh formulation, the unused capacity of a particular link j_0 is maximized. This link could be a bottleneck link or the most often traveled link in the network where capacity is considered scarce or of great value.

Formulation 7: (Unused capacity maximization) The following nonlinear problem of maximizing the unused capacity under SLA and network constraints is a convex optimization problem.

$$\begin{array}{lll} \text{maximize} & C_{j0} - \sum_{k\in K_{j_0}} b_k n_k & \\ \text{subject to} & \sum_{k\in K_j} b_k n_k & \leq \; C_j, \; \forall j \\ & \sum_{j\in J_k} \left(\frac{\sum_{i\in K_j} n_i}{C_j} \right) & \leq \; d_{k,UB}, \; \forall k \\ & \prod_{j\in J_k} p_j & \geq \; p_{k,LB}, \; \forall k \\ & b_k n_k & \geq \; R_k, \; \forall k \\ & b_{k^*} n_{k^*} & = \; C_{j^*} \\ & \frac{n_{k^*}}{C_{j^*}} & = \; d^*_{k,j} \\ & p_j & \leq \; p_{j,UB} \\ & b_k, C_j, p_j, d_{k,UB}, p_{k,LB} & \geq \; 0 \end{array} \quad (25)$$

The objective function is to maximize unused capacity of a link j_0 by keeping the used capacity to the minimum under all network and QoS constraints. The constraints are the same as in formulation 6.

In the eighth formulation, the total delay for a particular class of traffic is minimized.

Formulation 8: (Weighted Joint Capacity and Delay Minimization) The following nonlinear problem of minimizing transmission delay under SLA and network constraints is a convex optimization problem.

$$\begin{array}{lll} \text{minimize} & \sum_{j\in J_{k_0}} \frac{\sum_{i\in K_j} n_i}{C_j} + \alpha \left(\sum_j C_j \right) & \\ \text{subject to} & \sum_{k\in K_j} b_k n_k & \leq \; C_j, \; \forall j \\ & \sum_{j\in J_k} \left(\frac{\sum_{i\in K_j} n_i}{C_j} \right) & \leq \; d_{k,UB}, \; \forall k \\ & \prod_{j\in J_k} p_j & \geq \; p_{k,LB}, \; \forall k \\ & b_k n_k & \geq \; R_k, \; \forall k \\ & b_{k^*} n_{k^*} & = \; C_{j^*} \\ & \frac{n_{k^*}}{C_{j^*}} & = \; d^*_{k,j} \\ & p_j & \leq \; p_{j,UB} \\ & b_k, C_j, p_j, d_{k,UB}, p_{k,LB} & \geq \; 0 \end{array} \quad (26)$$

where α is the marginal tradeoff of capacity for delay. By increasing capacities available on each link at the relative cost α through dynamic bandwidth allocation or bandwidth leasing, delay of the most time sensitive QoS class can be decreased.

B. Joint capacity and delay minimization simulation

This simulation investigates the tradeoff between delay and cost of capacity in Formulation 8. For the network in Fig. 3 there are three classes of traffic. The first class is data traffic sent along path ABCD requiring a rate of 50 packets/second and a maximum delay of 0.2 seconds. The second class is also data traffic sent along path DFEA with the same rate and delay requirements. The third class of traffic is voice sent along path ABFD with a rate requirement of 250 packets/second. We want to minimize both the delay of the voice traffic and the cost of capacity that we must provision or lease. We accomplish this by minimizing a weighted sum of the voice traffic delay and the total capacity used, subject to meeting the rate constraints on all traffic classes and the delay constraints on the data traffic. For each value of α, the marginal tradeoff parameter between delay and capacity, we find the minimum delay achievable for the

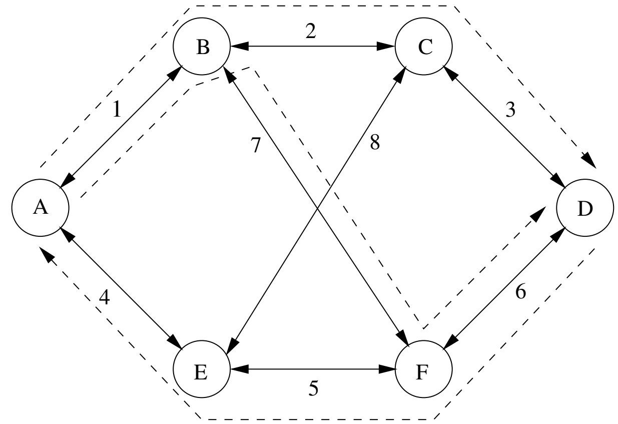

Fig. 3. Network Topology for Simulation

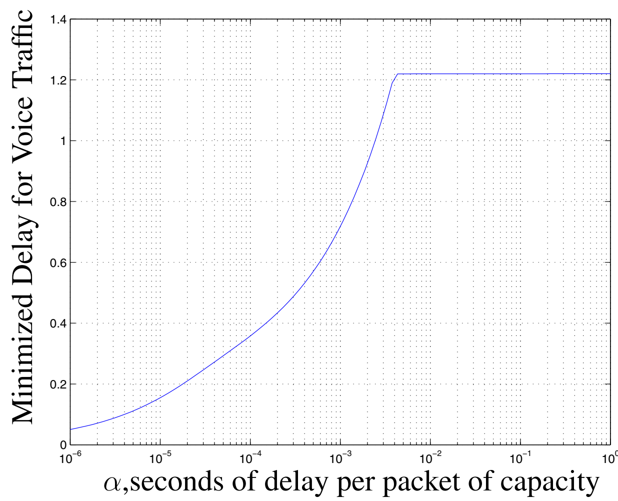

Fig. 4. Trade off between voice traffic delay and capacity cost

voice traffic given in Fig. 4 with log scale for the x-axis. The tradeoff curve shows that the minimum delay increases rapidly with increasing cost of capacity until it reaches the delay associated with the minimum capacity required to support the voice signal; from that point onwards the tradeoff curve is flat.

C. Extensions

A number of extensions can be made to Formulations 6 to 8. One extension is minimizing the maximum transmission delay for users. A second extension is accounting for queueing delay at the nodes. The formulations, proofs and interpretations are similar to those in section IV C and section V, respectively.

Another extension is weighted fairness formulations which can also be solved by geometric programming. The weight parameters can become variables, therefore fairness and QoS criteria can be jointly optimized using geometric programming. This is based on the observation that the posynomial form of the objective and constraint functions is maintained when weights of proportional fairness are variables; thus preserving the geometric programming framework. Solving these problems would give the globally optimal tradeoff between the weights attached to each user and the resources allocated among the users.

VIII. Simple Heuristics

Although geometric programming has highly efficient interior point algorithms that find the global optimal solution in polynomial time, in some practical systems, suboptimal simple heuristics with even lower computational load are desired. In this section, we briefly outline two heuristics for some of the formulations. Geometric programming now becomes an efficient tool for heuristic validation and testing.

A. A heuristic for cellular networks

Using the geometric programming formulations in section IV as an efficient tool of validating heuristics, we find that the following simple heuristic performs well with a small suboptimality gap for Formulation 1. Denote by D_{ij} the total path loss from the transmitter on link j to the receiver on link i that takes into account all factors other than power:

$$D_{ij} = \frac{d_{ij}^{-\gamma_{ij}} \alpha_{ij} K_s^{-1}}{n_i} \tag{27}$$

where d_{ij}, γ_{ij}, and α_{ij} are the distance, path loss exponent, and normalization constant from the transmitter on link i to the receiver on link j, respectively. Denote by D_i the mean of D_{ij} for $j \neq i$. The constraint set can then be rewritten as a system of inequalities in terms of P_i and D_i. Let the SIR of user i^* be the objective function to be maximized under this constraint set. For the heuristic, set P_{i^*} to its maximum $P_{i^*,UB}$, and make all constraints active by turning the system of inequalities into a system of equalities. Feasibility of the SIR requirements can be analytically determined, and if feasible, the system of equalities can easily be solved for $P_i, i \neq i^*$. Substituting the resulting set of powers into SIR_{i^*} gives the maximized SIR for user i^* under this heuristic. Note that the computational load for this heuristic is very small.

From empirical results in simulations, the difference in SIR, for both user i^* and all the other users, between the convex optimization algorithm and this heuristic is smaller than 5 percent for networks with sizes larger than 15.

B. A distributed heuristic for ad hoc networks

Extending similar ideas in [11], the following is a simple iterative heuristic for throughput maximization in ad hoc networks under outage probability constraints as in formulation 5. Briefly, the heuristic makes the constraints in formulation 5 active, and reexpress each constraint. For example, the constraint $1 - \prod_{s \in S} \prod_{k \neq s} \frac{1}{(1 + \frac{SIR_{th} G_{ik} P_k}{G_{ii} P_i})} \leq Pr_{out\_path\_s}$ becomes $\prod_{s \in S} \prod_{k \neq s} (1 + \frac{SIR_{th} G_{ik} P_k}{G_{ii} P_i}) = \frac{1}{1 - Pr_{out\_path\_s}}$. Next, taking the log of both sides and multiplying both sides by P_i, the constraints are rewritten as a system of linear equations, which can be put into matrix form $Q(P)P = \lambda P$. Thus, an initial power vector $P^{(0)}$ can be randomly selected, and the next iterate of the power vector $P^{(1)}$ can be computed as the Perron Frobenius eigenvector of $Q(P^{(0)})$. This heuristic becomes a sequence of Perron Frobenius eigenvector problems, which is computationally easy to solve.

IX. Summary

Various QoS provisioning problems are considered for wireless cellular and ad-hoc networks from a resource allocation point of view. Such formulations are nonlinear problems, but can be efficiently solved by convex optimization. The geometric programming framework makes possible formulations that include both throughput and delay as objective functions and allow for a variety of general network models. Proportional and minmax fairness algorithms are also possible. Simulations show that these algorithms can be used for admission control and pricing schemes that are based on the relative disturbances on the use of network resources by the new users. Additionally, two simple heuristics for cellular and ad hoc networks are outlined.

Acknowledgments

We would like to thank Nicholas Bambos and Andrea Goldsmith at Stanford University, and Lizhong Zheng at UC Berkeley for helpful discussions.

References

[1] M. Andersin, Z. Rosberg, and J. Zander, "Gradual removals in cellular PCS with constrained power control and noise." *Wireless Networks*, 2(1):27-43, 1996.

[2] M. Andersin, Z. Rosberg, and J. Zander, "Gradual removals in celllular PCS with constrained power control and noise." *IEEE/ACM Trans. in Networking*, 5(2):255-265, 1997.

[3] N. Bambos, "Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects." *IEEE Personal Communications Magazine*, 5(3):50-59, 1998.

[4] N. Bambos, S. Chen, and G. Pottie, "Radio link admission algorithms for wireless networks with power control and active link quality protections." *Proc. of the IEEE Infocom*, 1995.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization and Its Engineering Applications*, Stanford University EE 364 Course Reader 2001.

[6] K. Chawla and X. Qiu, "Throughput performance of adaptive modulation in cellular systems." *IEEE Proc. on Universal Personal Communications*, 2:945-950, 1998.

[7] M. Chiang, D. ONeill, D. Julian, and S. Boyd, "Resource allocation for QoS provisioning in wireless ad hoc networks." *Proc. IEEE Globecom*, San Antonio, TX, Nov. 2001.

[8] G. Foschini and Z. Miljanic, "A simple distributed autonomour power control algorithm and its convergence." *IEEE Trans. Vehicular Technology*, 42(4), 1993.

[9] G. Foschini and Z. Miljanic, "Distributed autonomous wireless channel assignment with power control." *IEEE Trans. Vehicular Technology*, 44(3):420-429, 1995.

[10] D. Julian, M. Chiang, and D. ONeill, "Robust and QoS constrained optimization of power control in wireless cellular networks." *Proc. IEEE VTC Fall*, Atlantic City, NJ, Oct. 2001.

[11] S. Kandukuri and S. Boyd, "Optimal power control in interference limited fading wireless channels with outage probability specifications." *To appear IEEE JSAC*, 2002.

[12] S. Lal and E. Sousa, "Distributed resource allocation for DS-CDMA based multimedia ad hoc wireless LAN's." *IEEE J. Sel. Areas in Communications*, 17(5):947-967, May 1999.

[13] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems." *Proceedings of 4th WINLAB Workshop* Rutgers University, NJ, 1993.

[14] Yu. Nesterov and A. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming,* SIAM 1994.

[15] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems." *IEEE Transaction on Communication*, pp. 884-895, June 1999.

[16] R. Ramanathan and R. Rosales-Hain, "Topology control of multihop wireless networks using transmit power adjustment." *Proc. IEEE Infocom*, March 2000.

[17] M.J. Shah and P.G. Flikkema, "Power-based leader selection in ad-hoc wireless networks." *IEEE Int. Performance, Computing and Communications Conf.*, pp. 134-139, 1999.

[18] S. Ulukus and R. Yates, "Adaptive power control and MMSE interference suppression." *ACM/Baltzer Wireless Networks*, 999(4), 1998.
[19] R. Yates, "A framework for uplink power control in cellular radio systems." *IEEE J. Sel. Areas in Comm.*, 13(7):1341-1347, 1995.