Robust Routing in Wireless Multi-Hop Networks*

Yuchen Wu, Alejandro Ribeiro and Georgios B. Giannakis[†]
Dept. of ECE, University of Minnesota
200 Union Street S.E., Minneapolis, MN 55455

***Abstract*— This paper introduces a robust approach to stochastic multi-hop routing for wireless networks when the quality of links is modelled through a reliability matrix R. Yielding to the practical constraint that link reliabilities have to be measured, we consider that R is random with known mean and variance. Thus, network utilities are also random quantities. Robust routing algorithms are then introduced to maximize an average utility subject to a variance constraint; or, alternatively, to minimize variance subject to a minimum utility yield. We prove that both problems can be solved by convex programming techniques. We further show that the robust routing optimization problems exhibit a separable structure enabling the proposal of routing protocols based on communication with one-hop neighbors only. Although the communication cost to compute the optimal routes is thus significantly reduced, we show that there is no performance penalty with respect to optimal routes computed by a centralized algorithm.**

Keywords: Networking, Signal Processing, Communications

I. Introduction

The design of a routing algorithm can be likened to finding a path between an origin-destination (OD) pair which is optimal in a certain sense. In wired networks, this can be almost always reduced to the problem of finding the shortest path route between the OD pair using a properly defined cost, i.e., distance, between individual hops. In multi-hop wireless networks, shortest path routing is certainly a useful approach, see e.g., [10] and references therein. However, the definition of a link in a wireless network is somewhat arbitrary, because there is no tangible connection among nodes.

To be specific, let R_{ij} denote the probability that a packet transmitted by the j-th user U_j is correctly decoded by the i-th user U_i. In a wired network the reliability R_{ij} is either very close to 1, if there is a link between U_j and U_i, or 0, if there is not. In a wireless network, the whole range of R_{ij} values is possible, as testified by experimental measurements [1]. To deal with links of intermediate reliabilities, a number of works have advocated link metrics that depend on R_{ij}, e.g., [7]. In particular, the link cost $1/R_{ij}$ that penalizes but does not preclude the use of unreliable links has found widespread acceptance [5]. An alternative approach to shortest path routing is to formulate routing algorithms as network utility optimization problems based on the matrix $\mathbf{R}$ with entries R_{ij}. It has been recently shown that many routing schemes can be formulated as convex optimization problems thus ensuring algorithmic tractability [8]. Furthermore, dual decomposition techniques can be used to solve these optimization problems in a distributed manner [9]; see also [6], [4].

*Work in this paper was prepared through collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

Email at `(wuxxx315, aribeiro, georgios)@umn.edu`

As matrix $\mathbf{R}$ is not a priori known, it must be estimated in practice. For instance, U_j can estimate R_{ij} by computing the ratio of the number of acknowledgements received from U_i over the total number of transmitted packets. Regardless of the specific estimator used, routing decisions must be taken based on the estimated $\hat{\mathbf{R}}$. Thus, the optimal network utility – expressed in terms of $\hat{\mathbf{R}}$ – is itself an estimate of the real utility. In this context, a problem of interest here is to formulate routing problems taking into account the uncertainty in network utility induced by the uncertainty in reliability estimates. Finding routes that exhibit robustness against estimation errors is the problem addressed in this paper.

In particular, we develop routing algorithms so that: (P1) the variance of a network utility is minimized subject to a constraint in the utility; and (P2) a network utility is maximized subject to a constraint in its variance (Section II). The problem of finding optimal robust routes may or may not be tractable. Conditions ensuring convexity – thus tractability – of the problems considered are introduced in Section III. It turns out that the class of mean and variance utilities ensuring convexity of the optimization problem includes many cases of practical interest (Section III-A). Even though convexity ensures problem tractability, the communication cost associated with collecting reliability estimates at a central location followed by percolation of the optimal routing matrix may be prohibitive. This motivates the introduction of routing protocols based on local communications only that as time progresses converge to the optimal routing matrix (Section IV). We finally present corroborating simulations (Section V) and conclude the paper (Section VI).

II. Problem Formulation

Consider a group of J wireless terminals $\{U_j\}_{j=1}^{J}$ that collaborate in routing packets to any out of a set of J_{ap} access points (AP) $\{U_j\}_{j=J+1}^{J+J_{\text{ap}}}$. At any given time slot, U_j services a packet with rate μ_j that we assume is determined by the medium access control (MAC) layer. If it decides to send a packet, U_j would transmit it to U_i with probability T_{ij}. Instead of a routing table, stochastic routing algorithms search for an optimal matrix of probabilities $\mathbf{T} \in \mathbb{R}^{(J+J_{\text{ap}})\times(J+J_{\text{ap}})}$ with entries T_{ij}. To model the evolution of packets through the network we introduce a matrix $\mathbf{K} \in \mathbb{R}^{(J+J_{\text{ap}})\times(J+J_{\text{ap}})}$ such that K_{ij} denotes the probability of a packet moving from U_j's to U_i's queue. Since the packet moves from U_j's to U_i's queue if and only U_j transmits it to U_i, and U_i correctly receives it, we have that

$$K_{ij} = T_{ij}R_{ij} \quad \text{for } i \neq j, \tag{1}$$

Further, note that since $\mathbf{T}$ and $\mathbf{K}$ are stochastic matrices we have $\mathbf{T}^T\mathbf{1} = \mathbf{1}$ and $\mathbf{K}^T\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes the $(J+J_{\text{ap}})\times 1$ vector of all ones. To simplify notation, define the set $\mathcal{K} := \{\mathbf{K} : K_{ij} = T_{ij}R_{ij} \text{for } i \neq j, \mathbf{T}^T\mathbf{1} = \mathbf{1}, \mathbf{K}^T\mathbf{1} = \mathbf{1}\}$ of feasible matrices $\mathbf{K}$

that can be implemented – by proper selection of $\mathbf{T}$ – in a network described by $\mathbf{R}$. Note that $\mathcal{K}$ is a convex polyhedron.

Packet arrivals at U_j are modeled as a stationary random process with rate ρ_j. In addition to its own packets, U_j receives packets from other nodes for a total arrival rate λ_j. A sufficient condition for stability (i.e., for the number of packets in the j-th queue to be bounded with probability 1) is $\lambda_j < \mu_j$. Upon defining vectors $\boldsymbol{\rho} := [\rho_1, \ldots, \rho_J]^T$, $\boldsymbol{\mu} := [\mu_1, \ldots, \mu_J]^T$ and $\boldsymbol{\lambda} := [\lambda_1, \ldots, \lambda_J]^T$, it can be further proved that all queues will be stable so long as the $\boldsymbol{\rho}$ vector of arrival rates satisfies [8]

$$\boldsymbol{\rho} = (\mathbf{I} - \mathbf{K}_0)\boldsymbol{\lambda}, \quad \text{for some} \quad \boldsymbol{\lambda} \in [\mathbf{0}, \boldsymbol{\mu}] \tag{2}$$

where $\mathbf{K}_0$ denotes the $J \times J$ upper left sub-matrix of $\mathbf{K}$. The description of the stable rate region in (2) involves a bilinear function of the problem variables $\mathbf{K}$ and $\mathbf{T}$. It can be proven [8] that an alternative characterization is

$$\boldsymbol{\rho} = (\mathbf{I} - \mathbf{K}_0)\mathbf{1}, \quad \mathbf{T}^T\mathbf{1} \leq \boldsymbol{\mu}. \tag{3}$$

The set of feasible $\mathbf{K}$ matrices changes accordingly to $\mathcal{K} := \{\mathbf{K} : K_{ij} = T_{ij}R_{ij} \text{for } i \neq j, \mathbf{T}^T\mathbf{1} \leq \boldsymbol{\mu}, \mathbf{K}^T\mathbf{1} = \mathbf{1}\}$. Since (3) is a characterization in terms of linear functions of the problem variables, it is more tractable from an optimization perspective.

To model the fact that $\mathbf{R}$ is estimated, we consider that $\mathbf{R}$ is random with known mean and variance:

$$\hat{\mathbf{R}} := \mathrm{E}(\mathbf{R}) \tag{4}$$
$$\Sigma_{i_1j_1,i_2j_2} := \mathrm{E}[(R_{i_1j_1} - \hat{R}_{i_1j_1})(R_{i_2j_2} - \hat{R}_{i_2j_2})] > 0$$

implying that the rate in (3) is also random. We further assume that reliability estimates are never perfect, i.e., $\Sigma_{ij,ij} > 0$ whenever $\hat{R}_{ij} \neq 0$. Our goal here is to design robust routing algorithms that we define as follows:

(P1) Maximize a function of the expected value $\mathrm{E}(\boldsymbol{\rho}) = \mathrm{E}[(\mathbf{I} - \mathbf{K}_0)\mathbf{1}]$ subject to a constraint in the maximum tolerable variance $\mathrm{var}(\boldsymbol{\rho}) = \mathrm{var}[(\mathbf{I} - \mathbf{K}_0)\mathbf{1}]$[1]

$$\begin{aligned} \mathbf{T}^* = \quad & \arg\max_{K \in \mathcal{K}} f_0\left[[\mathrm{E}(\boldsymbol{I} - \boldsymbol{K}_0)\mathbf{1}]\right] \\ & \text{s.t.} \quad g_i\left[\mathrm{var}[(\boldsymbol{I} - \boldsymbol{K}_0)\mathbf{1}]\right] \leq g_{0i} \; i \in [1, M] \end{aligned} \tag{5}$$

where $\{g_i[\mathrm{var}(\boldsymbol{\rho})], g_{0i}\}_{i=1}^M$ describes M prescribed tolerances on variance utilities.

(P2) Minimize a function of the variance $\mathrm{var}(\boldsymbol{\rho})$ subject to a minimum requirement on a function of $\mathrm{E}(\boldsymbol{\rho})$; i.e., we seek

$$\begin{aligned} \mathbf{T}^* = \quad & \arg\max_{K \in \mathcal{K}} g_0\left[\mathrm{var}[(\boldsymbol{I} - \boldsymbol{K}_0)\mathbf{1}]\right] \\ & \text{s.t.} \quad f_i\left[\mathrm{E}[(\boldsymbol{I} - \boldsymbol{K}_0)\mathbf{1}]\right] \geq f_{0i} \quad i \in [1, N] \end{aligned} \tag{6}$$

where $\{f_i[\mathrm{E}(\boldsymbol{\rho})], f_{0i}\}_{i=1}^N$ describes N pre-specified mean rate utility requirements.

The goal of this paper is to: (i) compute the mean $\mathrm{E}(\boldsymbol{\rho})$ and variance $\mathrm{var}[\boldsymbol{\rho}]$ as functions of $\mathbf{R}$, $\mathbf{T}$ and $\Sigma_{i_1j_1,i_2j_2}$; and (ii) establish cases in which (6) and (5) are convex; (iii) identify conditions for (6) and (5) to be equivalent, and (iv) introduce a distributed implementation of (6)-(5).

[1]For a random vector $\mathbf{v}$ we adopt the notation $\mathrm{var}(\mathbf{v})$ to denote a vector with components $[\mathrm{var}(\mathbf{v})]_j := \mathrm{E}[(v_j - \mathrm{E}(v_j))^2]$. This is to be distinguished from the covariance matrix $\mathbf{Cov}(\mathbf{v}) := \mathrm{E}[(\mathbf{v} - \mathrm{E}(\mathbf{v}))^T(\mathbf{v} - \mathrm{E}(\mathbf{v}))]$. The two are related by $\mathrm{var}(\mathbf{v}) = \mathrm{tr}[\mathbf{Cov}(\mathbf{v})]$.

III. Optimal Robust Routes

To compute the mean and variance of $\boldsymbol{\rho}$ in terms of $\hat{\mathbf{R}}$ and $\Sigma_{i_1j_1,i_2j_2}$, start by noting that the j-th component of $\boldsymbol{\rho}$ can be written as [cf. (3)]

$$\rho_j = 1 - \sum_{i=1}^{J} K_{ji} = \sum_{i=1, i\neq j}^{J+J_{\mathrm{ap}}} K_{ij} - \sum_{i=1, i\neq j}^{J} K_{ji} \tag{7}$$

where in the second equality we replaced $1 - K_{jj} = \sum_{i=1,i\neq j}^{J+J_{\mathrm{ap}}} K_{ij}$ which follows from $\mathbf{K}$ being a stochastic matrix. We can now use the constraint in (1) to write

$$\rho_j = \sum_{i=1}^{J+J_{\mathrm{ap}}} R_{ij}T_{ij} - \sum_{i=1}^{J} R_{ji}T_{ji} \tag{8}$$

which shows that $\boldsymbol{\rho}$ is a linear function of the reliability matrix $\mathbf{R}$.

Since $\mathbf{R}$ is usually sparse many terms in the sum in (8) are null. To make this explicit, we define the set $c(j) := \{i : R_{ij} > 0;\ i \neq j,\ i \in [1, J + J_{ap}]\}$, representing the indices of terminals $\{U_i\}_{i=1}^{J+J_{ap}}$ that are able to decode U_j's transmission with non-zero probability. Likewise, define $r(j) := \{i : R_{ji} > 0;\ i \neq j,\ i \in [1, J]\}$ as the set of indices corresponding to terminals $\{U_i\}_{i=1}^{J}$ whose transmission can be decoded by U_j with nonzero probability. We can thus rewrite (8) as

$$\rho_j = \sum_{i \in c(j)} T_{ij}R_{ij} - \sum_{i \in r(j)} R_{ji}T_{ji}. \tag{9}$$

To further simplify notation define $\boldsymbol{r}_j := \boldsymbol{R}_{c(j),j}$ and $\boldsymbol{s}_j := \boldsymbol{R}_{j,r(j)}$ containing the non-zero elements of the j-th column and row of $\boldsymbol{R}$, respectively. In the same way define $\boldsymbol{t}_j := \boldsymbol{T}_{c(j),j}$ and $\mathbf{t}'_j := \boldsymbol{T}_{j,r(j)}$ to write

$$\rho_j = \mathbf{r}_j^T\mathbf{t}_j - \mathbf{s}_j^T\mathbf{t}'_j. \tag{10}$$

From (10) we can readily express $\mathrm{E}(\rho_j)$ in terms of the mean $\hat{\mathbf{R}} := \mathrm{E}(\mathbf{R})$ in (4). Noting that $\mathrm{E}(\mathbf{r}_j) = \mathrm{E}(\mathbf{R}_{c(j)}, j) = \hat{\mathbf{R}}_{c(j),j}$ and $\mathrm{E}(\mathbf{s}_j) = \mathrm{E}(\mathbf{R}_{j,c(j)}) = \hat{\mathbf{R}}_{j,c(j)}$ we can take expected value in (10) to obtain

$$\mathrm{E}(\rho_j) = \hat{\mathbf{R}}_{c(j),j}T\mathbf{t}_j - \hat{\mathbf{R}}_{j,c(j)}^T\mathbf{t}'_j := \hat{\mathbf{r}}_j^T\mathbf{t}_j - \hat{\mathbf{s}}_j^T\mathbf{t}'_j \tag{11}$$

where we defined $\hat{\mathbf{r}}_j := \hat{\mathbf{R}}_{c(j),j}$ and $\hat{\mathbf{s}}_j := \hat{\mathbf{R}}_{j,c(j)}$.

The rate variance $\mathrm{var}(\rho_j)$ can be analogously expressed in terms of the $\Sigma_{i_1j_1,i_2j_2}$ in (4). Indeed, upon defining the covariance matrices $\mathbf{Cov}(\mathbf{r}_j) := \mathrm{E}[(\mathbf{r}_j - \hat{\mathbf{r}}_j)(\mathbf{r}_j - \hat{\mathbf{r}}_j)^T]$ and $\mathbf{Cov}(\mathbf{s}_j) := \mathrm{E}[(\mathbf{s}_j - \hat{\mathbf{s}}_j)(\mathbf{s}_j - \hat{\mathbf{s}}_j)^T]$ and the cross-covariance $\mathbf{Cov}(\mathbf{r}_j, \mathbf{s}_j) := \mathrm{E}[(\mathbf{r}_j - \hat{\mathbf{r}}_j)(\mathbf{s}_j - \hat{\mathbf{s}}_j)^T]$ we can write

$$\mathrm{var}(\rho_j) = \mathbf{t}_j^T\mathbf{Cov}(\mathbf{r}_j)\mathbf{t}_j - 2\mathbf{t}_j^T\mathbf{Cov}(\mathbf{r}_j, \mathbf{s}_j)\mathbf{t}'_j + \mathbf{t}'^T_j\mathbf{Cov}(\mathbf{s}_j)\mathbf{t}'_j. \tag{12}$$

The (i,k)-th entry of $\mathbf{Cov}(\mathbf{r}_j)$ is given by $[\mathbf{Cov}(\mathbf{r}_j)]_{ik} = \Sigma_{ij,kj}$. Likewise, we have $[\mathbf{Cov}(\mathbf{s}_j)]_{ik} = \Sigma_{ji,jk}$ and $[\mathbf{Cov}(\mathbf{r}_j, \mathbf{s}_j)]_{ik} = \Sigma_{ij,jk}$. For future reference, note that the rate variance in (12) is a positive definite quadratic form in the transmission probabilities $(\mathbf{t}_j, \mathbf{t}'_j)$

A particular case of practical relevance is when the estimation of $\mathbf{R}$ is carried out componentwise, e.g, when R_{ij} is estimated by U_j as the ratio between the number of acknowledgements received from U_i and the number of packets sent to U_i. In this case the $\hat{R}_{ij}$ estimates are independent implying that $\mathrm{E}[(R_{i_1j_1} - \hat{R}_{i_1j_1})(R_{i_2j_2} -$

$\hat{R}_{i_2 j_2})] = \Sigma_{i_1 j_1, i_2 j_2} = 0$ for $(i_1, j_1) \neq (i_2, j_2)$. Thus, the cross-covariance $\mathbf{Cov}(\mathbf{r}_j, \mathbf{s}_j) = \mathbf{0}$ and the covariances $\mathbf{Cov}(\mathbf{r}_j)$ and $\mathbf{Cov}(\mathbf{s}_j)$ are diagonal. We have

$$\mathrm{var}[\rho_j] = \sum_{i=1}^{J+J_{\mathrm{ap}}} T_{ij}^2 \Sigma_{ij} + \sum_{i=1}^{J} T_{ji}^2 \Sigma_{ji} \tag{13}$$

where we defined $\Sigma_{ij} := \mathrm{E}[(R_{ij} - \hat{R}_{ij})^2] = \Sigma_{ij,ij}$.

A. Convexity of robust routing problems

Substituting (11) and (12) (or (13) if the R_{ij} estimates are independent) into (6) and (5), we obtain an optimization problem that can, in principle, be solved to obtain the optimal routing matrix $\mathbf{T}^*$. Solving these optimization problems might, or might not be tractable. Under proper conditions however, we can guarantee that (6) and (5) are convex, as we assert in the following proposition.

Proposition 1 *Consider the optimal robust routing problems in* (6) *and* (5) *and assume that (h1) the functions $f_i[\mathrm{E}(\boldsymbol{\rho})]$ are concave for $i \in [0, N]$ and (h2) the functions $g_i[\mathrm{var}(\boldsymbol{\rho})]$ are convex and nondecreasing in each argument. Then, the optimization problems in* (6) *and* (5) *are convex.*

Proof: Since $\mathbf{T}$ is constrained by a set of linear inequalities ($\mathbf{T}^T\mathbf{1} \leq \boldsymbol{\mu}$ and $\mathbf{T} \geq \mathbf{0}$), to prove that the problem in (6) is convex, it suffices to prove that: i) $g_0[\mathrm{var}(\boldsymbol{\rho})]$ is a convex function of the routing matrix $\mathbf{T}$; and ii) $f_i[\mathrm{E}(\boldsymbol{\rho})]$ for $i \in [1, N]$ is a concave function of $\mathbf{T}$. Correspondingly, (5) will be convex as long as: iii) $f_0[\mathrm{E}(\boldsymbol{\rho})]$ is a concave function of $\mathbf{T}$; and iv) $g_i[\mathrm{var}(\boldsymbol{\rho})]$ for $i \in [1, N]$ a convex function of $\mathbf{T}$. Thus, the claim follows if (c1) $f_i[\mathrm{E}(\boldsymbol{\rho})]$ is a concave function of $\mathbf{T}$; and (c2) $g_i[\mathrm{var}(\boldsymbol{\rho})]$ a convex function of $\mathbf{T}$ for $i \in [0, N]$.

The latter follows from the composition rules of convex analysis [3, Sec.3.2.4]. Indeed, $\mathrm{E}[\boldsymbol{\rho}]$ is a linear function of $\mathbf{T}$. Composition of the concave function $f_i[\mathrm{E}(\boldsymbol{\rho})]$ [cf. (h1)] with the linear function $\mathrm{E}[\boldsymbol{\rho}]$ [cf. (11)], yields a concave function implying (c1). To prove (c2) recall that $\mathrm{var}(\rho_j)$ is a positive definite quadratic form with variables $\mathbf{t}_j$, and thus convex (indeed, strictly convex). The composition of the convex and nondecreasing in each argument function $g_i[\mathrm{var}(\boldsymbol{\rho})]$ [cf. (h2)] with the convex function $\mathrm{var}(\rho_j)$ [cf. (12)] is convex establishing (c2). ■

Under (mild) restrictions on the utility functions $f_i[\mathrm{E}(\boldsymbol{\rho})]$ and $g_i[\mathrm{var}(\boldsymbol{\rho})]$, Proposition (1) ensures tractability of (6) and (5). Consequently, interior point methods can be used to solve these problems with affordable complexity in the order $O(J^{3.5})$.

The conditions (h1) and (h2) are satisfied in many practical cases. Some examples are given next.

Maximum rate utility with bounded variance. A typical example of a problem of the form in (P1) is to consider the maximization of a weighted sum of rates $\mathbf{w}^T\mathrm{E}(\boldsymbol{\rho})$. The variance of the individual rates is further upper bounded by a certain tolerance v_{0j} yielding the problem

$$\begin{aligned} \max \quad & \sum_{j=1}^{J} w_j \left(\hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^T \mathbf{t}_j' \right) \\ \text{s.t.} \quad & \mathrm{var}(\rho_j) \leq v_{0j}, \quad \mathbf{t}_j \geq 0,\ \mathbf{t}_j^T \mathbf{1} \leq \mu_j, \quad j \in [1, J] \end{aligned} \tag{14}$$

The functions $f_0[\mathrm{E}(\boldsymbol{\rho})] := \mathbf{w}^T\mathrm{E}(\boldsymbol{\rho})$ and $g_j[\mathrm{var}(\boldsymbol{\rho})] = \mathrm{var}(\rho_j)$ satisfy the hypothesis (h1) and (h2) of Proposition 1 proving that the problem in (14) is convex. This can be verified by noting that the argument to be optimized is a linear function of the $\mathbf{t}_j$, and that the constraint $\mathrm{var}(\rho_j) \leq v_{0j}$ is a positive definite quadratic form on $\mathbf{t}_j$.

Different rate utilities can be used in the argument of (14). E.g., the minimum rate utility $\min_{j\in[1,J]}[\mathrm{E}(\rho_j)] = \min_{j\in[1,J]} \left(\hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^T \mathbf{t}_j' \right)$ is consider a fair alternative since it maximizes the rate of the least favored terminal. The sum of logarithms utility $\sum_{j=1}^J \log[\mathrm{E}(\rho_j)] = \sum_{j=1}^J \log \left(\hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^T \mathbf{t}_j' \right)$, is regarded as an intermediate point between weighted sum and minimum rate.

Minimum variance with rate guarantees. Alternatively, we may aim to comply with a minimum rate requirement ρ_{0j} for each terminal U_j, while minimizing the norm of the variance vector, i.e.,

$$\begin{aligned} \min \quad & \|\mathrm{var}(\boldsymbol{\rho})\| \\ \text{s.t.} \quad & \hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^T \mathbf{t}_j' \geq \rho_{j0}, \quad \mathbf{t}_j \geq 0,\ \mathbf{t}_j^T \mathbf{1} \leq \mu_j, \quad j \in [1, J], \end{aligned} \tag{15}$$

Any norm verifies (h1) and (h2) of Proposition 1 establishing the convexity of the problem in (15). As a particular case, for the 1-norm we have $\|\mathrm{var}(\boldsymbol{\rho})\|_1 = \sum_j^J \mathrm{var}(\rho_j)$ because variances are always non-negative. For the ∞-norm, $\|\mathrm{var}(\boldsymbol{\rho})\|_\infty = \max_{j\in[1,J]} \mathrm{var}(\rho_j)$. If the $\hat{\mathbf{R}}$ estimate were perfect, (15) would guarantee rates $\{\rho_{0j}\}_{j=1}^J$. In the presence of estimation uncertainty, (15) attempts the same while, in some sense, maximizing the likelihood of this actually happening.

We have shown that finding the optimal solution to (5)-(6) incurs affordable computational complexity. However, it requires the entire reliability estimate $\hat{\mathbf{R}}$ and all the variances $\Sigma_{i_1 j_1, i_2 j_2}$, to be available at a central location, so that the optimization problem can be solved and the optimal routing matrix $\mathbf{T}^*$ distributed to the individual nodes. The drawbacks of this centralized approach are: i) a large communication cost to collect $\hat{\mathbf{R}}$ and $\Sigma_{i_1 j_1, i_2 j_2}$ and to distribute $\mathbf{T}^*$; ii) considerable delay to compute $\mathbf{T}^*$; and iii) lack of resilience to changes in the statistics of $\mathbf{R}$. These motivates distributed algorithms that we pursue next.

IV. Robust routing protocols

To develop a robust routing protocol, we introduce iterative algorithms to solve (P1) and (P2) in a distributed fashion. In such distributed algorithm U_j is interested in computing a sequence of iterates $\mathbf{t}_j(n)$ such that as $n \to \infty$ we have $\lim_{n\to\infty} \mathbf{t}_j(n) = \mathbf{t}_j^*$, with $\{\mathbf{t}_j^*\}_{j=1}^J$ denoting the solution of a problem of the form (P1) or (P2) for given functions $f_i(\mathrm{E}(\boldsymbol{\rho}))$ and $g_i(\mathrm{var}(\boldsymbol{\rho}))$.

Throughout this section we assume that

[A1] Terminal U_i can decode U_j with non-zero probability if and only if the probability of U_j decoding U_j is also non-zero, i.e., $R_{ij} \neq 0$ if and only if $R_{ji} \neq 0$.

[A2] The reliability estimates $\hat{R}_{ij}$ are independent, i.e., $\mathrm{E}[(R_{i_1 j_1} - \hat{R}_{i_1 j_1})(R_{i_2 j_2} - \hat{R}_{i_2 j_2})] = \Sigma_{i_1 j_1, i_2 j_2} = 0$ for $i_1, j_1 \neq i_2, j_2$.

A distributed algorithm can be developed using dual decomposition techniques. For simplicity of exposition we concentrate on the problem in (6) with $g_0[\mathrm{var}(\boldsymbol{\rho})] = \sum_j^J \mathrm{var}(\rho_j)$, and $f_j[\mathrm{E}(\boldsymbol{\rho})] = \mathrm{E}(\rho_j)$ for $j \in [1, J]$. This problems amounts to minimizing the sum of rate variances while satisfying a (expected) rate requirement

ρ_{0j} for terminal U_j. We further assume that the R_{ij} estimates are independent so that we can write the sum of variances as [cf. (13)]:

$$g_o[\mathrm{var}(\boldsymbol{\rho})] := \sum_{j=1}^{J} \mathrm{var}(\rho_j) \tag{16}$$
$$= \sum_{j=1}^{J} \left[\sum_{i=1}^{J+J_{\mathrm{ap}}} (\mu_j T_{ij})^2 \Sigma_{ij} + \sum_{i=1}^{J} (\mu_i T_{ji})^2 \Sigma_{ji} \right]$$

Rearranging terms in (16) and defining the constants

$$\begin{aligned} \alpha_{ij} &= 2\mu_j^2 \Sigma_{ij} \qquad \text{for } \ i \in [1, J] \\ \alpha_{ij} &= \mu_j^2 \Sigma_{ij} \qquad \text{for } \ i \in [J+1, J+J_{\mathrm{ap}}] \end{aligned} \tag{17}$$

it is easy to see that we can write $g_o[\mathrm{var}(\boldsymbol{\rho})] = \sum_{j=1}^{J} \sum_{i=1}^{J+J_{ap}} \alpha_{ij} T_{ij}^2$. We emphasize that being Σ_{ij} the variance of the R_{ij} estimate the constant α_{ij} is readily available at U_j.

Substituting the α_{ij} definitions in (16) into (17), and the result in (15) we obtain

$$\begin{aligned} \min \quad & \sum_{j=1}^{J} \mathbf{t}_j^t \mathbf{A}_j \mathbf{t}_j \\ \text{s.t.} \quad & \boldsymbol{t}_j \geq 0; \quad \mathbf{t}_j^T \mathbf{1} \leq \mu_j; \quad \hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^t \mathbf{t}_j' \geq \rho_{0j} \quad \forall j \end{aligned} \tag{18}$$

where $\boldsymbol{A}_j = \mathrm{diag}\{\alpha_{c(j)j}\}$, a diagonal matrix whose diagonal entries are $\alpha_{c(j),j}$.

A distributed optimization algorithm separates the original problem in J sub-problems involving variables $\mathbf{t}_j$ only. Even though the objective of the optimization problem in (18) is a summation of quadratic functions that depend on the local variables $\mathbf{t}_j$ and the constraints $\mathbf{t}_j \geq 0$ and $\mathbf{t}_j^T \mathbf{1} \leq \mu_j$ depend on $\mathbf{t}_j$ only, such separation cannot be achieved for the problem in (18). Indeed, the constraints $\hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^t \mathbf{t}_j' \geq \rho_{0j}$ involve the local variable $\mathbf{t}_j$, and the variable $\mathbf{t}_j'$ that contains transmission probabilities corresponding to U_j's neighboring terminals.

This problem can be overcome by resorting to the dual problem. Define a vector of Lagrange multipliers $\boldsymbol{\gamma} := [\gamma_1, \ldots, \gamma_J]^T$ with the multiplier γ_j associated with the constraint $\hat{\mathbf{r}}_j^T \mathbf{t}_j - \hat{\mathbf{s}}_j^t \mathbf{t}_j' \geq \rho_{0j}$ and write the Lagrangian:

$$\mathcal{L}(\{\mathbf{t}_j\}_{j=1}^{J}, \boldsymbol{\gamma}) = \sum_{j=1}^{J} \mathbf{t}_j^T \boldsymbol{A}_j \mathbf{t}_j + \sum_{j=1}^{J} \gamma_j \left(\rho_{0j} - \hat{\mathbf{r}}^t \mathbf{t}_j + \hat{\mathbf{s}}^T \mathbf{t}_j' \right) \tag{19}$$

that is defined over the set $\{\mathbf{t}_j : \mathbf{t}_j \geq 0, \mathbf{t}_j^T \mathbf{1} \leq \mu_j\}$.

The dual function is then defined as

$$q(\boldsymbol{\gamma}) = \min_{\{\mathbf{t}_j : \mathbf{t}_j \geq 0, \mathbf{t}_j^T \mathbf{1} \leq \mu_j\}} \mathcal{L}(\{\mathbf{t}_j\}_{j=1}^{J}, \boldsymbol{\gamma}). \tag{20}$$

Since the problem in (18) is convex strong duality holds and the optimal value of (18) can be found as $\max_{\boldsymbol{\gamma} \geq 0} q(\boldsymbol{\gamma})$.

An interesting observation is that the Lagrangian can be separated in J "local" Lagrangian containing $\mathbf{t}_j$ variables only. This property also allows a separable computation of the gradient of the dual function. These two properties are introduced in the following proposition.

Proposition 2 *Define the vector $\boldsymbol{\gamma}_j' := \boldsymbol{\gamma}_{c(j)}$ containing the multipliers of U_j's neighbors and the diagonal matrix $\boldsymbol{\Gamma}_j := \gamma_j \mathbf{I} - \mathrm{diag}(\boldsymbol{\gamma}_j')$. Define then the local Lagrangian:*

$$\mathcal{L}_j(\mathbf{t}_j, \gamma_j, \boldsymbol{\gamma}_j') = \mathbf{t}_j^T \boldsymbol{A}_j \mathbf{t}_j - \hat{\mathbf{r}}_j^T \boldsymbol{\Gamma}_j \mathbf{t}_j + \gamma_j \rho_{0j}. \tag{21}$$

We then have that

(i) The (global) Lagrangian $\mathcal{L}(\{\mathbf{t}_j\}_{j=1}^{J}, \boldsymbol{\gamma})$ is the sum of the local Lagrangians in (21), i.e.,

$$\mathcal{L}(\{\mathbf{t}_j\}_{j=1}^{J}, \boldsymbol{\gamma}) = \sum_{j=1}^{J} \mathcal{L}_j(\mathbf{t}_j, \gamma_j, \boldsymbol{\gamma}_j'). \tag{22}$$

(ii) Let $\mathbf{t}_j(\boldsymbol{\gamma}) := \arg\min_{\{\mathbf{t}_j \geq 0,\ \mathbf{t}_j^T \mathbf{1} \leq \mu_j\}_{j=1}^{J}} \mathcal{L}(\{\mathbf{t}_j\}_{j=1}^{J}, \boldsymbol{\gamma})$ be the arguments minimizing the Lagrangian in (20) for a given multiplier $\boldsymbol{\gamma}$. Then $\mathbf{t}_j(\boldsymbol{\gamma})$ is the argument minimizing the local Lagrangian, i.e.,

$$\mathbf{t}_j(\boldsymbol{\gamma}) = \arg \min_{\mathbf{t}_j \geq 0,\ \mathbf{t}_j^T \mathbf{1} \leq \mu_j} \mathcal{L}_j(\mathbf{t}_j, \boldsymbol{\gamma}). \tag{23}$$

(iii) The derivative of the dual function with respect to γ_j is given by

$$\frac{\partial q(\boldsymbol{\gamma})}{\partial \gamma_j} = \rho_{0j} - \hat{\mathbf{r}}^t \mathbf{t}_j(\boldsymbol{\gamma}) + \hat{\mathbf{s}}^T \mathbf{t}_j'(\boldsymbol{\gamma}) \tag{24}$$

with $\mathbf{t}_j(\boldsymbol{\gamma})$ and $\mathbf{t}_j'(\boldsymbol{\gamma})$ the solutions of (23).

Proof: Consider the sum of the local Lagrangians in (21) and use the definition of $\boldsymbol{\Gamma}_j := \gamma_j \mathbf{I} - \mathrm{diag}(\boldsymbol{\gamma}_j')$ to write.

$$\sum_{j=1}^{J} \mathcal{L}_j(\mathbf{t}_j, \gamma_j, \boldsymbol{\gamma}_j') = \sum_{j=1}^{J} \mathbf{t}_j^T \boldsymbol{A}_j \mathbf{t}_j + \gamma_j \rho_{0j} - \gamma_j \hat{\mathbf{r}}_j^T \mathbf{t}_j + \sum_{j=1}^{J} \hat{\mathbf{r}}_j^t \mathrm{diag}(\boldsymbol{\gamma}_j') \mathbf{t}_j \tag{25}$$

The last sum in (25) can be written as

$$\begin{aligned} \sum_{j=1}^{J} \hat{\mathbf{r}}_j^t \mathrm{diag}(\boldsymbol{\gamma}_j') \mathbf{t}_j &= \sum_{j=1}^{J} \sum_{i \in c(j)} \gamma_i \hat{R}_{ij} T_{ij} \\ &= \sum_{i=1}^{J} \gamma_i \sum_{j \in c(i)} \hat{R}_{ij} T_{ij} \\ &= \sum_{i=1}^{J} \gamma_i \hat{\mathbf{s}}_i^T \mathbf{t}_i' \end{aligned} \tag{26}$$

where in the first equality we use the definitions of $\hat{\mathbf{r}}_j^t$ and $\mathbf{t}_j$; in the second equality we interchange the order of summations and use the assumption $c(j) = r(j)$; and in the last equality we use the definitions of $\hat{\mathbf{s}}_j^t$ and $\mathbf{t}_j'$.

Substituting (26) in (25) we obtain

$$\sum_{j=1}^{J} \mathcal{L}_j(\mathbf{t}_j, \gamma_j, \boldsymbol{\gamma}_j') = \sum_{j=1}^{J} \mathbf{t}_j^T \boldsymbol{A}_j \mathbf{t}_j + \gamma_j \rho_{0j} - \gamma_j \hat{\mathbf{r}}_j^T \mathbf{t}_j + \gamma_j \hat{\mathbf{s}}_j^T \mathbf{t}_j' \tag{27}$$

Since the right hand sides of (27) and (20) coincide, (22) follows.

To prove (23) use (22) to write the optimal argument $\mathbf{t}_j(\boldsymbol{\gamma})$ as

$$\mathbf{t}_j(\boldsymbol{\gamma}) = \arg \min_{\{\mathbf{t}_j \geq 0,\ \mathbf{t}_j^T \mathbf{1} \leq \mu_j\}_{j=1}^{J}} \sum_{j=1}^{J} \mathcal{L}_j(\mathbf{t}_j, \gamma_j, \boldsymbol{\gamma}_j'). \tag{28}$$

The only term in the summation in (28) that contains $\mathbf{t}_j$ is $\mathcal{L}_j(\mathbf{t}_j, \gamma_j, \boldsymbol{\gamma}_j')$ from where (23) follows.

To obtain the result in (24) note that since $\mathbf{A}_j$ is positive definite ($\mathbf{A}_j$ is a diagonal matrix with strictly positive elements) it is invertible. Thus, for any $\boldsymbol{\gamma}$ there exists a unique minimizer $\mathbf{t}_j(\boldsymbol{\gamma})$ of $\mathcal{L}(\{\mathbf{t}_j\}_{j=1}^J$. The result then follows from Danskin's Theorem, [2, pp. 737] stating that: i) if there is a unique Lagrangian minimizer – i.e., a unique set $\{\mathbf{t}_j(\boldsymbol{\gamma})\}_{j=1}^J$ solving (20) – the dual function is differentiable; and ii) the derivative of the dual function with respect to γ_j is given by $\rho_{0j} - \hat{\mathbf{r}}^t \mathbf{t}_j(\boldsymbol{\gamma}) + \hat{\mathbf{s}}^T \mathbf{t}_j'(\boldsymbol{\gamma})$. ■

Using the separability properties described by Proposition 2 we can propose a distributed routing protocol. Indeed, since a gradient of the dual function can be computed using local and neighboring iterates, we can define a routing protocol by using a gradient ascent algorithm on the dual function. The routing protocol and the corresponding optimality claim is presented in the next proposition

Proposition 3 *Consider a routing protocol in which terminal U_j updates local primal and dual iterates denoted by $\mathbf{t}_j(n)$ and $\gamma_j(n)$ respectively. The iterates are updated according to the following rules:*

[P1] *Receive dual iterates $\gamma_j'(n)$ from neighboring terminals $\{U_i\}_{i\in c(j)}$.*

[P2] *Update the local primal iterates $\mathbf{t}_j(n)$ using*

$$\mathbf{t}_j(n) = \frac{1}{2}\left[\boldsymbol{A}_j^{-1}\left(\delta_j(n)\mathbf{1} + \boldsymbol{\Gamma}_j\right)\right]^+. \tag{29}$$

where $[\cdot]^+$ denotes projection to the nonnegative orthant and $\delta_j(n) \geq 0$ is chosen so that $\boldsymbol{t}_j^T(n)\mathbf{1} = \mu_j$. If for making $\boldsymbol{t}_j^T(n)\mathbf{1} = \mu_j$ we require $\delta_j(n) < 0$ we set $\delta_j(n) = 0$.

[P3] *Transmit the primal iterates $\mathbf{t}_j(n)$ to neighboring terminals $\{U_i\}_{i\in c(j)}$.*

[P4] *Receive primal iterates $\mathbf{t}_j(n)$ from neighboring terminals $\{U_i\}_{i\in c(j)}$.*

[P5] *Update the local dual iterates using*

$$\gamma(n+1) = \left[\gamma(n) + c\left(\rho_{0j} - \hat{\mathbf{r}}^t \mathbf{t}_j(n) + \hat{\mathbf{s}}^T \mathbf{t}_j'(n)\right)\right]^+. \tag{30}$$

with $c > 0$ a properly selected step size.

[P6] *Transmit the dual iterate $\gamma(n+1)$ to neighboring terminals $\{U_i\}_{i\in c(j)}$.*

For sufficiently small step size c, as $n \to \infty$, the local iterates $\mathbf{t}_j(n)$ converge to the optimal robust routes $\mathbf{t}_j^*$ solving the optimization problem in (18), i.e.,

$$\lim_{n\to\infty} \mathbf{t}_j(n) = \mathbf{t}_j^* \tag{31}$$

Proof: Start by noting that (29) is the solution of the local Lagrangian optimization in (21), i.e., $\mathbf{t}_j(n)$ in (29) is such that

$$\mathbf{t}_j(n) = \arg \min_{\mathbf{t}_j \geq 0,\ \mathbf{t}_j^T \mathbf{1} \leq \mu_j} \mathcal{L}_j(\mathbf{t}_j, \boldsymbol{\gamma}(n)). \tag{32}$$

Thus, according to (24) we have

$$\frac{\partial q[\boldsymbol{\gamma}(n)]}{\partial \gamma_j} = \rho_{0j} - \hat{\mathbf{r}}^t \mathbf{t}_j(n) + \hat{\mathbf{s}}^T \mathbf{t}_j'(n). \tag{33}$$

Consequently, the iteration in (30) is tantamount to gradient ascent for optimizing $q[(\boldsymbol{\gamma})]$, implying that for sufficiently small c

$$\lim_{n\to\infty} \gamma_j(n) = \gamma_j^* \tag{34}$$

with $\boldsymbol{\gamma}^* := [\gamma_1, \ldots, \gamma_J] = \arg\max_{\boldsymbol{\gamma}\geq 0} q(\boldsymbol{\gamma})$. But since the dual function is differentiable, convergence of $\gamma_j(n)$ implies convergence of $\mathbf{t}_j(n)$.

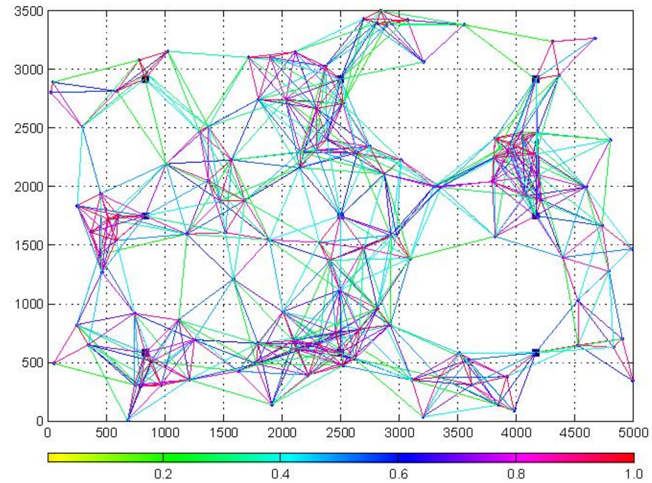

Fig. 1. Representation of network reliability estimates $\hat{\mathbf{R}}$, the color index represents the value of $\hat{R}_{ij}$ ($J = 120$ nodes shown as dots and $J_{\mathrm{ap}} = 9$ APs shown as boxes).

Steps [P1] and [P6] simply ensure that dual iterates are properly communicated and received. Steps [P3] and [P4] do the same for the primal iterates. ■

The distributed routing protocol [P1]-[P6] overcomes the limitations of a centralized implementation detailed at the end of Section III. Indeed, note that the proposed protocol i) requires communication with one-hop neighbors only, and ii) relies on knowledge of R_{ij} estimates and variances that either U_j or U_j's neighbors have available. Interestingly, there is no optimality penalty associated with this reduction in communication cost. The optimal routing matrix solving (18) and its corresponding optimal utility are eventually achieved by [P1]-[P6].

V. Simulations

We present simulation results for the problem in (18). We consider a network with 120 user terminals randomly placed within a rectangle area of size $5000m \times 3500m$. The nodes collaborate to forward packets to 9 APs. We set all the requested rates to $\rho_{0j} = 0.2$. We form $\boldsymbol{R}$ using the empirical distribution in [1] and then generate the $\hat{R}_{ij}$ estimates uniformly distributed in $[(1-0.25)R_{ij}, (1+0.25)R_{ij}]$. This entails a 25% uncertainty in reliability estimates. The variance of the estimates is $\Sigma_{ij} = (0.5R_{ij}^2)/12$. The resulting estimated matrix $\hat{\boldsymbol{R}}$ is shown schematically in Fig. 1.

After running the robust routing protocol defined in proposition 3 for 200 iterations, we obtained the optimal routing matrix $\mathbf{T}$ shown in Fig. 2. We see that nodes split their traffic in different routes, ensuring less sensibility of network utility to estimation errors.

An alternative representation of the data flow throughout the network, is shown in Fig 3 where we plot the inbound traffic (green dots) and outbound traffic (red circles) of each node. We see that near each AP, there are many nodes with large inbound traffic. This again indicates that nodes farther away from the AP are splitting their traffic between many different routes. For any node, the difference between the red dot and the green circle depicts the data rate of that node.

To study the convergence rate of our algorithm, we define the

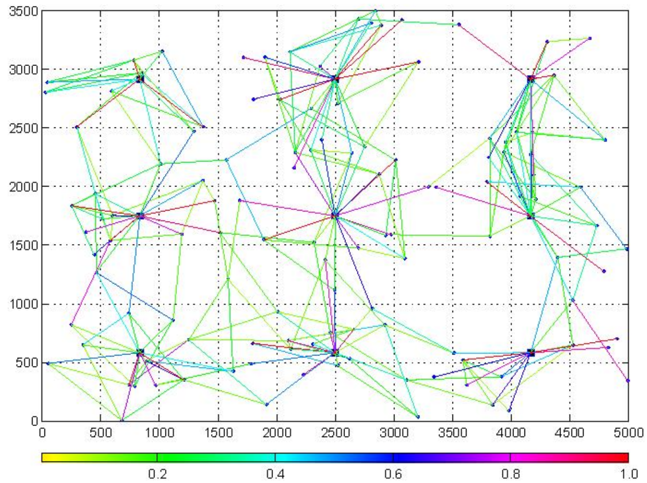

Fig. 2. Optimal robust routing matrix $\mathbf{T}$ obtained as a solution of (18). Nodes use different routes to reduce the variance of the estimated rates.

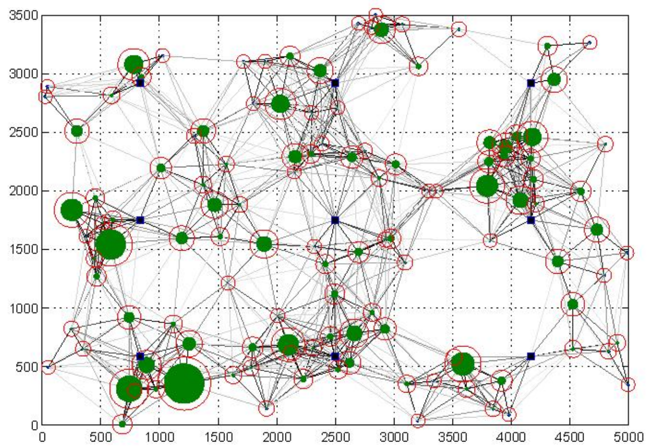

Fig. 3. The inbound traffic of each terminal is proportional to the size of the green dot. The outbound traffic is represented by the red circle ($\hat{\mathbf{R}}$ is shown in the background).

residual as

$$\epsilon(k) := \|\boldsymbol{\gamma}(k+1) - \boldsymbol{\gamma}(k)\|_2 \tag{35}$$

which is an indicator of the optimality of the dual iterates. The evolution of $\epsilon(k)$ is shown in Fig. 4. The algorithm converges slowly at the beginning, but once it gets close enough to the optimum the convergence rate is very fast.

VI. Conclusions

We introduced a robust approach to stochastic routing in wireless multi-hop networks. Since in any practical implementation link performance metrics have to be estimated we posed problems in which average rate utilities (given as functions of estimated performance metrics) are maximized. To ensure robustness against estimation errors we further constraint variance utilities (expressed as functions of the estimator's variances) to lie within a certain region.

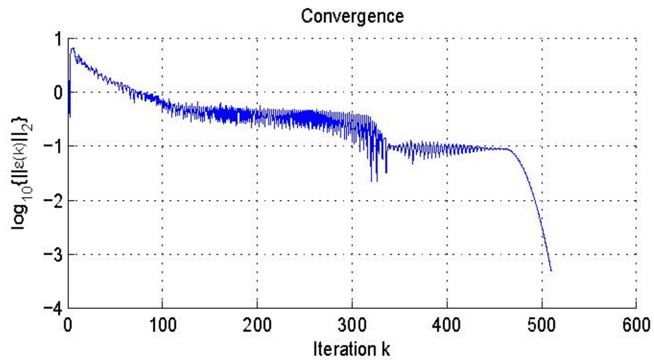

Fig. 4. Convergence rate of the robust routing protocol [P1]-[P6] of Proposition 3.

From a practical point of view, we showed that the optimization problems involved are convex, and thus can be solved efficiently using interior point methods. Furthermore, using dual decomposition and associated computational methods, we developed a distributed solution to the robust routing problem.

References

[1] D. Aguayo, J. Bicket, S. Biswas, G. Judd, and R. Morris, "Link-level measurements from an 802.11b mesh network," *ACM SIGCOMM Computer Commun. Review*, vol. 34, pp. 121–132, October 2004.

[2] D. Bertsekas, *Nonlinear programming*. Athena Scientific, 2nd ed., 1999.

[3] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[4] 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, pp. 104–116, January 2005.

[5] D. De Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proc. Int. ACM Conf. Mobile Computing, Networking*, pp. 134–146, San Diego, CA, September 14-19, 2003.

[6] B. Johansson, P. Soldati, and M. Johansson, "Mathematical decomposition techniques for distributed cross-layer optimization of data networks," *IEEE J. Sel. Areas Commun.*, vol. 24, pp. 1535–1547, August 2006.

[7] H. Lundgren, E. Nordstrom, and C. Tschudin, "The gray zone problem in IEEE 802.11b based ad hoc networks," *ACM SIGMOBILE Mobile Computing and Commun. Review*, vol. 3, pp. 104–105, July 2002.

[8] A. Ribeiro, Z.-Q. Luo, N. D. Sidiropoulos, and G. B. Giannakis, "A general optimization framework for stochastic routing in wireless multihop networks," *IEEE Trans. Signal Process.*, August 2006 (submitted). Available at http:// www.ece.umn.edu/ users/ aribeiro/ research/ pubs.html.

[9] A. Ribeiro, N. D. Sidiropoulos, and G. B. Giannakis, "Optimal distributed stochastic routing algorithms for wireless multihop networks," *IEEE Trans. Wireless Commun.*, October 2006 (submitted). Available at http:// www.ece.umn.edu/ users/ aribeiro/ research/ pubs.html.

[10] E. M. Royer and C.-K. Toh, "A review of current routing protocols for ad-hoc mobile wireless networks," *IEEE Pers. Commun.*, vol. 6, pp. 46–55, April 1999.