

Optimal and Distributed Protocols for Cross-Layer Design of Physical and Transport Layers in MANETs

John Papandriopoulos, *Student Member, IEEE*, Subhrakanti Dey, *Member, IEEE*, and Jamie Evans, *Member, IEEE*

Abstract—We seek distributed protocols that attain the global optimum allocation of link transmitter powers and source rates in a cross-layer design of a mobile ad hoc network. Although the underlying network utility maximization is nonconvex, convexity plays a major role in our development. We provide new convexity results surrounding the Shannon capacity formula, allowing us to abandon suboptimal high-SIR approximations that have almost become entrenched in the literature. More broadly, these new results can be back-substituted into many existing problems for similar benefit.

Three protocols are developed. The first is based on a convexification of the underlying problem, relying heavily on our new convexity results. We provide conditions under which it produces a globally optimum resource allocation. We show how it may be distributed through message passing for both rate- and power-allocation. Our second protocol relaxes this requirement and involves a novel sequence of convex approximations, each exploiting existing TCP protocols for source rate allocation. Message passing is only used for power control. Our convexity results again provide sufficient conditions for global optimality. Our last protocol, motivated by a desire of power control devoid of message passing, is a near optimal scheme that makes use of noise measurements and enjoys a convergence rate that is orders of magnitude faster than existing methods.

Index Terms—Congestion control, cross-layer optimization, mobile ad hoc network, network utility maximization, outage probability, power control, Rayleigh fading.

I. INTRODUCTION

TRADITIONAL paradigms for communication network design often call upon a layered approach: each slice within the network stack should only utilize services from those below, for design and implementation simplicity. Recent research efforts have shown there exists a significant performance benefit in undertaking a cross-layer networking design, by optimizing functionality of the stack across the layers.

Nonlinear optimization has been instrumental in this process. For example, in a multi-hop wired network, transmit control protocol (TCP) algorithms are used to address problems of congestion at the transport layer. These algorithms have recently

been shown to implement approximate solutions to an underlying network utility maximization (NUM)

$$\begin{aligned} & \max_{\mathbf{x} \geq 0} \sum_s U_s(x_s) \\ & \text{s.t.} \quad \sum_{s:l \in L(s)} x_s \leq c_l, \forall l, \end{aligned} \quad (1)$$

where each source s attains some nonlinear utility $U_s(x_s)$ by transmitting at data-rate x_s , doing so without overwhelming the set of intermediate links $L(s)$ along its route, and where each link l contributes a shared capacity of c_l . While this problem has been well studied in the context of wired networks having fixed link capacities [16], [19], more recent efforts have focused on a cross-layer design of source rate allocation and capacity provisioning [31].

The link capacities c_l can be conveniently optimized through power control in a mobile ad hoc network (MANET): by increasing the transmitter power on link l , its capacity also increases and vice versa. It is therefore of great benefit to consider maximizing the utility (1) over both source rates *and* link powers through a cross-layer design. In an interference limited system—for example, one based on code-division multiple-access (CDMA)—this optimization becomes nontrivial since the power allocation is coupled across the entire network.

In [10], Chiang analyzed such a joint congestion and power control problem by considering the Lagrangian dual of (1), where the link capacities are functions of their signal-to-interference ratio (SIR). A high-SIR approximation was taken in order to formulate a convex problem that was solved with a gradient-based algorithm and made distributed with message passing. These messages carried congestion state and noise measurements at each link. Network nodes would repeatedly broadcast such messages to the network and each receiving node would use the state for their own power-allocation. It turned out that the “high-SIR optimal” rate-allocation was achieved by the existing TCP algorithm, coupled to the power-allocation through the link congestion state.

In that work, an underlying assumption was fixed (snapshot) or very slowly varying wireless channels. Under such conditions, it is conceivable that such an algorithm is able to track any changes in the channel as one would have the luxury of iterating at a faster rate than the fading dynamics. Now consider a more realistic scenario where the fading rate is increased, due to mobility of the nodes and/or environment. The iteration update-rate would also need to increase to keep track of the dynamic fading state and so either the message passing overhead would become

Manuscript received October 25, 2006; revised June 12, 2007. First published February 25, 2008; current version published December 17, 2008. Approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor P. Marbach. CUBIN is an affiliated program of National ICT Australia (NICTA).

The authors are with the ARC Special Research Centre for Ultra-Broadband Information Networks (CUBIN), Department of Electrical and Electronic Engineering, University of Melbourne, Melbourne VIC 3010, Australia (e-mail: jpap@ee.unimelb.edu.au; sdey@ee.unimelb.edu.au; jse@ee.unimelb.edu.au).

Digital Object Identifier 10.1109/TNET.2008.918099

excessive, or the instantaneous channel state varies too quickly to track properly—at this point the scheme would collapse.

Our motivation for this paper came from situations where the fading falls into this fast-varying dynamic category. Rather than a search for optimum source data-rates based on the instantaneous link capacities, we allow the network to experience a limited amount of fading-induced congestion, and by doing so, avoid the fast-update problems outlined above. We make this concept rigorous in this paper through the notion of a rate-outage probability: the probability of experiencing fading-induced congestion.

The resulting cross-layer design problem becomes one to jointly allocate power and source rates such that the rate-outage probability is within some arbitrarily small target tolerance. It turns out that resources should be allocated according to some notion of an “average SIR”, rather than the instantaneous SIR that is the basis for Chiang’s snapshot model. The two paradigms are similar as far as the underlying optimization is concerned; the main point of departure between the development in this paper and related work lies in its solution. We remove the need for a high-SIR approximation, showing that in a dynamic fading environment we can no longer afford to presume as much anyway.

The underlying canonical optimization problem dealt with here is an interesting one and deserves specific treatment: solution methods for the particular snapshot or dynamic fading environments easily flows from any positive development. The problem was previously thought to be nonconvex, yet there is much structure present. We show that for a restricted class of utility functions, the canonical problem is indeed convex under a transformation. We proceed with a Lagrangian solution and observe that message passing for both power- and rate-allocation is required.

Convexity plays an important role in this paper. Indeed, some of our general convexity results can be back-substituted into other power control problems found in the literature (for example, [9], [10], [13], [21]) where a high-SIR assumption has been a popular but unnecessary path to solution. With the convexity results herein, such problems can now be solved for true global optimality.

Our most exciting results involve an effort to preserve the existing TCP stack. Consequently, measurement-based source rate allocation becomes possible—for example, with the delay-based Vegas protocol [20]. Our method involves relaxing the nonconvex problem formulation into a sequence of convex approximations. The solution to each approximation involves the existing TCP stack, each producing successively more accurate allocations until the approximation becomes exact and converges. This sequence is shown to always converge to a Karush–Kuhn–Tucker (KKT) solution of the original problem, and is a consequence of its careful construction. Unbeknownst to us, the underlying idea can be traced to the optimization community as far back as 1966 [1], [22], [29]—our scheme appears to be a rediscovery of a very useful mathematical technique that we hope finds a greater prominence in communication engineering circles. Our developments go further still, where we provide sufficient conditions for the sequence to produce a globally optimum allocation.

All developments thus far mentioned require some form of broadcast message passing for power-allocation. We recognize that in a practical setting it may be desirable to avoid such overhead and so we include a near-optimal scheme that makes use of autonomous SIR measurements at each link for power-allocation.

This paper is organized as follows. In Section II we introduce the system and fading model and derive an expression for the rate outage probability. Section III outlines the underlying optimization problem for the joint source rate- and power-allocation, for both a snapshot channel model and one based on a composite fading channel.

We review the high-SIR approximation in Section IV before introducing some new convexity properties of the Shannon link capacity in Section V. These are subsequently used in Section VI to prove that the underlying problem is convex under a transformation, when the utility functions are within a suitable family. In Section VII we do away with this transformation by employing a sequence of convex approximations—each making use of the existing TCP stack—to solve the original nonconvex problem formulation. Both of these schemes require message passing for power-allocation and is the focus of Section VIII where we outline a third near-optimal protocol making use of noise measurements instead.

Illustrative numerical simulation results are provided in Section IX that highlight the convergence speed, tracking abilities and optimality of the protocols developed in this paper. Our concluding remarks are given in Section X.

II. SYSTEM MODEL

We consider a MANET having $\mathcal{L} = \{1, \dots, L\}$ logical links, shared by $\mathcal{S} = \{1, \dots, S\}$ sources.

A. Network Stack Decomposition

1) *Network Layer*: For each source $s \in \mathcal{S}$ there exists a destination node and we denote the path (route) from this source to the destination as an ordered set of links $L(s) \subseteq \mathcal{L}$.

In this work, we assume that this layer is fixed. i.e., we assume static routes. In practice, routing may be adapted and optimized, with these improved routes communicated to the transport layer. Provided that the time-scale of such updates is much slower than the relevant time-scale of the cross-layer optimization presented here, any such adaption will not adversely interfere with the results provided.

2) *Transport Layer*: We make the usual fluid-flow model assumption where each source has an infinite amount of data to send. In this context, each source s attains a utility $U_s(x_s)$ when allocated a data-rate $x_s \geq 0$, where $U_s : \mathbb{R}_+ \rightarrow \mathbb{R}$ is an increasing strictly-concave function. In practice, however, sources are often bursty. Our developments in the sequel are iterative in nature and are nonetheless capable of handling such dynamics.

3) *Physical Layer*: We extend the physical-layer model considered in [10] to incorporate composite fading. This model utilizes CDMA so that each link in the system may simultaneously communicate within the same spectrum allocation, at the expense of multiple-access interference.

This model has an underlying assumption that nodes are able to transmit and receive simultaneously. From an information-

theoretic perspective, such a mode of operation is indeed possible: *two-way channels* have double the capacity of the corresponding one-way channel [24]. Moreover, real implementations of simultaneous transceivers have been demonstrated in practice [8], [30], where the self-interference problem has been mitigated through RF isolators and echo-cancelers, coupled with base-band digital filtering (e.g., utilizing a CDMA spreading gain).

The alternative, that we will not consider in this paper for simplicity, is to emulate full-duplex operation with distinct transmit and receive transmission modes (time-division duplex). This is only possible if each node in the network is globally synchronized through a *link-activity schedule* and is akin to a hybrid CDMA-TDMA system. Finding optimal schedules is a very difficult problem as it is combinatorial in nature, and the topic is an important area of research in its own right. Our results can easily be applied to such a paradigm, through a similar time-slotted system model as outlined in [17], [21] and references therein.

We consider fixed CDMA spreading sequences \mathbf{s}_l having length N and unit energy (i.e., $\mathbf{s}_l^T \mathbf{s}_l = 1$). These sequences are preassigned to each link l where matched filtering is employed. Additionally, we make the simplifying assumption that the self-interference cancellation at each node is perfect.

The instantaneous capacity of each link $l \in \mathcal{L}$

$$c_l(\mathbf{P}) = W \log(1 + K \text{SIR}_l(\mathbf{P})) \quad (2)$$

is modeled on the Shannon capacity, offset by the ‘‘SIR-gap’’ K that reflects a particular modulation and coding scheme [11]. For simplicity of notation in the sequel, we will assume without loss of generality, a gap $K = 1$ unless otherwise stated. The base-band (unspread) bandwidth W is assumed normalized by a fixed packet size, resulting in the capacity (2) taking units ‘‘packets/sec’’.

The SIR is defined as

$$\text{SIR}_l(\mathbf{P}) = \frac{P_l F_{ll} S_{ll} (\mathbf{s}_l^T \mathbf{s}_l)^2}{\sum_{j \neq l} P_j F_{lj} S_{lj} (\mathbf{s}_l^T \mathbf{s}_j)^2 + \sigma^2},$$

where $\mathbf{P} = [P_1, \dots, P_L]^T$ is a vector of transmitter powers, $F_{ij} S_{ij}$ is the *instantaneous gain* from the transmitter on link j to the receiver on link i , and we assume thermal noise power σ^2 at each receiver.

Without loss of generality, we will absorb the $(\mathbf{s}_i^T \mathbf{s}_j)^2$ terms into the gain terms S_{ij} using the ‘‘effective gains’’ $G_{ij} = S_{ij} (\mathbf{s}_i^T \mathbf{s}_j)^2$ to simplify the SIR expression

$$\text{SIR}_l(\mathbf{P}) = \frac{P_l F_{ll} G_{ll}}{\sum_{j \neq l} P_j F_{lj} G_{lj} + \sigma^2}. \quad (3)$$

B. Fading Model

We can decompose each instantaneous channel gain $F_{ij} G_{ij}$ into fast- and slowly-varying components F_{ij} and G_{ij} respectively. We assume Rayleigh fast-fading, where random variables F_{ij} are iid exponentially distributed with unit mean. The terms G_{ij} model slow-fading (such as distance-dependent path-loss and/or log-normal shadowing) and are assumed constant over the time-scale of interest.

The independence assumption on each F_{ij} is justified since they represent the fading on *distinct* paths between nodes in the network. The unity mean is without loss of generality, as any non-unity value can be absorbed into the corresponding G_{ij} component.

It will be useful to define the ‘‘average SIR’’,

$$\begin{aligned} \overline{\text{SIR}}_l(\mathbf{P}) &= \frac{E[P_l F_{ll} G_{ll}]}{E\left[\sum_{j \neq l} P_j F_{lj} G_{lj} + \sigma^2\right]} \\ &= \frac{P_l G_{ll}}{\sum_{j \neq l} P_j G_{lj} + \sigma^2} \end{aligned} \quad (4)$$

where we have made use of the unity mean of each F_{ij} .

C. Rate-Outage Probability

Define the *rate-outage probability* as the probability that ingress rate R_l to link l exceeds its randomly time-varying capacity c_l , resulting in fading-induced congestion. It is written

$$\begin{aligned} &Pr\{R_l > c_l(\mathbf{P})\} \\ &= Pr\left\{F_{ll} < \frac{R_l^{\text{th}}}{P_l G_{ll}} \left(\sum_{j \neq l} P_j F_{lj} G_{lj} + \sigma^2\right)\right\}, \end{aligned} \quad (5)$$

where we define the data-rate threshold $R_l^{\text{th}} = [\exp(R_l/W) - 1]$.

For our Rayleigh model, we can write (5) in closed form (cf. [15], [25]) as

$$\begin{aligned} &Pr\{R_l > c_l(\mathbf{P})\} \\ &= 1 - \exp\left(\frac{-\sigma^2 R_l^{\text{th}}}{P_l G_{ll}}\right) \prod_{j \neq l} \left[1 + \frac{R_l^{\text{th}} P_j G_{lj}}{P_l G_{ll}}\right]^{-1} \end{aligned} \quad (6)$$

$$\leq 1 - \exp(-R_l^{\text{th}} / \overline{\text{SIR}}_l(\mathbf{P})) \quad (7)$$

where the upper bound was derived in [25] and has been shown to be very tight.

III. CROSS-LAYER DESIGN PROBLEMS

A. The Canonical Problem

Making use of our notion of link capacity (2), we can write the joint power and congestion control problem as the following canonical problem

$$\begin{aligned} &\max_{\mathbf{x}, \mathbf{P} \geq \mathbf{0}} \sum_{s \in \mathcal{S}} U_s(x_s) - \omega \sum_{l \in \mathcal{L}} P_l \\ &\text{s.t.} \quad \sum_{s: l \in L(s)} x_s \leq c_l(\mathbf{P}), \forall l \in \mathcal{L}, \end{aligned} \quad (8)$$

where we jointly optimize over the vector of nonnegative source rates $\mathbf{x} = [x_1, \dots, x_S]^T$ and link transmitter powers \mathbf{P} . The constraints ensure that the network is devoid of congestion: that the rate-allocation of sources do not overwhelm any intermediate links, where the link-capacity is now a quantity that is optimized through the power-allocation.

Much like the general NUM problem (1), the objective here is to maximize the utility of the network. In a MANET, however, power also becomes an important consideration and so the above optimization contains a second objective to minimize

power (helping to reduce interference and prolong the life of nodes relying on batteries). We incorporate the fixed weight $\omega \in [0, \infty)$ that allows these conflicting objectives to be traded by a specified amount, a common way in which to treat multi-criterion optimization problems [7, Sec. 4.7].

Should we choose $\omega = 0$ and ignore the fast-varying gains F_{ij} (i.e., assume them fixed and absorbing their value into the slowly-varying gains G_{ij}), this problem becomes the NUM optimization that is the basis for Chiang’s work in [10].

This problem formulation is based on a somewhat implicit assumption that all channel gains F_{ij} and G_{ij} are fixed: a so-called snapshot channel model. Any optimum resource allocation is then tied directly to the fading state. To cater for a time-varying channel, a completely new search could be undertaken for each fading state. From a practical perspective, however, such an approach may be excessive and so iterative algorithms designed to solve the snapshot optimization are often employed even though the channel may change during execution. To track such changes, it is often the case that the iteration rate need be “fast-enough”, often greater than the fading-rate.

B. Composite Fading Formulation

When the channel is fast-varying, it can become prohibitively difficult to undertake a rate- and power-allocation across the entire network. To illustrate, consider the rate-allocation of a source s . The congestion state along the route $L(s)$ must be communicated back to the source, either implicitly by measurement, or explicitly through message passing. The congestion state is time-varying as it depends directly on the link-capacities en route, that is in turn linked to the fading state of each link. As the fading rate increases, more timely feedback of the end-to-end congestion state becomes increasingly difficult.

To alleviate this problem, we could allow the network to experience a limited amount of fading-induced congestion. The constraints of the canonical problem (8) are no longer adequate: they ensure a zero-level tolerance on fading-induced congestion. We instead consider the rate-outage probability and aim to constrain it to a small tolerable value. By doing so, we show that such a problem in fact becomes decoupled from the instantaneous state of the composite fading channel. Resources can then be allocated on a much slower and manageable time-scale.

Incorporating rate-outage constraints results in the new NUM problem

$$\begin{aligned} & \max_{\mathbf{x}, \mathbf{P} \geq 0} \sum_{s \in \mathcal{S}} U_s(x_s) - \omega \sum_{l \in \mathcal{L}} P_l \\ \text{s.t.} \quad & \Pr \left\{ \sum_{s: l \in L(s)} x_s > c_l(\mathbf{P}) \right\} \leq \Omega_l^{\text{rate}}, \forall l \in \mathcal{L}, \quad (9) \end{aligned}$$

where we limit the probability to a maximum value of $\Omega_l^{\text{rate}} \in (0, 1)$ on each link l . It turns out that at the optimum solution, all constraints are met with equality and so we refer to these as rate-outage probability targets. They are fixed parameters of the system.

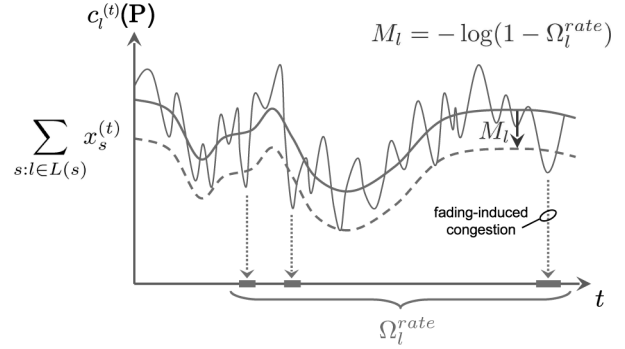


Fig. 1. Concept of the rate outage probability. The ingress rate to a link (dotted) is adjusted below the average link capacity by a fading-margin M_l that limits the severity of any fading-induced congestion to within a target probability Ω_l^{rate} .

Although the expressions for rate-outage probability are known in closed form for the Rayleigh channel [cf. (6)], we proceed by making use of the bound (7) to reformulate our constraints as follows:

$$\Pr \left\{ \sum_{s: l \in L(s)} x_s > c_l(\mathbf{P}) \right\} \leq \underbrace{1 - \exp \left(\frac{-R_l^{\text{th}}}{\overline{\text{SIR}}_l(\mathbf{P})} \right)}_{\text{new rate-outage constraint}} \leq \Omega_l^{\text{rate}}.$$

This reformulation significantly simplifies subsequent analysis while still ensuring that the original rate-outage probabilities are met.

Taking the logarithm of both sides of this constraint and simplifying, we form a new optimization problem:

$$\begin{aligned} & \max_{\mathbf{x}, \mathbf{P} \geq 0} \sum_{s \in \mathcal{S}} U_s(x_s) - \omega \sum_{l \in \mathcal{L}} P_l \\ \text{s.t.} \quad & \sum_{s: l \in L(s)} x_s \leq W \log(1 + M_l \overline{\text{SIR}}_l(\mathbf{P})), \forall l \in \mathcal{L}, \quad (10) \end{aligned}$$

where we have scaled the average SIR by a positive constant M_l that is a function of the fixed rate-outage target probabilities as given by

$$M_l = -\log(1 - \Omega_l^{\text{rate}}). \quad (11)$$

This optimization has the same form as the canonical problem (8), however we now deal with *average* SIR. We are not concerned at all with the instantaneous fading state of the channel, only its average. To cater for the variation in the channel, we employ the fading-margin (11), and allocate source rates so that the total ingress traffic to any link does not exceed this margin. In a Rayleigh fading environment, this margin is “just the right amount” to meet the rate-outage probability targets Ω_l^{rate} , illustrated in Fig. 1.

The NUM problems (8) and (8) are mathematically equivalent, in the sense that a technique to solve one can be used to solve the other. In the interest of notational clarity, we will concentrate on the canonical problem (8) in the sequel, with reference to the composite fading formulation as appropriate.

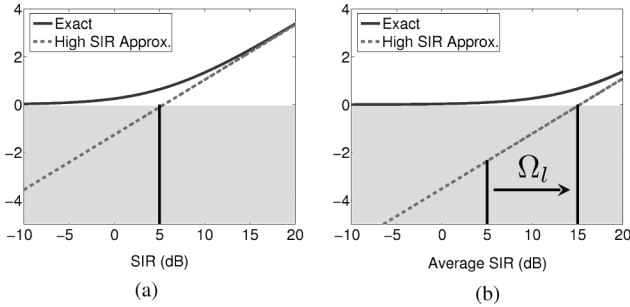


Fig. 2. Accuracy of the high-SIR assumption with and without fading margins. With a fading margin, the minimum SIR to support a rate of zero kbps on a link is over 15 dB. For positive link-rates, an exceedingly high SIR is required for accuracy. (a) Without fading margin. (b) With fading margin.

IV. PRIOR TREATMENT: HIGH-SIR APPROXIMATIONS

The canonical problem (8) is not jointly convex in $\{\mathbf{x}, \mathbf{P}\}$. The nonconvexity stems from the link capacity terms (2) within the constraints: they are neither convex nor concave in \mathbf{P} . Consequently, local optima may not necessarily be globally optimum solutions and under appropriate regularity conditions, the well-known KKT optimality conditions are merely necessary and not sufficient.

Prior work in this area employed a “high-SIR approximation” on these link capacities, followed by a variable transformation $\tilde{P}_l = \log P_l$. The resulting problem was then convex in the new $\{\mathbf{x}, \tilde{\mathbf{P}}\}$ -space so that any locally optimum solution was then declared globally optimum, up to the approximation [9], [10], [13], [14], [21].

That procedure can be interpreted as a lower bound approximation

$$\underbrace{\sum_{s:l \in L(s)} x_s}_{\text{lower bound approximation}} \leq W \log(\text{SIR}_l(\mathbf{P})) \leq c_l(\mathbf{P}) \quad (12)$$

that reduces the size of the constraint set in relation to the original problem. As such, there is no guarantee that the feasible region of the high-SIR approximate problem includes the true globally optimal point(s). The tightness of the bound is what determines how close the solution to the high-SIR approximate problem is to the true global optimum. In this case, the bound becomes tight when $\text{SIR} \gg 1$.

Consider the following typical operating parameters:

- $K = \frac{-1.5}{\log_5 \text{BER}}$ for MQAM (see [27], [14], [11]), with $\text{BER} = 10^{-3}$;
- $\Omega_l^{\text{rate}} = 10\%$ (when a fading-margin is employed).

Fig. 2 illustrates the tightness of the bound over different SIR operating points. For the canonical problem without fading margins, a 10–15 dB SIR is required for reasonable tightness. When fading margins are employed, this jumps toward 20 dB average SIR, and registers positive link-rates only after a staggering 15 dB average SIR.

Such high operating points are undesirable for CDMA systems, as we must limit the interference between links by employing extremely large spreading gains, or complex interference suppression schemes so that channels appear near-orthogonal. Moreover, it is often unclear whether one would know

a-priori whether a system will operate in a high-SIR regime, as the SIR (3) itself depends on the power-allocation and cross-gains G_{ji} between links i and j .

On the surface, a high-SIR assumption may seem reasonable in some circumstances. Nevertheless, it can also be embarrassingly unacceptable—especially when a fading margin comes into play. Fortunately such an approximation is not required, as we next demonstrate.

V. SHANNON CAPACITY CONVEXITY PROPERTIES

In this section we expose some convexity properties of the Shannon link capacity (2) in preparation for the development of a method attaining a global solution to the canonical problem (8) in the subsequent section. At the heart of our investigations herein is a more general notion of convexity that has been the subject of extensive research spanning the last 50+ years.

One such important type of generalized convexity is the class of quasiconvex functions. While not necessarily convex, these functions have the property that their lower level sets are convex [2, Ch. 3]. Consequently, optimization problems comprising quasiconvex constraint functions have the desirable property that any local optimum is also global and the KKT conditions are sufficient for global optimality, much like a regular convex program.

Result 1: The Shannon link capacity (2) is a quasiconcave function.

Proof: We can re-write the Shannon link capacity as

$$c_l(\mathbf{P}) = W \log \left(\frac{\mathbf{a}_l \mathbf{P} + \sigma^2}{\mathbf{a}_{-l} \mathbf{P} + \sigma^2} \right),$$

where $\mathbf{a}_l = [F_{l1}G_{l1}, F_{l2}G_{l2}, \dots, F_{lL}G_{lL}]$ is a row-vector of channel gains, and \mathbf{a}_{-l} is exactly the same, with the l th element zeroed.

The argument to the logarithm is known as a linear fractional and is quasilinear (both quasiconvex and quasiconcave) [2, Sec. 5.2]. The required result follows by composition with the nondecreasing concave function $W \log(\cdot)$ [2, Prop. 5.1]. ■

In the context of the canonical problem (8), Result 1 might lead us to believe that the constraint functions (comprising a sum of linear terms and the negative of the link capacity) is a quasiconvex function, in turn, giving us hope of sufficient conditions on global optimality. Unfortunately quasiconvex functions are not closed under addition, in contrast to the convex case. Despite this fact, we include Result 1 as it may prove useful in solving other problems involving the Shannon link capacity.

A more promising aspect of the growing body of research on generalized convexity relates to the family of convex transformable functions. These are functions that can be turned into convex ones by a one-to-one transformation of their domain and/or range. In an optimization context, such functions prove useful as seemingly nonconvex optimization problems can be convexified and solved with relative ease, using convex programming techniques such as duality.

The most general notion of convexifiability was introduced by Ben-Tal in 1977 [3]. A function f on $D \subset \mathbb{R}^n$ (not necessarily convex), having a domain and range transformation by functions h and ϕ respectively, is said to be (h, ϕ) -convex if the

combined transformation given by its corresponding function $\hat{f}(\hat{x}) = \phi(f(h^{-1}(\hat{x})))$ is convex on $h(D)$.

We now proceed with the following general convexity results before providing our main Shannon link capacity convexity results.

Result 2: The function $h(x) = \log(\log(1 + e^x))$ is nondecreasing and concave in $x \in \mathbb{R}$.

Proof: Omitted due to space restrictions; see [26]. ■

Result 3: The high-SIR (lower bound) approximation of the Shannon link capacity $g(\mathbf{P}) = W \log(\text{SIR}_l(\mathbf{P}))$ is (\log, x) -concave.

Proof: Under the transformation $\tilde{P}_l = \log P_l$, the approximation can be written

$$\tilde{g}(\tilde{\mathbf{P}}) = W \left[\tilde{P}_l + \log(F_{lj}G_{ll}) - \log \left(\sum_{j \neq l} F_{lj}G_{lj}e^{\tilde{P}_j} + \sigma^2 \right) \right].$$

The sum of a linear and concave negative-log-sum-exp term is clearly concave. ■

Result 4: The Shannon link capacity is (\log, \log) -concave.

Proof: Under the range transformation, we can write the Shannon link capacity as the composition $h(g_l(\mathbf{P}))$ where $h(\cdot)$ and $g_l(\cdot)$ are given as in Result 2 & 3. With Results 2 & 3 at hand, the proof follows after domain transformation and scalar composition [7, pg. 84]. ■

Result 5: The Shannon link capacity is a so-called d.c. function (difference of two concave functions).

Proof: The proof follows after rewriting it as

$$c_l(\mathbf{P}) = W \log(\mathbf{a}_l \mathbf{P} + \sigma^2) - W \log(\mathbf{a}_{-l} \mathbf{P} + \sigma^2),$$

where \mathbf{a}_l and \mathbf{a}_{-l} are defined in the Proof of Result 1. ■

Result 4 implies that we might transform the nonconvex optimization (8) into a much easier convex problem. On the other hand, Result 5 implies that in general, the Shannon link capacity may lead to great difficulties, as it is known that optimization problems having d.c. constraints are NP-hard [12]. We resolve these seemingly conflicting statements in the next section.

VI. GLOBAL OPTIMALITY VIA CONVEXIFICATION

Inspired by Result 4, we transform the canonical problem (8) as follows:

$$\begin{aligned} & \max_{\tilde{\mathbf{x}}, \tilde{\mathbf{P}}} \sum_{s \in \mathcal{S}} U_s(\exp(\tilde{x}_s)) - \omega \sum_{l \in \mathcal{L}} \exp(\tilde{P}_l) \\ & \text{s.t. } \log \left(\sum_{s: l \in L(s)} \exp(\tilde{x}_s) \right) \leq \log \left(c_l(\exp(\tilde{\mathbf{P}})) \right), \forall l \in \mathcal{L}. \end{aligned} \quad (13)$$

The link powers P_l are transformed logarithmically to ensure that the right-hand side of the constraints are concave. The source rates x_s are similarly transformed, to ensure convexity of the left-hand side (recall that log-sum-exp is convex). We therefore conclude that the constraint set is convex.

Have we succeeded in transforming the nonconvex canonical problem (8) into convex form? We address this question in the following theorems.

Theorem 1: The objective of the transformed problem (13) comprises a sum of quasiconcave functions.

Proof: The function $f(x) = \exp(x)$ is quasilinear (both quasiconcave and quasiconvex), since the lower level set $\underline{\mathcal{S}}^c = \{x \in \mathbb{R} | f(x) \leq c\}$ and the upper level set $\overline{\mathcal{S}}^c = \{x \in \mathbb{R} | f(x) \geq c\}$ are each convex. Consequently, its negative $-f(x)$ is also quasilinear.

We can write the objective as the sum

$$\sum_{s \in \mathcal{S}} U_s(f(\tilde{x}_s)) + \sum_{l \in \mathcal{L}} -f(\tilde{P}_l).$$

Each term in the first sum is quasiconcave by the concavity of the utility and composition [2, Prop. 5.1]. The proof is completed once it is recognized that each $-f(x)$ term appearing in the second sum is also quasiconcave. ■

Unfortunately quasiconcave functions are not closed under addition, and in general, their sum is nonconvex. Theorem 1 tells us that in general, the transformed problem remains nonconvex. This should not be surprising, given the d.c. nature of the original problem as noted above. Fortunately there exists an important exception that is the subject of our next result.

Theorem 2: The transformed problem (13) is convex if the utility functions are all (\log, x) -concave over their domain.

Proof: The objective is separable in each variable, and clearly concave in $\tilde{\mathbf{P}}$ (sum of negative exponentials). Concavity in $\tilde{\mathbf{x}}$ follows from the concavity of $U_s(\exp(\tilde{x}_s))$, and is satisfied whenever the utility functions are (\log, x) -concave. ■

Theorem 2 essentially tells us that when each of the quasiconcave functions comprising the objective are also concave, the transformed problem is indeed convex. This depends directly on the structure of the utility functions. Fortunately, as the following examples show, many utility functions of interest satisfy this property to a large degree.

A. Examples

1) *Vegas:* Associated with TCP Vegas is the utility function

$$U_s^{(v)}(x_s) = \alpha_s d_s \log x_s,$$

where α_s is a parameter of Vegas and d_s is the round trip propagation delay of source s [20]. This utility function is clearly (\log, x) -concave.

2) *Reno:* The following utility functions are associated with TCP Reno:

$$\begin{aligned} U_s^{(r1)}(x_s) &= \frac{\sqrt{3/2}}{D_s} \tan^{-1}(\sqrt{2/3} D_s x_s); \\ U_s^{(r2)}(x_s) &= \frac{1}{D_s} \log \left(\frac{D_s x_s}{2D_s x_s + 3} \right) \end{aligned}$$

where D_s is the propagation plus queueing delay of source s . The first utility function models an older behavior of Reno, while the second relates to newer variants. The differences relate to particular ways in which a multiplicative-decrease in rate is implemented upon detection of a packet mark. Full details and derivations are found in [18].

We are interested whether these utility functions are (\log, x) -concave. Taking the first function, we can write

the second derivative as follows, after a logarithmic domain transformation:

$$\tilde{U}_s^{(r1)}(\tilde{x}_s) = \frac{e^{\tilde{x}_s} [1 - (2/3D_s)^2 e^{2\tilde{x}_s}]}{[1 + (2/3D_s)^2 e^{2\tilde{x}_s}]^2}.$$

This second derivative is nonpositive only when $(2/3D_s)^2 e^{2\tilde{x}_s} \geq 1$. Unraveling the domain transformation and rearranging reveals that $U_s^{(r1)}(x_s)$ is (\log, x) -concave only on the domain $x_s \geq \sqrt{\frac{3}{2D_s}}$. This utility function can only satisfy the condition of Theorem 2 if we restrict its domain to some minimum rate that is a function of the total propagation and queueing delay. Nonetheless, this minimum rate is usually small in practice: just over 8 packets/sec for a delay of $D_s = 20$ ms.

The second utility function can be re-written as

$$U_s^{(r2)}(x_s) = \frac{1}{D_s} \log(D_s x_s) - \frac{1}{D_s} \log(2D_s x_s + 3)$$

and is clearly (\log, x) -concave.

3) α -Bandwidth Utility: A general utility function given by

$$U_s^\alpha(x_s) = \begin{cases} \log x_s, & \text{for } \alpha = 1 \\ (1 - \alpha)^{-1} x_s^{1-\alpha}, & \text{otherwise} \end{cases}$$

was introduced in [23] for the general NUM problem (1). It provides a so-called α -bandwidth rate-allocation through selection of the α -parameter: proportional fairness with $\alpha = 1$, harmonic mean fairness with $\alpha = 2$, and that of max-min fairness as $\alpha \rightarrow \infty$ [6], [13].

We have already established the (\log, x) -concavity for the case of $\alpha = 1$ (Vegas). Concentrating on the other case, we write its corresponding function after domain transformation as follows:

$$\tilde{U}_s^\alpha(\tilde{x}_s) = (1 - \alpha)^{-1} \exp((1 - \alpha)\tilde{x}_s).$$

By inspection, it can be seen that the second derivative $\tilde{U}_s^{\prime\prime\alpha}(\cdot)$ is nonpositive only when $\alpha \geq 1$. We conclude that this utility function is (\log, x) -concave only when $\alpha \geq 1$. Nonetheless, this α -range captures all notions of fairness that are of interest.

B. Globally Optimal Solution

Assuming that the condition of Theorem 2 is satisfied, we now derive an algorithm that solves the transformed problem (13) for the globally optimum rate- and power-allocation. We will refer to the following development as Algorithm A.

The Lagrangian dual of (13) is given by

$$\min_{\boldsymbol{\lambda} \geq \mathbf{0}} q_{\tilde{\mathbf{P}}}(\boldsymbol{\lambda}) + q_x(\boldsymbol{\lambda}) \quad (14)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_L]$ is a vector of dual variables, and (15) are partial dual functions, thanks to the separable nature of the problem:

$$q_{\tilde{\mathbf{P}}}(\boldsymbol{\lambda}) = \min_{\tilde{\mathbf{P}}} \left\{ L_{\tilde{\mathbf{P}}}(\tilde{\mathbf{P}}, \boldsymbol{\lambda}) \triangleq \sum_{l \in \mathcal{L}} \omega \exp(\tilde{P}_l) - \lambda_l \log(c_l(\exp(\tilde{\mathbf{P}}))) \right\} \quad (15a)$$

$$q_x(\boldsymbol{\lambda}) = \max_{\tilde{\mathbf{x}}} \left\{ L_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}}, \boldsymbol{\lambda}) \triangleq \sum_{s \in \mathcal{S}} U_s(\exp(\tilde{x}_s)) - \sum_{l \in \mathcal{L}} \lambda_l \log \left(\sum_{s: l \in L(s)} \exp(\tilde{x}_s) \right) \right\}. \quad (15b)$$

The following result tells us that we can solve this dual problem in lieu of the primal problem (13) for the globally optimum result.

Lemma 1: The solution to the dual problem (14) has zero duality gap.

Proof: The objective of the primal problem (13) is concave, under the (\log, x) -concave utility assumption of Theorem 2, while the constraints are convex. There exists a strictly feasible point (cf. $\{\tilde{x}_s = -\infty, \tilde{P}_l = -\infty | \forall s, l\}$) and so Slater's constraint qualification holds. The proof follows from [4, Prop. 5.3.1]. \blacksquare

We will take an iterative approach to solve the dual problem (14): at iteration t , the inner minimizations (15) are solved for fixed $\boldsymbol{\lambda}$, then a gradient method is employed to update the dual variables. Subsequent iterations $t + 1, t + 2, \dots$ would repeat this process until convergence to the globally optimum resource allocation.

1) *Link Algorithm:* Link powers are obtained by solving (15a).

Lemma 2: Repeated transmitter power updates

$$P_l^{(t+1)} = \frac{\Delta_l^{(t)}}{\omega + \sum_{n \neq l} G_{nl} \mathcal{M}_n^{(t)}} \quad (16a)$$

with

$$\Delta_n^{(t)} = \lambda_n^{(t)} \frac{\text{SIR}_n^{(t)}}{1 + \text{SIR}_n^{(t)}} \frac{1}{\log(1 + \text{SIR}_n^{(t)})}, \quad (16b)$$

$$\mathcal{M}_n^{(t)} = \Delta_n^{(t)} \frac{\text{SIR}_n^{(t)}}{G_{nn} P_n^{(t)}} \quad (16c)$$

solve (15a), for a specified $\boldsymbol{\lambda}$.

Proof: Omitted due to space restrictions; see [26]. \blacksquare

Each \mathcal{M}_n are messages in \mathbb{R}_+ comprising local information at each link n : a scaled measurement of the receiver interference-plus-noise.

These equations combine into a distributed power-update protocol through message passing, in a conceptually similar manner to the protocol described in [10]: each **receiver** on link n broadcasts their message \mathcal{M}_n . In turn, each **transmitter** l receives each of the broadcasts and estimate G_{nl} through training sequences. They utilize the quantity to update their power with (16a).

Due to strict concavity and strict convexity of the partial Lagrangian functions $L_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}}, \boldsymbol{\lambda})$ and $L_{\tilde{\mathbf{P}}}(\tilde{\mathbf{P}}, \boldsymbol{\lambda})$ respectively, the optimizations (15) have unique solutions, for $\boldsymbol{\lambda}$ fixed. By [4, Prop. 6.1.1], the dual functions $q_x(\boldsymbol{\lambda})$ and $q_{\tilde{\mathbf{P}}}(\boldsymbol{\lambda})$ are differentiable everywhere, therefore we can employ the gradient-descent

$$\lambda_l^{(t+1)} = \left[\lambda_l^{(t)} + \epsilon \left\{ \log \left(\sum_{s: l \in L(s)} x_s^{(t)} \right) - \log c_l(\mathbf{P}^{(t)}) \right\} \right]^+ \quad (17)$$

to solve the dual problem (14). This update is shown after transformation back to the original co-ordinate space and where ϵ is a sufficiently small step-size. Only local information is required for this update: a measurement of the $\text{SIR}_l^{(t)}$ and the ingress rate.

In a similar spirit to prior work in congestion control, the dual variables λ_l can be interpreted as ‘‘congestion prices’’ [10], [20], here in logarithmic form.

2) *Source Algorithm:*

Lemma 3: The source rate update

$$x_s^{(t+1)} = x_s^{(t)} \exp \left(x_s^{(t)} \epsilon \left[U'_s(x_s^{(t)}) - \sum_{l \in L(s)} \Lambda_l^{(t)} \right] \right) \quad (18a)$$

solve the inner maximization (15b), where ϵ is a sufficiently small step size and

$$\Lambda_l^{(t)} = \frac{\lambda_l^{(t)}}{\sum_{m: l \in L(m)} x_m^{(t)}} \quad (18b)$$

are normalized congestion prices. The function $U'_s(\cdot)$ is the first derivative of the utility.

Proof: The inner Lagrangian maximization (15b) is strictly concave for fixed λ , and so we can use a gradient ascent with a sufficiently small fixed step-size to find the maximizer. The ascent direction is given by

$$\frac{\partial L_{\tilde{x}}}{\partial \tilde{x}_s} = x_s \left(U'_s(x_s) - \sum_{l \in L(s)} \frac{\lambda_l}{\sum_{m: l \in L(m)} x_m} \right),$$

shown after returning to the original co-ordinate space.

The proof follows after the ascent direction in the transformed space

$$\tilde{x}_s^{(t+1)} = \tilde{x}_s^{(t)} + \epsilon \frac{\partial L_{\tilde{x}}}{\partial \tilde{x}_s}$$

is placed back into the original space:

$$x_s^{(t+1)} = x_s^{(t)} \exp \left(\epsilon \frac{\partial L_{\tilde{x}}}{\partial \tilde{x}_s} \right). \quad \blacksquare$$

This update can figure into a distributed protocol by making use of explicit message passing. A field is reserved in an acknowledgment (ACK) packet header that is sent from the receiver back to the source. Along the path, each intermediate link m accumulates its normalized congestion price Λ_m into this field. When the ACK reaches the source, this reserved field forms the summation within (18a).

VII. PRESERVING THE EXISTING TCP STACK

Alg. A requires explicit message passing for both power- and rate-allocation. In this section we develop a new scheme that makes use of the existing TCP stack for rate-allocation. Each source can then control their rate in a distributed manner through (implicit) measurement of the congestion prices en-route. The associated power-allocation remains along similar lines to previous developments.

Our new scheme guarantees a globally optimum resource allocation when sources’ utility functions are (\log, x) -concave, and otherwise provides a solution that satisfies the first-order necessary condition for optimality. Unlike related work, we make no assumptions on the SIR regime and therefore aim to

solve the nonconvex canonical problem (8) for the global optimum, and doing so in a distributed fashion. Analogous results in [10] are based on a fixed (suboptimal) convex approximation under a high-SIR regime.

A. A Series of Convex Approximations

We begin by making use of the inequality

$$\alpha \log(z) + \beta \leq \log(1 + z) \quad (19)$$

that is tight (exact) at $z = z_0$ when the approximation constants are chosen as

$$\alpha = \frac{z_0}{1 + z_0}, \quad (20a)$$

$$\beta = \log(1 + z_0) - \frac{z_0}{1 + z_0} \log z_0. \quad (20b)$$

These relations are easily derived by equating the slope and function values at z_0 and therefore a unique correspondence exists between each z_0 and the pair $\{\alpha, \beta\}$.

We now form a new lower bound approximation to the constraint of problem (8):

$$\underbrace{\sum_{s: l \in L(s)} x_s}_{\text{new lower bound approximation}} \leq W(\alpha_l \log(\text{SIR}_l(\mathbf{P})) + \beta_l) \leq c_l(\mathbf{P}), \quad (21)$$

where we have the benefit of making this approximation exact for a given $z_0 = \text{SIR}_l(\mathbf{P})$ by choosing $\{\alpha_l, \beta_l\}$ as in (20).

Making use of the parameterized lower-bound capacity, defined as

$$\check{c}_l(\mathbf{P}; \alpha, \beta) = W(\alpha \log(\text{SIR}_l(\mathbf{P})) + \beta), \quad (22)$$

optimization (8) becomes

$$\begin{aligned} \max_{\mathbf{x}, \mathbf{P} \geq \mathbf{0}} \quad & \sum_{s \in \mathcal{S}} U_s(x_s) - \omega \sum_{l \in \mathcal{L}} P_l \\ \text{s.t.} \quad & \sum_{s: l \in L(s)} x_s \leq \check{c}_l(\mathbf{P}; \alpha_l, \beta_l), \quad \forall l \in \mathcal{L}, \end{aligned} \quad (23)$$

where the approximation vectors $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L]$ and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_L]$ are fixed.

Taking a logarithmic transformation of co-ordinates $\tilde{P}_l = \log P_l$ **in the link powers only** results in an equivalent convex problem: the objective is a sum of exponentials and negative-concave (thus convex) utilities; the constraint set is convex by Result 3.

Not surprisingly, we arrive at the same high-SIR problem as Chiang [10] when we fix $\boldsymbol{\alpha} = \mathbf{1}, \boldsymbol{\beta} = \mathbf{0}$. In our formulation (23), we have choice over these approximation vectors. The following procedure for ‘‘tuning’’ them becomes natural:

- 1: Initialize all source rates $\mathbf{x}^{(0)} = \mathbf{0}$; link powers $\mathbf{P}^{(0)} = \mathbf{0}$
- 2: Initialize $\boldsymbol{\alpha}^{(0)} = \mathbf{1}, \boldsymbol{\beta}^{(0)} = \mathbf{0}$ (a high-SIR approximation)
- 3: Initialize iteration counter $\tau = 1$
- 4: **repeat**
- 5: *Maximize:* solve (23) for solution $\{\mathbf{x}^{(\tau)}, \mathbf{P}^{(\tau)}\}$
- 6: *Tighten:* update $\boldsymbol{\alpha}^{(\tau+1)}, \boldsymbol{\beta}^{(\tau+1)}$ at $z_0 = \text{SIR}_l(\mathbf{P}^{(\tau)})$ with (20)
- 7: Increment τ
- 8: **until** convergence

Lemma 4: Each iteration τ results in a monotonically improving objective. The sequence always converges, at which point the lower bound approximation (21) becomes exact.

Proof: We first establish that, at the τ th iteration, the previous iterate $\{\mathbf{x}^{(\tau-1)}, \mathbf{P}^{(\tau-1)}\}$ is a feasible solution to subproblem (23). Clearly the all-zero vector is feasible for the first subproblem when $\tau = 1$. For $\tau > 1$, we have that

$$\begin{aligned} \sum_{s:l \in L(s)} x_s^{(\tau)} &\stackrel{a}{=} \check{c}_l(\mathbf{P}^{(\tau)}; \alpha_l^{(\tau)}, \beta_l^{(\tau)}) \\ &\stackrel{b}{\leq} c_l(\mathbf{P}^{(\tau)}) \\ &\stackrel{c}{=} \check{c}_l(\mathbf{P}^{(\tau)}; \alpha_l^{(\tau+1)}, \beta_l^{(\tau+1)}) \end{aligned}$$

for each link l , and our desired result follows.

We show (a) by contradiction. On the left-hand side is the ingress rate; our approximation (21) of the link capacity is on the right-hand side. Suppose that at the optimum solution to the subproblem, the ingress rate is instead strictly less than the approximate capacity. We could then lower the link power P_l so that we achieve equality. This increases the objective of subproblem (23) and so our solution was clearly not optimum.

Inequality (b) follows from the definition of the bound (19), while equality (c) is a consequence of the tightening step (“T-step”) and (20).

We therefore conclude that the maximization (“M-step”) will either improve the objective on the τ th iteration, or remain at the same point as the previous $(\tau - 1)$ th iterate (since it is feasible). The T-step can only lead to an improved objective on the next iteration: recall that all constraints are satisfied with equality at the optimum solution to each subproblem; where they are not, an improved objective can be had. By the above development, we observe that the point $\{\mathbf{x}^{(\tau)}, \mathbf{P}^{(\tau)}\}$, while feasible for the $(\tau + 1)$ th iteration, may not be optimal for the $(\tau + 1)$ th subproblem, as the constraints may not be met with equality at that point. Consequently, an improved objective could be had.

Convergence is brought about when (b) becomes an equality. That is, the T-step results in the same approximation vectors for the following iteration: $\alpha^{(\tau)} = \alpha^{(\tau+1)}$ and $\beta^{(\tau)} = \beta^{(\tau+1)}$. Clearly then, performing an M-step would be futile as the solution would also remain the same.

We are also guaranteed convergence, since the monotonically improving objective is bounded above by the global optimum of the original canonical problem (8), and each subproblem remains within its feasible region due to the lower bound constraints (21). ■

The Proof of Lemma 4 does not change if, rather than finding the optimum to each subproblem during an M-step, a feasible point that improves the objective is found instead. This is the key to the formation of a distributed protocol: in practice, each T-step is achieved through local information (a measurement of the SIR_l) and each link l need not know when “subproblem τ has been solved”; each link performs a T-step when the constraint function

$$g_l(\mathbf{x}, \mathbf{P}) = \sum_{s:l \in L(s)} x_s - \check{c}_l(\mathbf{P}; \alpha_l, \beta_l)$$

is sufficiently close to zero, corresponding to an equilibrium between the ingress- and egress-rates at that link. A first-order approximation to this rule could be a periodic T-step every D iterations, where D is chosen sufficiently large so that each subproblem safely converges within the D iterates.

B. Subproblem Solution

The dual problem associated with each subproblem (23) is

$$\min_{\boldsymbol{\mu} \geq \mathbf{0}} \hat{q}_{\tilde{\mathbf{P}}}(\boldsymbol{\mu}) + \hat{q}_x(\boldsymbol{\mu}) \quad (24)$$

where $\boldsymbol{\mu} = [\mu_1, \dots, \mu_L]$ is a vector of dual variables, and

$$\hat{q}_{\tilde{\mathbf{P}}}(\boldsymbol{\mu}) = \min_{\tilde{\mathbf{P}}} \left\{ \hat{L}_{\tilde{\mathbf{P}}}(\tilde{\mathbf{P}}, \boldsymbol{\mu}) \triangleq \sum_{l \in \mathcal{L}} \omega e^{\tilde{P}_l} - \mu_l \check{c}_l(\tilde{\mathbf{P}}; \alpha_l, \beta_l) \right\} \quad (25a)$$

$$\hat{q}_x(\boldsymbol{\mu}) = \max_{\mathbf{x} \geq \mathbf{0}} \left\{ \hat{L}_x(\mathbf{x}, \boldsymbol{\mu}) \triangleq \sum_{s \in \mathcal{S}} U_s(x_s) - \sum_{l \in \mathcal{L}} \mu_l \sum_{s:l \in L(s)} x_s \right\} \quad (25b)$$

are partial dual functions, since each subproblem is separable.

1) *Link Algorithm:* Solving (25a) gives the optimum link powers for a fixed $\boldsymbol{\mu}$. The fixed-point update

$$P_l^{(t+1)} = \frac{W \mu_l^{(t)} \alpha_l}{\omega + \sum_{n \neq l} G_{nl} \mathcal{M}_n^{(t)}} \quad (26)$$

with

$$\mathcal{M}_n^{(t)} = W \mu_n^{(t)} \alpha_n \frac{\text{SIR}_n(\mathbf{P}^{(t)})}{G_{nn} P_n^{(t)}} \quad (27)$$

can be shown to converge to such a solution, where the proof is similar to the one provided for Lemma 2, available in [26].

Links update their power according to (26), with messages $\mathcal{M}_n^{(t)}$ passed via broadcasts from other links, in a similar way to the operation of Alg. A.

Like the convex problem (13) of the previous section, the partial Lagrangian functions $\hat{L}_x(\mathbf{x}, \boldsymbol{\mu})$ and $\hat{L}_{\tilde{\mathbf{P}}}(\tilde{\mathbf{P}}, \boldsymbol{\mu})$ are strictly concave and strictly convex, respectively, for $\boldsymbol{\mu}$ fixed. The solutions to (25) are therefore unique and the dual problem (24) can be solved by updating the link congestion prices μ_l by the gradient descent

$$\mu_l^{(t+1)} = \left[\mu_l^{(t)} + \epsilon \left\{ \sum_{s:l \in L(s)} x_s^{(t)} - \check{c}_l(\mathbf{P}^{(t)}; \alpha_l, \beta_l) \right\} \right]^+, \quad (28)$$

where ϵ is a sufficiently small step-size. We observe that these congestion prices are updated in exactly the same way as the well-known duality-model solution [20], except that the link capacity is no longer a constant: it is our approximation (22). Should the egress-rate of each link be artificially restricted to this approximate capacity $\check{c}_l(\mathbf{P}^{(t)}; \alpha_l, \beta_l)$ instead of the true value $c_l(\mathbf{P}^{(t)})$ given by (2), then it follows from [20] that the

network implicitly performs this update: at equilibrium, the congestion prices are given by $\mu_l^* = b_l^*/\check{c}_l^*$, where b_l is the buffered backlog at the link.

2) *Source Algorithm*: Solving (25b) gives the optimum source rate allocation for a fixed $\boldsymbol{\mu}$. We need only consider the stationary point of the Lagrangian function $\hat{L}_x(\mathbf{x}, \boldsymbol{\mu})$ since is strictly concave in \mathbf{x} . This results in the well-known duality-model solution [20]:

$$x_s^{(t+1)} = U_s'^{-1} \left(\sum_{l \in L(s)} \mu_l^{(t)} \right), \quad (29)$$

where $U_s'^{-1}(\cdot)$ is the inverse of the first derivative of the utility. The inverse is always guaranteed to exist, since each utility function is strictly increasing.

This is an important result. It implies that existing TCP algorithms can be used to allocate source rates in this cross-layer optimization. We can therefore reuse existing TCP algorithms that are fully distributed, such as TCP Vegas that employs delay measurements for estimating the total congestion $\lambda^s = \sum_{l \in L(s)} \lambda_l^{(t)}$ en-route appearing in (29). Going further still, it enables a MANET cloud to become transparently interconnected between other wired internet networks that use existing TCP algorithms, and doing so without breaking end-to-end semantics.

C. Global Optimality

Although the canonical problem (8) has nonconvex form, the following results tell us that Alg. B can truly converge to a globally optimum resource allocation.

Lemma 5: All constraints are active at any (local or global) optimum solution to each of the canonical (8), convexified (13) and convex subproblems (23).

Proof: Observe that the objective function associated with each problem is a decreasing function of the link powers. In each problem, each constraint function can be represented by the inequality $f_l(\mathbf{x}) \leq g_l(\mathbf{P})$ where we understand that \mathbf{x} and \mathbf{P} are the logarithmically transformed variables in the case of the convexified problem. Suppose that at any (locally or globally) optimum solution, these inequalities are strict. We could then lower the link power P_l (or \hat{P}_l) so that we achieve equality. By

our earlier observation, this increases the objective and so our solution was clearly not (locally or globally) optimum. ■

Theorem 3: The series of convex approximations converges to a KKT-point of the original canonical problem (8).

Proof: Due to the separability of the original canonical problem, and each subproblem, we can investigate the KKT conditions for \mathbf{x} and \mathbf{P} separately.

We begin with the link powers. By Lemma 4, the sequence of convex subproblems converges when iteration t satisfies:

$$\check{c}_l(\mathbf{P}^{(t)}; \alpha_l^{(t)}, \beta_l^{(t)}) = c_l(\mathbf{P}^{(t)}) = \check{c}_l(\mathbf{P}^{(t)}; \alpha_l^{(t+1)}, \beta_l^{(t+1)}).$$

It follows that the solution $\mathbf{P}^* = \exp(\tilde{\mathbf{P}}^{(t)})$ satisfies (30), shown below, where (30a) is the KKT condition for the t th convex subproblem, and (30b) follows from (20a). It is easily verified that (30b) is exactly the KKT condition for the original canonical problem (8).

We now consider the source rates. The associated KKT condition of the t th convex subproblem is given by

$$\frac{\partial \hat{L}_x}{\partial x_s} = 0 = U_s'(x_s) - \sum_{l \in L(s)} \mu_l, \quad (31)$$

and is exactly the same condition for the original canonical problem (8).

We have demonstrated that any solution $\{\mathbf{x}^*, \mathbf{P}^*\}$ of the final subproblem in the series of approximations satisfies the KKT conditions of the original canonical problem (8). The proof is completed by recognizing that complementary slackness is always satisfied, since all constraints are active (by Lemma 5) and dual variables satisfy $\mu_l \geq 0$. ■

Corollary 1: When the condition of Theorem 2 is satisfied, the series of convex approximations converges to a globally optimum solution to the nonconvex canonical problem formulation (8).

Proof: With the condition of Theorem 2 satisfied, the problem (13) is convex. Therefore the associated KKT conditions become sufficient for global optimality. That is, the primal-dual triplet $\{\mathbf{x}^*, \mathbf{P}^*, \boldsymbol{\lambda}^*\}$ solves the convex problem (13) if conditions (32), shown below, are satisfied and $\{\mathbf{x}^*, \mathbf{P}^*\}$ is primal-feasible. Conditions (32a) and (32b) arise from the stationary points of the partial Lagrangian functions $L_{\tilde{\mathcal{P}}}$ and

$$\frac{\partial \hat{L}_{\tilde{\mathcal{P}}}}{\partial \hat{P}_l} = 0 = \omega P_l^* - W \mu_l \alpha_l + \sum_{n \neq l} W \mu_n \alpha_n \text{SIR}_n(\mathbf{P}^*) \frac{G_{nl} P_l^*}{G_{nn} P_n^*} \quad (30a)$$

$$= \omega P_l^* - W \mu_l \frac{\text{SIR}_l(\mathbf{P}^*)}{1 + \text{SIR}_l(\mathbf{P}^*)} \sum_{n \neq l} W \mu_n \frac{\text{SIR}_n^2(\mathbf{P}^*)}{1 + \text{SIR}_n(\mathbf{P}^*)} \frac{G_{nl} P_l^*}{G_{nn} P_n^*} \quad (30b)$$

$$\omega P_l - \frac{\lambda_l^*}{W \log(1 + \text{SIR}_l(\mathbf{P}^*))} \frac{\text{SIR}_l(\mathbf{P}^*)}{1 + \text{SIR}_l(\mathbf{P}^*)} + \sum_{n \neq l} \frac{\lambda_n^*}{W \log(1 + \text{SIR}_n(\mathbf{P}^*))} \frac{\text{SIR}_n^2(\mathbf{P}^*)}{1 + \text{SIR}_n(\mathbf{P}^*)} \frac{G_{nl} P_l^*}{G_{nn} P_n^*} = 0 \quad (32a)$$

$$U_s'(x_s) - \sum_{l \in L(s)} \frac{\lambda_l^*}{\sum_{m: l \in L(m)} x_m^*} = 0 \quad (32b)$$

$$\lambda_l^* \geq 0 \quad (32c)$$

L_x respectively (shown after returning to the $\{\mathbf{x}, \mathbf{P}\}$ -space). Condition (32c) follows from Lemma 5, and implies complementary slackness.

Suppose now that the sequence of convex approximations converge to the triplet $\{\hat{\mathbf{x}}^*, \hat{\mathbf{P}}^*, \boldsymbol{\mu}^*\}$, along with approximation vectors $\{\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*\}$. By our supposition, this solution is feasible and satisfies the KKT conditions (30b) and (31). Recall that the constraints are all active (Lemma 5) and that our link capacity approximation is exact (Lemma 4). That is, for all links $l \in \mathcal{L}$:

$$\sum_{m:l \in L(m)} \hat{x}_m^* = \check{c}_l(\hat{\mathbf{P}}^*; \boldsymbol{\alpha}_l^*, \boldsymbol{\beta}_l^*) = c_l(\hat{\mathbf{P}}^*). \quad (33)$$

We can now construct new dual-variables

$$\hat{\lambda}_l^* = \mu_l^* c_l(\hat{\mathbf{P}}^*) = \mu_l^* \sum_{m:l \in L(m)} \hat{x}_m^* \geq 0, \quad (34)$$

by making use of condition (33) and recognizing that each term is nonnegative. By inspection, the new triplet $\{\hat{\mathbf{x}}^*, \hat{\mathbf{P}}^*, \hat{\boldsymbol{\lambda}}^*\}$ satisfies the KKT conditions (32a)–(32c), where we remind the reader of the link capacity definition (2). It then follows that $\{\hat{\mathbf{x}}^*, \hat{\mathbf{P}}^*\}$ is a globally optimum rate- and power-allocation. ■

Corollary 1 tells us that the multipliers $\boldsymbol{\mu}^*$ arising from the sequence of approximations are a scalar multiple of the optimum multipliers $\boldsymbol{\lambda}^*$ associated with the convex optimization (13). A straightforward rearrangement of (34) further reveals that they are exactly the normalized congestion prices (18b).

VIII. DISPOSING OF POWER-ALLOCATION OVERHEADS

In pursuit of the (globally) optimum solution to the canonical problem (8), algorithms A and B require explicit message passing between links for power-allocation. While broadcasts are a viable way in which to realize such inter-link communication, an ideal scheme would use some kind of indirect measurement to achieve a similar goal.

In this section, we formulate an algorithm having this property: each link's power-allocation is based on the locally measured interference level caused by other links. No explicit message passing is required. For the logarithmic utility functions of TCP Vegas, we outline a scheme that makes use of limited message passing only for source rate allocation. These messages are embedded in a header field within ACK packets that traverse the source's reverse-path and presents little overhead.

Of course, we can opt to simply and completely omit passing of messages \mathcal{M}_n in an implementation of Alg. B, creating a fully distributed algorithm at the expense of optimality. How its subsequent performance compares to the following development is an interesting direction of future research.

1) *Link Algorithm*: The following development pivots on the fact that, at the optimum solution, the ingress rate to a link is matched to its capacity (all constraints are active). Considering a particular link l , we can rearrange the constraint in terms of the link transmit power, resulting in the following power update:

$$P_l^{(t+1)} = \frac{P_l^{(t)}}{\text{SIR}_l(\mathbf{P}^{(t)})} \left[\exp \left(\frac{1}{W} \sum_{s:l \in L(s)} x_s \right) - 1 \right]. \quad (35)$$

This update rule is very simple. The fraction outside of the brackets is a scaled noise measurement, observed to be independent of $P_l^{(t)}$ on comparison with (3), while the bracketed quantity makes use of an estimate of the ingress rate. With \mathbf{x} fixed, we can interpret the bracketed quantity as a SIR-threshold; convergence of the update then follows from [32].

2) *Source Algorithm: General Utility*: We now obtain a rate-allocation under the assumption that link powers are fixed. Substituting (35) into (8) results in the unconstrained optimization

$$\max_{\mathbf{x} \geq \mathbf{0}} \sum_{s \in \mathcal{S}} U_s(x_s) - \omega \sum_{l \in \mathcal{L}} \frac{P_l}{\text{SIR}_l(\mathbf{P})} \left[\exp \left(\sum_{s:l \in L(s)} \frac{x_s}{W} \right) - 1 \right]. \quad (36)$$

This approach is not entirely ideal: it does not guarantee that we will arrive at the jointly optimum rate- and power-allocation of the original problem (8). To do so would require that we instead consider the above optimization with P_j replaced by the function $P_j^*(\mathbf{x})$. This function, in vector form $\mathbf{P}^*(\mathbf{x}) = [P_1^*(\mathbf{x}), \dots, P_L^*(\mathbf{x})]$, gives the simultaneous minimum power solution to (35) for the specified \mathbf{x} . Unfortunately we do not have a closed-form expression for this function: it is the solution to L coupled nonlinear equations. In the previous sections, the congestion prices λ_l played a go-between the physical and transport layers and helped us avoid this issue.

Returning to (36) with \mathbf{P} fixed, we observe that the objective is strictly concave in \mathbf{x} . We can therefore make use of the source-rate update

$$x_s^{(t+1)} = \left[x_s^{(t)} + \epsilon \partial_s^{(t)} \right]^+ \quad (37)$$

to find the optimum rate-allocation, where ϵ is a sufficiently small step-size and

$$\partial_s^{(t)} = U'_s(x_s^{(t)}) - \frac{\omega}{W} \sum_{l \in L(s)} \frac{P_l}{\text{SIR}_l(\mathbf{P})} \exp \left(\frac{1}{W} \sum_{n:l \in L(n)} x_n^{(t)} \right)$$

is an ascent direction for source s .

3) *Source Algorithm: Vegas Utilities*: An alternative to the gradient-update (37) is the nonlinear Gauss-Seidel algorithm that solves (36) by successively maximizing the objective in each component x_s while holding all others fixed [5].

For the logarithmic utility function $U_s(x_s) = \alpha_s d_s \log x_s$ of TCP Vegas, we can find the component-wise maximizer by considering the stationary point of the objective, with all other source rates fixed. This leads to the fixed-point equation

$$x_s = \frac{W}{\omega} \frac{\alpha_s d_s}{\sum_{l \in L(s)} \Phi_l^s(\mathbf{x}_{-s})} \exp(-x_s/W) \quad (38)$$

where

$$\Phi_l^s(\mathbf{x}_{-s}) = \frac{P_l}{\text{SIR}_l(\mathbf{P})} \exp \left(\frac{1}{W} \sum_{m:l \in L(m), m \neq s} x_m \right)$$

is considered fixed, as it depends on the vector \mathbf{x}_{-s} of source rates that does not include x_s . It can be shown that the associated fixed-point is not attractive over all $x_s \geq 0$ and so the

associated iterative rate-update may not converge. Fortunately, our next result provides a remedy.

Result 6: With $\phi \geq 0$ fixed, the equation

$$z = \log\left(\frac{\phi + e^z}{1 + z}\right) \quad (39)$$

has a unique fixed-point, that is attractive for $z \geq 0$. Moreover, it shares the same fixed-point as

$$z = \phi e^{-z}. \quad (40)$$

Proof: Omitted due to space restrictions; see [26]. ■

Making use of Result 6 with

$$\begin{cases} \phi = \frac{1}{\omega} \sum_{l \in L(s)} \frac{\alpha_s d_s}{\Phi_l^s(\mathbf{x}_{-s})} \\ z = x_s / W \end{cases}$$

results in the convergent source rate update:

$$x_s^{(t+1)} = x_s^{(t)} + W \log\left(\frac{1 + \frac{\alpha_s d_s}{\omega \mathcal{M}_s^{(t)}}}{1 + x_s^{(t)} / W}\right), \quad (41)$$

where

$$\begin{aligned} \mathcal{M}_s^{(t)} &= \sum_{l \in L(s)} \Phi_l^s(\mathbf{x}_{-s}) \exp\left(-x_s^{(t)} / W\right) \\ &= \sum_{l \in L(s)} \frac{P_l}{\text{SIR}_l(\mathbf{P})} \exp\left(\frac{1}{W} \sum_{m: l \in L(m)} x_m\right) \end{aligned} \quad (42)$$

are messages that are accumulated from each link along the route of a source s . Each of these terms involves a noise measurement that is scaled by the total ingress rate, a per-link, locally measurable quantity. As in Section VI-B2, these terms may be progressively accumulated into an ACK packet header as it traverses from the receiver back to the source.

This source rate update is guaranteed to converge should (41) and (42) be iterated until convergence for a single source s , holding fixed other sources, and before proceeding to the next source. In practice, we have observed it to converge even when these messages are used for continuous asynchronous updates.

IX. NUMERICAL RESULTS

In this section, we provide illustrative numerical examples of Algs. A–C, where we highlight the speed of convergence, suitability under composite-fading and tracking ability.

We consider a MANET with topology illustrated in Fig. 3. Each link is assigned a random CDMA spreading sequence of unit energy having length $N = 8$. A transmission bandwidth of 1 MHz is selected, giving a $W = 125$ kHz baseband on each link. We model the SIR gap as $K = -1.5 / \log(5 \text{ BER})$ with $\text{BER} = 10^{-3}$ corresponding to MQAM modulation [11]. Vegas logarithmic utilities are employed throughout, for a fair comparison between algorithms. A scaled base-RTT of $\alpha_s d_s = 20$ ms

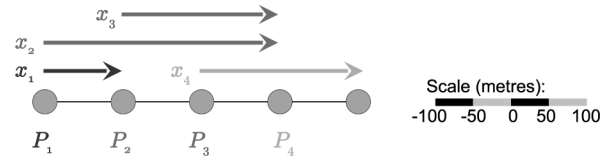


Fig. 3. MANET with four sources having routes shown. Nodes are spaced 100 m apart.

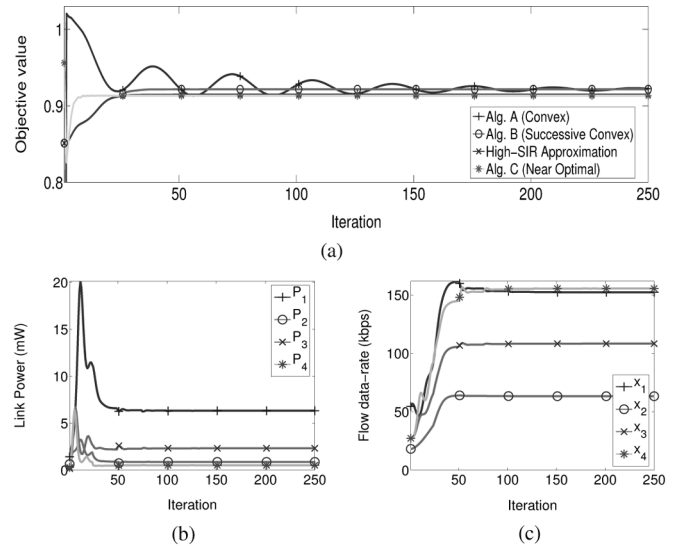


Fig. 4. Illustrative example (snapshot channel gains). (a) Convergence comparison (higher values are superior). (b) Link transmitter powers (Alg. B). (c) Source rates (Alg. B).

is arbitrarily chosen for each source. An equal trade-off between power-allocation and network-utility is chosen with $\omega = 1$. All simulations are initialized from an all-zero state.

A. Static Channel

We begin with the snapshot channel model, where fading is ignored (all $F_{ij} = 1$) and the slowly-varying gains are assumed fixed with $G_{ij} = S_{ij}(\mathbf{s}_i^T \mathbf{s}_j)^2$. Each $S_{ij} = d_{ij}^{-4}$ is a loss depending on distance d_{ij} from transmitter on link j to the receiver on link i .

Fig. 4(a) compares the evolution of all algorithms. Alg. B converges much faster than Alg. A, and both to the global optimum within 50 and 250 iterations respectively. This is significant, as Alg. B has solved the nonconvex canonical problem formulation (8) for the global optimum with a distributed algorithm, while preserving the existing TCP Vegas stack: no explicit message passing is required for source rate allocation. Corresponding power- and rate-allocations are shown in Figs. 4(b) and (c).

While Alg. C is clearly not optimal, it is very close. Performance is observed to be similar to the high-SIR based protocol, although it was observed to use roughly double the transmission power as compared to the optimal (Alg. A and B). Of great importance is the observed rate of convergence—even though only one sub-iteration of the fixed-point updates (41) and (42) are taken.

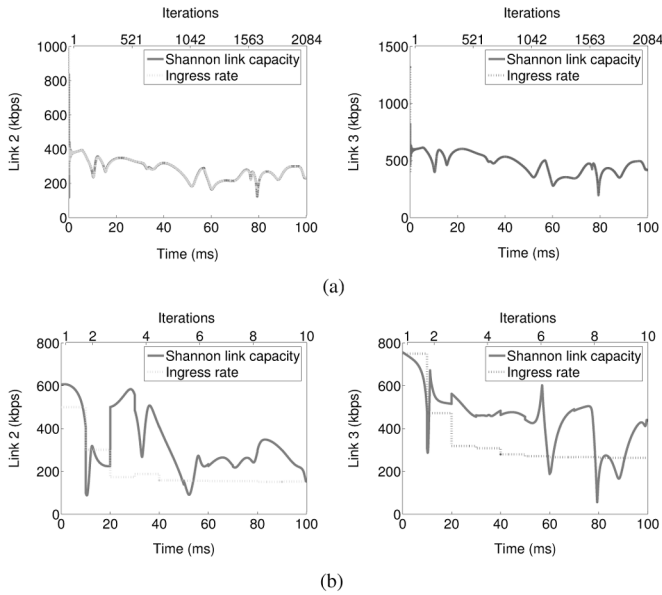


Fig. 5. Snapshot- and composite-fading algorithm comparison. (a) Snapshot formulation (10 kHz update rate). (b) Composite formulation (0.1 kHz update rate).

B. Composite Fading Channel

We now consider a channel subject to fading, where we compare the snapshot- and composite fading-based schemes. Only Alg. C is considered, for brevity and due to its attractive convergence rate and fully-distributed power-updates.

At each symbol period t , fading realizations $F_{ij}^{(t)} G_{ij}$ are computed with Clarke's model [28, Ch. 5] at a carrier frequency of 2.4 GHz and maximum node velocity 5 km/hr.¹ Each slowly-varying gain G_{ij} is fixed (taking the same value as the static scenario). Gains are normalized so the time-series $F_{ij}^{(t)}$ has unity mean for all i, j . We make use of the same sequence of channel realizations for each simulation, under the assumption of perfect channel knowledge.

Although the snapshot-based scheme assumes fixed channels, we implement it blindly though changes occur between iterates. Experimentally, we found a minimum update frequency of 10 kHz sufficient for the algorithm to adequately track the fading without severe fading-induced outage, illustrated in Fig. 5(a) for links 2 & 3. Recall that each update includes message-passing of (42) along each route. At a rate of 10 kHz, this presents an prohibitive implementation overhead.

An update frequency of just 0.1 kHz was found to be sufficient for the composite fading based algorithm to converge within just 50 ms from an all-zero state, as illustrated in Fig. 5(b). In comparison to the snapshot-based algorithm, this update frequency is much more manageable, although at the expense of lower source rates and a limited amount of fading-induced congestion, in this case, controlled by a rate-outage target of $\Omega_l^{\text{rate}} = 20\%$ for all links.

In reality, perfect channel state is not available and we must obtain estimates of the slowly-varying gains G_{ij} from the composite instantaneous fading state $F_{ij}^{(t)} G_{ij}$. This can be achieved

¹We use $v = 10$ km/hr to calculate the maximum Doppler frequency for the worst-case situation of nodes moving in opposite directions.

for example, by averaging out the fast-fading gains $F_{ij}^{(t)}$ with a first-order filter. The iterative algorithm would then track estimates of G_{ij} until some degree of accuracy is achieved. In practice, the gains G_{ij} also vary with time and so the algorithm should run continuously, allowing changes in these gains to be tracked.

X. CONCLUSION

This paper has revisited a cross-layer design problem for MANETs involving power- and rate-allocation. The primary focus was on the analysis and solution of the underlying canonical optimization problem. Complicated by Shannon link capacity terms that are neither convex nor concave, previous attempts at solution involved high-SIR approximations that can result in allocations far from the true global optimum.

We have shown that the Shannon link capacity is in fact quasiconcave and, under a suitable range and domain transformation, can be "concavified" without resorting to approximation. These new developments revealed that the canonical problem is actually convex, under a transformation and suitable choice of NUM utility function. A new optimal protocol was subsequently derived, devoid of any high-SIR approximation.

Due to message-passing requirements of the optimal protocol, a second new solution method was proposed, involving a series of convex approximations. Each made use of the existing TCP stack for rate-allocation—a desirable attribute since these TCP protocols are distributed in nature. Convergence results to the global optimum of the underlying nonconvex canonical problem formulation were given when utilities are within the family already mentioned. It was further outlined how each approximate problem may be solved only partially, leading to a new distributed protocol.

A third protocol was also developed, having near-optimal performance in simulation with a convergence rate orders of magnitude faster than previous developments. It enjoyed the additional advantage of fully-distributed measurement based power-allocation, greatly simplifying implementation of such cross-layer designs in practice.

The results of this paper extend beyond the applications described. Many existing problems in the literature relying on high-SIR convex approximations can now be solved for true global optimality with the results and ideas presented herein.

REFERENCES

- [1] M. Avriel and A. C. Williams, "Complementary geometric programming," *SIAM J. Appl. Math.*, vol. 19, no. 1, pp. 125–141, Jul. 1970.
- [2] M. Avriel, W. E. Diewert, S. Schaible, and I. Zang, *Generalized Convexity*. New York, London: Plenum Press, 1988.
- [3] A. Ben-Tal, "On generalized means and generalized convex functions," *J. Opt. Theory Appl.*, vol. 21, no. 1, pp. 1–13, Jan 1977.
- [4] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995.
- [5] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [6] T. Donald and L. Massoulié, "Impact of fairness on internet performance," in *Proc. ACM SIGMETRICS*, 2001, pp. 82–91.
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [8] S. Chen, M. A. Beach, and J. P. McGeehan, "Division-free duplex for wireless applications," *Electron. Lett.*, vol. 34, no. 2, pp. 147–148, Jan 1998.

- [9] M. Chiang and J. Bell, "Balancing supply and demand of bandwidth in wireless cellular networks: Utility maximization over powers and rates," in *Proc. IEEE INFOCOM*, 2004, vol. 4, pp. 2800–2811.
- [10] M. Chiang, "Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 104–116, 2005.
- [11] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
- [12] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, 2nd ed. New York: Springer-Verlag, 1993.
- [13] M. Johansson, L. Xiao, and S. Boyd, "Simultaneous routing and power allocation in CDMA wireless data networks," in *IEEE Int. Conf. Communications*, May 2003.
- [14] D. Julian, M. Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," in *Proc. IEEE INFOCOM*, June 2002, vol. 2, pp. 477–486.
- [15] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, 2002.
- [16] F. P. Kelly, "Charging and rate control for elastic traffic," *Euro. Trans. Telecom.*, vol. 8, pp. 33–37, Jan 1997.
- [17] U. C. Kozat, I. Koutsopoulos, and L. Tassiulas, "A framework for cross-layer design of energy-efficient communication with QoS provisioning in multi-hop wireless networks," in *Proc. IEEE INFOCOM*, Mar. 2004, vol. 2, pp. 1446–1456.
- [18] S. Low, "A duality model of TCP and queue management algorithms," *J. ACM*, vol. 11, no. 4, pp. 525–536, 2003.
- [19] S. Low and D. Lapsley, "Flow control, I: Basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861–875, Dec. 1999.
- [20] S. Low, L. Peterson, and L. Wang, "Understanding TCP Vegas: A duality model," *J. ACM*, vol. 49, no. 2, pp. 207–235, 2002.
- [21] R. Madan, S. Cui, S. Lall, and A. Goldsmith, "Cross-layer design for lifetime maximization in interference-limited wireless sensor networks," in *Proc. IEEE INFOCOM*, Mar. 2005.
- [22] B. R. Marks and G. P. Wright, "A general inner approximation method for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.
- [23] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
- [24] K.-H. Pan, H.-K. Wu, R.-J. Shang, F. Lai, and Y.-W. Lin, "Communications over two-way waveform channels in wireless networks," in *Proc. 1999 IEEE Canadian Conf. Electrical and Computer Engineering*, Edmonton, Canada, May 1999, pp. 45–50.
- [25] J. Papandriopoulos, J. S. Evans, and S. Dey, "Optimal power control for Rayleigh-faded multiuser systems with outage constraints," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 2705–2715, Nov. 2005.
- [26] J. Papandriopoulos, "Resource optimization in multiuser communication networks," Ph.D. dissertation, University of Melbourne, Melbourne, Australia, 2006.
- [27] X. Qiu and K. Chawla, "On the performance of adaptive modulation in cellular systems," *IEEE Trans. Commun.*, vol. 47, no. 6, pp. 884–895, Jun. 1999.
- [28] T. S. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.
- [29] J. B. Rosen, "Iterative solution of nonlinear optimal control problems," *SIAM J. Control*, vol. 4, pp. 223–244, 1966.
- [30] K. Tsubouchi, H. Nakase, A. Namba, and K. Masu, "Full duplex transmission operation of a 2.45-GHz asynchronous spread spectrum using a SAN convolver," *IEEE Trans. Ultrason. Ferroelectr. Freq. Contr.*, vol. 40, no. 5, pp. 478–482, Sep. 1993.
- [31] J. Wang, S. H. L. Li, and J. C. Doyle, "Cross-layer optimization in TCP/IP networks," *IEEE/ACM Trans. Netw.*, vol. 13, no. 3, pp. 582–268, Jun. 2005.
- [32] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, 1995.



John Papandriopoulos (S'01) was born in Melbourne, Australia, in 1978. He received the combined B.E. degree in communications engineering and the B.App. Sci. degree in computer science from the Royal Melbourne Institute of Technology (RMIT University), in 2001. He received the Ph.D. degree in electrical engineering from the University of Melbourne, Australia, in 2006.

His research interests are in nonconvex optimization, DSL & CDMA resource allocation and cross-layer design. He enjoys delving into practical engineering implementations, and has spent some time with Telstra, Agilent Technologies, and the 3G Mobile R&D Division of NEC Australia.

Dr. Papandriopoulos received the J. N. McNicol Prize (University Medal) upon graduating from RMIT University in 2002, and the University of Melbourne Chancellor's Prize for excellence in his Ph.D. thesis in 2007. In 2005, he received a Victoria Fellowship from the Victoria State Government of Australia. From 2003, he has served as the Chairperson of the University of Melbourne IEEE Student Branch for two years.



Subhrakanti Dey (S'94–M'96) was born in Calcutta, India, in 1968. He received the B.Tech. and M.Tech. degrees from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, India, in 1991 and 1993, respectively, and the Ph.D. degree from the Department of Systems Engineering, Research School of Information Sciences and Engineering, Australian National University, Australia, in 1996.

He has been with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Australia, since February 2000, where he is currently a full Professor. From September 1995 to September 1997, and September 1998 to February 2000, he was a Postdoctoral Research Fellow with the Department of Systems Engineering, Australian National University. From September 1997 to September 1998, he was a Postdoctoral Research Associate with the Institute for Systems Research, University of Maryland, College Park. His current research interests include signal processing for telecommunications, wireless communications, and networks, performance analysis of communication networks, stochastic and adaptive estimation and control, and statistical and adaptive signal processing.



Jamie Evans (S'93–M'98) was born in Newcastle, Australia, in 1970. He received the B.S. degree in physics and the B.E. degree in computer engineering from the University of Newcastle, in 1992 and 1993, respectively, where he received the University Medal upon graduation. He received the M.S. and the Ph.D. degrees from the University of Melbourne, Parkville, Australia, in 1996 and 1998, respectively, both in electrical engineering, and was awarded the Chancellor's Prize for excellence for his Ph.D. thesis.

From March 1998 to June 1999, he was a Visiting Researcher in the Department of Electrical Engineering and Computer Science, University of California, Berkeley. He returned to Australia to take up a position as Lecturer at the University of Sydney, Australia, where he stayed until July 2001. Since that time, he has been with the Department of Electrical and Electronic Engineering, University of Melbourne, where he is now an Associate Professor. His research interests are in communications theory, information theory, and statistical signal processing with current focus on wireless communications networks.