This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE INFOCOM 2007 proceedings.

Modelling and Optimization of Stochastic Routing for Wireless Multihop Networks

Alejandro Ribeiro, Georgios. B. Giannakis and Zhi-Quan Luo
Dept. of Electrical and Computer Engineering
University of Minnesota
200 Union St. SE, Minneapolis MN-55455
Email: {aribeiro, georgios, luozq}@ece.umn.edu

Nikolaos D. Sidiropoulos
Dept. of Electrical and Computer Engineering
Technical University of Crete,
Chania - Crete, Greece 731 00.
Email: nikos@telecom.tuc.gr

***Abstract*—We introduce a novel approach to multi-hop routing in wireless networks. Instead of the usual graph description we characterize the network by the packet delivery ratio matrix whose entries represent the probability that a given node decodes the packet transmitted by any other node. The model lends itself naturally to the formulation of stochastic routing protocols in which packets are randomly routed to neighboring nodes; and routing algorithms search for a matrix of routing probabilities according to properly defined optimality criteria. The goal of the paper is to show that this novel framework offers a useful model to aid in the design of optimal routing algorithms. In particular, it is established that: i) performance is improved with respect to graph descriptions; and ii) optimal routes can be obtained as the solution of optimization problems, many of which turn out to be convex and can thus be solved in polynomial time using interior point methods.**

Keywords: Routing, Wireless Networks, Markov chains, Convex Optimization, Linear programming

I. Introduction

Despite the fact that energy is a scarce resource in wireless networks it has not been until recently that it has jumped to the forefront of design requirements [10], [11]. While energy can be saved in different ways, multi-hopping is heralded as a promising alternative. Considering that received power decays exponentially with distance as $d^{-\alpha}$, with α between 3 and 4 – depending on the environment – the numbers are staggering. Splitting for instance a single hop in two co-linear hops saves about 10 dB in energy; and dividing a route in ten hops consumes in the order of a thousandth of the energy consumed by the original single hop. Even though this rough assessment ignores the extra energy cost due to collision and retransmission protocols and extra processing required at each node, it is not difficult to see that by reducing the average distance between communicating pairs of nodes, multi-hop routing secures significant power savings, when not the feasibility of the communication link itself.

As [10] correctly points out "we all have learned to draw a graph to depict a communication network" and not surprisingly routing algorithms for wireless networks have evolved from the accumulated knowledge about these graph models. But since a link in a wireless network does not entail a tangible connection, its definition can be somewhat arbitrary. Nonetheless, many useful multi-hop routing algorithms adhere to the so called "disk routing models" which typically proceed in three stages: i) define a communication radius for each node; ii) draw the corresponding connectivity graph; and iii) utilize network optimization tools, e.g., shortest path routing, to find the optimal route. Most of the differences in multi-hop routing algorithms arise in the definition of the associated link metrics. These include path reliability, transmitted power, and mutual interference to name a few; see e.g., [12], [22] and references therein.

The limitations of disk models are well documented, and by now it has become clear that the simplistic assumption of the disk model propagates through analysis to yield unrealistic conclusions, e.g., [13], [14], [24]. An implicit assumption in the disk model is that the reliability of a link is either 0 when no packet is communicated successfully or 1 when all packets are. However, measurements in experimental networks have shown that many links have intermediate loss rates [1], [17]. In order to overcome the inadequacy of disk models different approaches have been proposed as in, e.g., [6], [9]. Of particular importance to the present paper [7] advocated a fully connected graph model in which the weight of the arc connecting two nodes is inversely proportional to the success probability of their communication link.

An alternative approach is to reinterpret the graph as a matrix and consider multi-hop routing as an optimization problem based on the delivery-ratio (or pairwise packet-success-probability) matrix $\mathbf{R}$ whose (i, j)-th entry R_{ij} represents the probability that a packet transmitted by the j-th user U_j is correctly received by the i-th user U_i. This is what we introduce in this paper in the context of multi-hop routing for a multiple access wireless channel.

The inherent uncertainty captured by $\mathbf{R}$ lends itself to stochastic routing protocols (SRPs). User node U_j in an SRP transmits (i.e., routes) its packet through U_i with a certain probability T_{ij}. Instead of a routing table, a stochastic routing algorithm finds a matrix of probabilities $\mathbf{T}$ – with entries T_{ij} – according to some optimality criteria (Section II). We first consider a per-session model of routing in which a *single* node is interested in finding an optimal route to its intended destination given the matrix $\mathbf{R}$. We start showing that the random route can be described by the evolution of a suitably defined Markov chain allowing a characterization of deliverability in terms of known properties of absorbing Markov chains (Section II). Using this result we assess the merit of a routing protocol in terms of the convergence rate to the Markov chain's steady state distribution, which readily suggests routing protocols that maximize this rate (Section II-B). The latter approach optimizes the worst case scenario since it implicitly maximizes the packet delivery probability for a fixed (sufficiently large) delay. An alternative figure of merit is the average convergence rate. This leads to the minimization of the expected delay, a problem that as we will show, turns out to be

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. The work of Z.-Q. Luo is supported in part by the National Science Foundation, Grant No. DMS-0312416. The work of N. D. Sidiropoulos was partially supported by a bilateral cooperative research grant of the Greek Secretariat for Research and Technology.

equivalent to finding the shortest path in a fully connected graph with arc weights equal to $1/R_{ij}$ (Section II-C). This coincides with, and thus motivates, the heuristic algorithm originally proposed in [7].

We then look at joint session optimization by considering arrival rates at each user and designing routing protocols to maximize, in some sense, the (vector of) arrival rates. This setup calls for finding conditions for queue stability (Section III). Based on these conditions we define optimal routes as those maximizing: i) the arrival rate of the worst user (max-min routing); ii) the sum of the rates of all users (max-sum rate); and iii) the product of all users rates (max-product). We demonstrate that all these optimal routes can be found as solutions of convex optimization problems via the highly efficient interior point methods (Section III-B). As a byproduct, we also show that max-min routing with a bound on the *total* network traffic is also equivalent to finding the shortest path in a fully connected graph with arc weights equal to $1/R_{ij}$ (Section III-C).

Overall, the contribution of this paper is to establish that SRPs based on the delivery ratio matrix $\mathbf{R}$ hold great potential for multi-hop routing in wireless networks since they: i) offer a more accurate model of the wireless network thus achieving better performance than routing algorithms based on graph models; ii) in many cases optimal routes can be obtained as the solution of convex optimization problems that can be solved in polynomial time using interior point methods [4]; and iii) they subsume some existing routing algorithms based on graph models as particular cases.

II. Stochastic Routing Protocols (SRP) – Per session model

Consider a wireless network with $J+1$ user nodes $\{U_j\}_{j=1}^{J+1}$ in which the first J users $\{U_j\}_{j=1}^{J}$ participate in routing packets to the destination $D \equiv U_{J+1}$. The physical and medium access layers are such that if a packet is transmitted by U_j it is correctly *received* by U_i with probability R_{ij} that we arrange in the matrix $\mathbf{R}$. Note that in the presence of fading the probabilities R_{ij} are averaged over all fading states. We first consider a *per-session* model of routing in which a user node establishing a session is confronted with the routing decisions of its peers that determine the entries R_{ij} of $\mathbf{R}$. Supposing that the probabilities in $\mathbf{R}$ remain invariant over the duration of a session, our goal is to find a stochastic routing strategy that is optimal in a suitable sense. Note that this model is also applicable in a low traffic scenario, where at any time there is only one packet in the network.

Let $e_j(n)$ indicate the binary (0/1) event that the packet is at U_j at time n whose probability we denote by $p_j(n) := \Pr\{e_j(n) = 1\}$. Correspondingly, we define the vectors $\mathbf{e}(n) := [e_1(n), \ldots, e_{J+1}(n)]^T$ and $\mathbf{p}(n) := [p_1(n), \ldots, p_{J+1}(n)]^T$, where ^{T} denotes transposition. If the packet is generated at a known source U_s for some $s \in [1, J]$ we have that $p_s(0) = 1$. In general, the packets are generated at a random source with initial distribution $\mathbf{p}(0)$.

Routing is carried on according to a matrix $\mathbf{T}$ whose (i,j)-th entry T_{ij} is the probability that U_j decides to *transmit* (i.e., route) the packet to U_i. If U_j receives the packet at a certain time n, i.e., if $e_j(n) = 1$, U_j will select a random candidate destination from the set $\{U_i\}_{i=1}^{J+1}$ such that U_i is chosen with probability T_{ij}. If the transmitted packet is correctly decoded by U_i we have that $e_i(n+1) = 1$; otherwise, the packet is kept by U_j, i.e., $e_j(n+1) = 1$, and the random selection and transmission process is repeated. To describe the stochastic percolation of the packet throughout the network we define the matrix $\mathbf{K}$ with (i,j)-th entry $K_{ij} := \Pr\{e_i(n+1)|e_j(n)\}$ denoting the probability that the packet moves from U_j to U_i between times n and $n+1$. Note that $\mathbf{T}$ and $\mathbf{K}$ are related through $\mathbf{R}$. Indeed, for $i \neq j$ the packet moves from U_j to U_i if and only if it is routed through U_i *and* is correctly decoded; since these two events are independent we have

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

Because $\mathbf{K}$ and $\mathbf{T}$ are stochastic matrices, columns must sum up to 1 implying that $\mathbf{K}^T\mathbf{1} = \mathbf{1}$ and $\mathbf{T}^T\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes the all-one column vector. These two constraints and (1) imply that since $\mathbf{R}$ is prescribed by the physical layer, $\mathbf{K}$ is uniquely determined by $\mathbf{T}$ (but not vice versa).

Since the $(J+1)$-st user is the destination it will not route the packet, from which we infer that $T_{i(J+1)} = 0$, $\forall i \in [1, J]$; and after taking (1) into account we arrive at $K_{i(J+1)} = 0$, $\forall i \in [1, J]$. Arguing similarly, it follows that $R_{(J+1)(J+1)} = T_{(J+1)(J+1)} = K_{(J+1)(J+1)} = 1$. Summing up, with properly defined $\mathbf{k}_D \in \mathbb{R}^J$ and $\mathbf{K}_D \in \mathbb{R}^{J \times J}$ we can partition $\mathbf{K}$ as

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_D & \mathbf{0} \\ \mathbf{k}_D^T & 1 \end{pmatrix}_{(J+1)\times(J+1)}, \tag{2}$$

where $\mathbf{0}$ denotes the all-zero column vector. Let $\mathbf{c}_{J+1} := [0, \ldots, 0, 1]$ denote the $(J+1)$-st vector in the canonical basis of $\mathbb{R}^{J+1}$. It follows easily by direct substitution that (2) holds if and only if $\mathbf{K}\mathbf{c}_{J+1} = \mathbf{c}_{J+1}$, i.e., if and only if $\mathbf{c}_{J+1}$ is an eigenvector of $\mathbf{K}$ associated with the eigenvalue 1.

For future reference, we define the set of transmit probability matrices in $\mathbb{R}^{(J+1)^2}$ as

$$\mathcal{T} = \{\mathbf{T} \in \mathbb{R}^{(J+1)^2} \;:\; \mathbf{T}^T\mathbf{1} = \mathbf{1}, T_{ij} \geq 0, \forall i, j\}. \tag{3}$$

The constraints on $\mathbf{K}$ can be written as $\mathbf{K} \in \mathcal{K}$ with

$$\mathcal{K} = \{\mathbf{K} \in \mathcal{T} : K_{ij} = T_{ij}R_{ij}, \text{for } i \neq j, \mathbf{T} \in \mathcal{T}; \\ \mathbf{K}\mathbf{c}_{J+1} = \mathbf{c}_{J+1}\}. \tag{4}$$

Note that the set $\mathcal{K}$ is a convex polyhedron in $\mathbb{R}^{(J+1)^2}$.

We can characterize the evolution of $\mathbf{p}(n)$ in terms of $\mathbf{K}$. Indeed, note that due to the law of total probability $p_i(n) = \sum_{j=1}^{n} \Pr\{e_i(n)|e_j(n-1)\}p_j(n-1) = \sum_{j=1}^{n} K_{ij}p_j(n-1)$, that we can write in vector-matrix form as

$$\mathbf{p}(n) = \mathbf{K}\mathbf{p}(n-1) = \mathbf{K}^n\mathbf{p}(0). \tag{5}$$

That is, $\mathbf{p}(n)$ represents the probability evolution of a Markov chain characterized by $\mathbf{K}$ in which the j-th state represents the presence of the packet at user node U_j.

A. Deliverability

A basic requirement for the routing matrix $\mathbf{T}$ is to ensure that packets are eventually delivered to the destination $D \equiv U_{J+1}$, i.e.,

$$\lim_{n\to\infty} \mathbf{p}(n) = \mathbf{c}_{J+1}, \tag{6}$$

Since it is meaningful to focus on routing matrices that, at least, satisfy (6), we introduce the following definition.

Definition 1: A routing matrix $\mathbf{T}$ ensures deliverability if and only if (6) holds for any initial distribution $\mathbf{p}(0)$.

Building on (5), it is possible to find conditions to ensure deliverability of an SR matrix as we describe in the following theorem.

Theorem 1: The following statements are equivalent:

(i) The routing matrix $\mathbf{T}$ ensures deliverability.

(ii) Matrix $\mathbf{K}$ describes the probability evolution of an absorbing Markov chain whose unique absorbing state is $J+1$.

(iii) The spectral radius of $\mathbf{K}_D$ is strictly smaller than one, i.e., with $\mathrm{eig}(\mathbf{K}_D)$ denoting the set of eigenvalues of $\mathbf{K}_D$ we have $\rho(\mathbf{K}_D) := \max|\mathrm{eig}(\mathbf{K}_D)| < 1$.

(iv) The matrix $\mathbf{K}_D$ and the vector $\mathbf{k}_D$ in (2) satisfy $\mathbf{k}_D^T(\mathbf{I}-\mathbf{K}_D)^{-1} = \mathbf{1}^T$.

Proof: See [20]. ∎

Theorem 1 gives necessary and sufficient conditions for an SR matrix to have guaranteed deliverability. None of these conditions is difficult to achieve and, in general, simple routing algorithms, e.g., a random walk through the network with $T_{ij} = 1/J$, can ensure deliverability. A more interesting problem is how to obtain a matrix which guarantees that the limit in (6) is practically achieved with n as small as possible. This motivates different routing algorithms that we can obtain from (5) and analyze next.

B. Fastest convergence rate routing

The rate of convergence can be either measured on average or for the worst possible initial distribution $\mathbf{p}(0)$. These metrics lead to different criteria for optimal routing. Optimal routing in an average sense will be considered in Section II-C. What we expect from an optimal routing matrix $\mathbf{T}$ is for the convergence rate in (6) to be as fast as possible. The distance – in some sense – between $\mathbf{p}(n)$ and $\mathbf{c}_{J+1}$ can be measured by the p-norm $\|\mathbf{p}(n)-\mathbf{c}_{J+1}\|_p$ which is to be compared with the original distance $\|\mathbf{p}(0)-\mathbf{c}_{J+1}\|_p$ leading to the following expression for the worst-case convergence rate:

$$\xi_p = \sup_{\mathbf{p}(0)\neq\mathbf{c}_{J+1}} \lim_{n\to\infty} \left(\frac{\|\mathbf{p}(n)-\mathbf{c}_{J+1}\|_p}{\|\mathbf{p}(0)-\mathbf{c}_{J+1}\|_p}\right)^{1/n}. \quad (7)$$

This cannot be computed in closed-form for any p-norm. For $p=2$, corresponding to the Euclidean norm, the argument in (7) is maximized by the eigenvector associated with the second largest eigenvalue of $\mathbf{K}$. A meaningful routing algorithm is thus to look for the matrix $\mathbf{K}\in\mathcal{K}$ such that

$$\min_{\mathbf{K}\in\mathcal{K}} |\mathrm{eig}_2(\mathbf{K})| = \min_{\mathbf{K}\in\mathcal{K}} \max|\mathrm{eig}(\mathbf{K}_D)| = \min_{\mathbf{K}\in\mathcal{K}} \rho(\mathbf{K}_D), \quad (8)$$

where $\mathrm{eig}_2(\mathbf{K})$ denotes the second largest eigenvalue of $\mathbf{K}$ and $\mathrm{eig}(\mathbf{K}_D)$ the set of eigenvalues of $\mathbf{K}_D$. In establishing the first equality in (8) we used that all the eigenvalues of $\mathbf{K}_D$ are eigenvalues of $\mathbf{K}$ [cf. (2)]; in fact, $\mathrm{eig}(\mathbf{K}) = \mathrm{eig}(\mathbf{K}_D)\cup\{1\}$. The second equality follows from the definition of spectral radius.

Unfortunately, minimizing the spectral radius of a non-symmetric matrix is a notoriously difficult problem, intractable except for small-medium values of J [4]. This motivates an alternative measure of convergence rate based on the vector $\mathbf{p}_D(n) := [p_1(n),\ldots,p_J(n)]^T$ containing the probabilities that the packet is at a certain node other than the destination. The norm of $\mathbf{p}_D(n)$ measures the probability of the packet *not* being delivered at time n. This suggests the metric

$$\zeta_p = \max_{\mathbf{p}_D(n)} \frac{\|\mathbf{p}_D(n+1)\|_p}{\|\mathbf{p}_D(n)\|_p}, \quad (9)$$

which amounts to the worst-case one-step relative reduction of the vector $\mathbf{p}_D(n)$ which we want converging to zero [cf. (6)]. Similarly to ξ_p, we can define optimal routing in terms of minimizing ζ_p.

If we further recall that $\mathbf{p}_D(n+1) = \mathbf{K}_D\mathbf{p}_D(n)$, another class of optimal SRPs stemming from (9) can be designed to achieve

$$\min_{\mathbf{K}\in\mathcal{K}} \max_{\mathbf{p}_D(n)} \frac{\|\mathbf{K}_D\mathbf{p}_D(n)\|_p}{\|\mathbf{p}_D(n)\|_p} = \min_{\mathbf{K}\in\mathcal{K}} \|\mathbf{K}_D\|_p, \quad (10)$$

where the equality follows from the definition of the p-norm of a matrix. Different from (8), the optimization in (10) is a convex problem for all p since: i) due to the triangle inequality, norms are convex functions of their arguments; and ii) the set $\mathcal{K}$ is a convex polyhedron [cf. (4)]. For the usual norms, $p = 1, 2, \infty$, solving (10) is either a simple linear program (LP) for $p = 1, \infty$, or, a semi-definite program (SDP) for $p = 2$ [4].

In general, (8) and (10) are optimized by different matrices $\mathbf{T}$, and the pertinent comparisons are discussed in the following remark.

Remark 1: Requiring the solution of convex optimization problems – indeed, canonical optimization problems – (10) is tractable for networks with even a large number of users J; whereas (8) is only tractable for small-to-medium scale networks. On the other hand, (8) is more meaningful than (10), since the former compares the asymptotic behavior with the initial state while the latter compares two consecutive states. In practical protocol designs, (10) can be viewed as a tractable approximation to (8).

C. Minimum expected delay routing

An alternative approach to optimal routing is to consider the packet delivery time measured by the number of hops, and look for the matrix $\mathbf{T}$ that minimizes the average packet delay. Packet delay is simply the time n at which the packet is received by $D \equiv U_{J+1}$ and is given by:

$$\delta = \min\{n \;:\; e_{J+1}(n) = 1\} = \sum_{n=0}^{\infty}[1 - e_{J+1}(n)] \quad (11)$$

where the second equality is true since $1-e_{J+1}(n) = 1$ if $n < \delta$ and $1 - e_{J+1}(n) = 0$ for $n \geq \delta$; we thus have δ terms equal to 1 in the summation in (11). Starting from (11), the expected delay can be computed as we describe in the following theorem.

Theorem 2: For a routing matrix ensuring deliverability, the expected delay is given by

$$\bar{\delta} := \mathrm{E}(\delta) = \mathbf{1}^T(\mathbf{I}-\mathbf{K}_D)^{-1}\mathbf{p}_D(0), \quad (12)$$

where $\mathbf{p}_D(0) := [p_1(0),\ldots,p_J(0)]^T$ is the initial distribution for the first J users.

Proof: See [20]. ∎

The expected delay $\bar{\delta}$ is a function of the routing matrix $\mathbf{K}$ *and* the initial distribution $\mathbf{p}_D(0)$. Using the result in Theorem 2, we can find the matrix that minimizes the expected delay as the argument solving the optimization problem

$$\mathbf{K}^*[\mathbf{p}_D(0)] = \arg\min_{\mathbf{K}\in\mathcal{K}} \bar{\delta} = \arg\min_{\mathbf{K}\in\mathcal{K}} \mathbf{1}^T(\mathbf{I}-\mathbf{K}_D)^{-1}\mathbf{p}_D(0). \quad (13)$$

Conceptually, (13) appears difficult to solve. Interestingly, it turns out that (13) is equivalent to a shortest path routing algorithm as we establish in the ensuing theorem.

Theorem 3: Define the expected delay vector $\bar{\boldsymbol{\delta}} := [\bar{\delta}_1,\ldots,\bar{\delta}_J] := \mathbf{1}^T(\mathbf{I}-\mathbf{K}_D)^{-1}$ in which $\bar{\delta}_j$ is the expected delay when the packet starts at U_j, i.e., when $\mathbf{p}(0) = \mathbf{c}_j$; and let $\bar{\delta}_{J+1} = 0$. If there exists a matrix $\mathbf{K}$ ensuring deliverability, there exists a matrix $\mathbf{K}^\dagger \in \mathcal{K}$ such that

$$\bar{\delta}_j = \min_i \left\{\frac{1}{R_{ij}} + \bar{\delta}_i\right\}, \quad \bar{\delta}_{J+1} = 0, \quad (14)$$

which minimizes the expected delay for any initial distribution, i.e., $\mathbf{K}^*[\mathbf{p}_D(0)] = \mathbf{K}^\dagger$ for any $\mathbf{p}_D(0)$ and its corresponding $\mathbf{K}^*[\mathbf{p}_D(0)]$ as in (13).

Proof: See [20]. ∎

Characterizing the solution as in (14) indicates that $\mathbf{K}_D^*$ in (13) satisfies Bellman's principle of optimality which is known to characterize the shortest path route [2, Chap.5]. Thus, $\mathbf{K}_D^*$ in (13) can be found as the shortest path route (SPR) in a fully connected graph

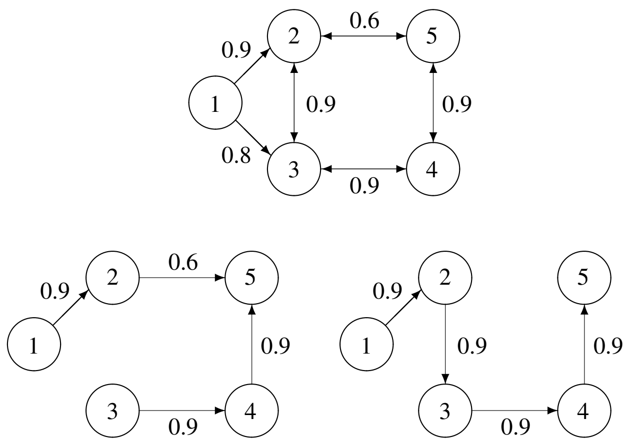

Fig. 1. For a simple connectivity graph (top, $\mathbf{R}$ matrix shown) the minimum expected delay routing algorithm in (13) tends to select short routes (left, minimum expected delay $\mathbf{K}$), while fastest convergence rate routing as per (8) selects longer routes with more reliable hops (right, fastest convergence rate $\mathbf{K}$).

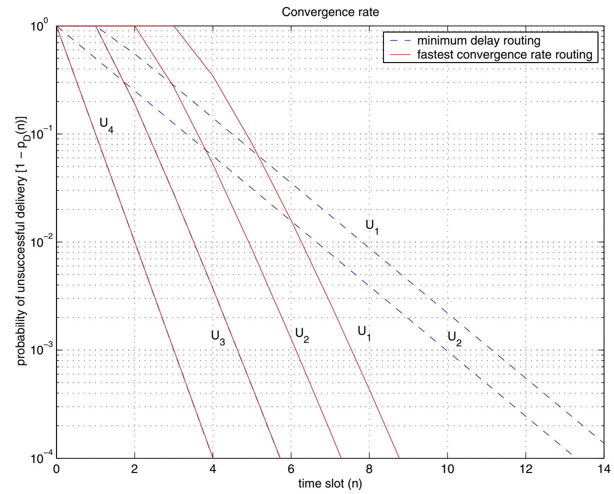

Fig. 2. Convergence rate for the network in Fig. 1. For a fixed time delay fastest convergence rate routing yields a smaller packet error probability.

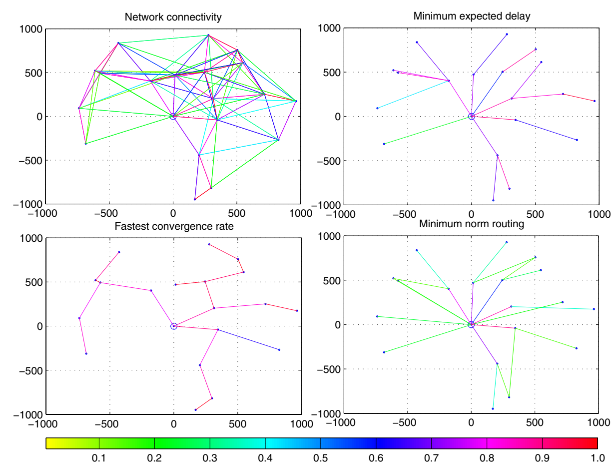

Fig. 3. A randomly generated network with 20 nodes, the color scale represents the elements of the matrix $\mathbf{K}$. Note how fastest convergence rate routing selects routes with large values of K_{ij}.

with the arc between U_i and U_j having weight $1/R_{ij}$. This implies that the solution to minimum expected delay routing can be found in $O(J^2)$ steps using dynamic programming tools, e.g., Bellman-Ford, Dijkstra, or Floyd-Warshall algorithms; see e.g., [2, Chap.5].

Also important, and contrary to what (13) suggested, minimum expected delay routing does not depend on the initial distribution. The average delays $\bar{\delta}[\mathbf{p}(0)]$ for different initial distributions $\mathbf{p}(0)$ are different, but there exists a matrix that minimizes $\bar{\delta}[\mathbf{p}(0)]$ for *all* $\mathbf{p}(0)$. Among other optimization problems, such a matrix is the solution of

$$\mathbf{K}^* = \arg\min_{\mathbf{K}\in\mathcal{K}} \mathbf{1}^T(\mathbf{I}-\mathbf{K}_D)^{-1}\mathbf{1} \tag{15}$$

obtained by making $\mathbf{p}_D(0) = \mathbf{1}/J$ in (13). Note that for a given $\mathbf{p}(0)$ there might exist alternative solutions to (13), but none will outperform $\mathbf{K}^*$ in (15). The matrix $\mathbf{K}^*$ in (15) can be obtained as the SPR in a fully connected graph with the arc between U_i and U_j having weight $1/R_{ij}$, a fact that we will later exploit in making pertinent comparisons between different routing algorithms.

D. Numerical examples and simulations

The fastest convergence rate SR algorithm in (8) maximizes the packet delivery probability for a given, sufficiently large, time index n. On the other hand, minimum expected delay routing as per (13) minimizes the expected time elapsed until packet delivery. The subtle differences between these two approaches are exemplified in Figs. 1 and 2.

The resulting routing matrices for minimum expected delay and fastest convergence rate routing are shown in Fig. 1. We can see that the former algorithm tends to select short routes sometimes containing unreliable hops (left) as verified by the link $U_2 \to U_5$ used to route U_1 and U_2's packets. Whereas, the latter uses longer routes but tends to use more reliable hops (right), as we can see from the use of the $U_2 \to U_3$ link to route U_1 and U_2's traffic. This is a manifestation of the different optimization criteria. The expected delay for routing U_2's packets is 1.67 for minimum expected delay routing and 3.33 for fastest convergence rate routing. The difference in convergence rate is shown in Fig. 2. To achieve a packet error probability of $1 - p_D(n) = 10^{-4}$, U_2's delay is 7.2 for fastest convergence rate routing and 13.1 for minimum expected delay routing.

Similar conclusions are reached for the more realistic example in Fig. 3 representing a randomly generated network with 20 nodes. In this figure, we depict the connectivity graph as well as the result of the minimum expected delay, fastest convergence rate, and minimum 2-norm SRP obtained from (10) with $p = 2$. Here it is also true that minimum expected delay prefers shorter routes, while fastest convergence rate prefers longer routes containing more reliable hops. Minimum 2-norm routing is the only algorithm considered that yields routing matrices implying non-deterministic routing, i.e., having $T_{ij} \neq 1, 0$ for some i, j.

For real time delay-sensitive applications, e.g., audio and/or video conferencing, fastest convergence routing is a better alternative. This is corroborated by Fig. 4 (top) showing the convergence rate for the network in Fig. 3. For a delay of 14 hops, fastest convergence rate routing yields a packet error probability of 10^{-4} for the least favored user; for the same delay, minimum expected delay routing achieves a packet error probability of 10^{-2}. For delay-tolerant applications, e.g., file transfers, the average delay metric is better suited since to deliver a large number of packets, the total number of required hops is significantly smaller – and consequently, the total energy required

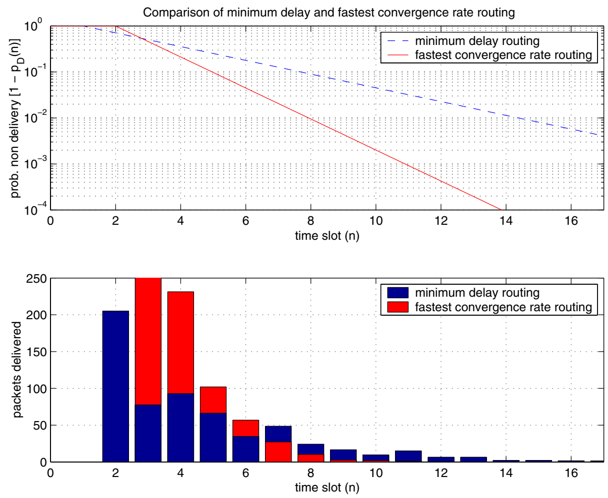

Fig. 4. Convergence rate of the least favored user for the network in Fig. 3 (top) and histogram of packet delivery times for a randomly chosen user (bottom). Fastest convergence rate routing is favored for time sensitive traffic.

for the session also is. This is illustrated in Fig. 4 (bottom) where we see that for minimum expected delay routing most packets are delivered in a few hops and a few packets take a long time to be delivered. For fastest convergence rate routing, none of the packets took more than 8 hops to be delivered but the total number of hops required to deliver all the packets was larger.

III. A Saturated System Approach

The approach in Section II ignores the effect of packet queuing at individual terminals. To incorporate this effect for heavily loaded networks, we consider that each user has an infinite-long queue. Packets arrive randomly at a rate $r_j \in (0, 1]$ packets per packet slot, to be delivered from terminal U_j to the destination U_{J+1}. The arrival process is assumed stationary. In any slot users with non-empty queues will transmit a packet at random with probability $\mu_j \leq 1$ as dictated by the medium access (MAC) sub-layer. As in Section II R_{ij} denotes the probability that U_i decodes U_j's packets. Because of possible interference $\mathbf{R} = \mathbf{R}(\boldsymbol{\mu})$ is a function of the transmission probabilities $\boldsymbol{\mu} := [\mu_1, \ldots, \mu_J]^T$ but we assume it given for the purposes of finding the optimal routes.

If a packet is indeed transmitted it is directed towards terminal U_i with probability T_{ij}. Our goal is to find conditions for the arrival rates r_j to yield stable queues and to design routing matrices $\mathbf{T}$ that maximize the sustainable r_j in some sense. Besides its own packets, U_j receives packets from other nodes for an *aggregate* arrival rate λ_j. Note that the departure rate $\lambda_j^{(o)}$ from U_j coincides with λ_j if the queue is stable and is smaller than λ_j if the queue is not stable. If, as in Section II, we let $K_{ij} = T_{ij}R_{ij}$ denote the probability that a packet moves from U_j to U_i between times n and $n+1$ we have that (see also Fig. 5)

$$\lambda_i = r_i + \sum_{j=1}^{J} K_{ij}\lambda_j^{(o)} \leq r_i + \sum_{j=1}^{J} K_{ij}\lambda_j \tag{16}$$

with equality achieved when all queues are stable. Notice that the sum in (16) includes the packets that fail to leave U_i in the term $K_{ii}\lambda_i$. Upon defining the vectors of (external) arrival rates $\mathbf{r} := [r_1, \ldots, r_J]^T$ and aggregate arrival rates $\boldsymbol{\lambda} := [\lambda_1, \ldots, \lambda_J]^T$, we

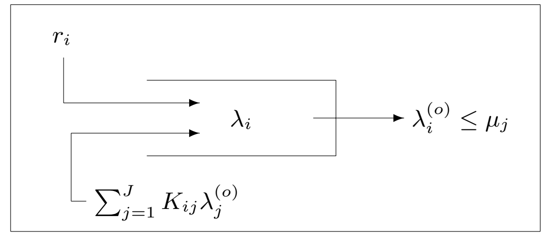

Fig. 5. Queue balance equations.

can express (16) in vector-matrix form as

$$\boldsymbol{\lambda} \preceq \mathbf{r} + \mathbf{K}_D\boldsymbol{\lambda}, \tag{17}$$

with $\mathbf{K}_D$ denoting the $J \times J$ upper left corner of $\mathbf{K}$ as in (2), and $\preceq$ denoting componentwise inequality. The first problem of interest is to find conditions under which pairs $(\mathbf{T}, \mathbf{r})$ of routing matrices and arrival rates are stable in the sense that the probability of any queue having more than an arbitrarily large number of packet is zero. Such a condition is given by the following theorem.

Theorem 4: Consider a wireless network with reliability matrix $\mathbf{R} = \mathbf{R}(\boldsymbol{\mu})$, and let $\mathbf{K}$ be the corresponding matrix whose entries K_{ij} denote the probability that a packet moves from U_j's to U_i's queue. $\mathbf{R}(\boldsymbol{\mu})$ and $\mathbf{K}$ are related as in (4). Assuming stationary arrival processes with strictly positive rates $\mathbf{r} \succ \mathbf{0}$ and stationary service processes with rates $\boldsymbol{\mu} \preceq \mathbf{1}$ we have that:

(i) for having a stable system, i.e., for all queues to be stable, it is sufficient to have
$$(\mathbf{I} - \mathbf{K}_D)^{-1}\mathbf{r} \prec \boldsymbol{\mu} \tag{18}$$

(ii) for having a stable system it is necessary to have invertible $(\mathbf{I} - \mathbf{K}_D)$ and
$$(\mathbf{I} - \mathbf{K}_D)^{-1}\mathbf{r} \preceq \boldsymbol{\mu}. \tag{19}$$

Proof: See [20]. ■

Theorem 4 provides a condition for having stable queues and in that sense it is the counterpart of (6). Given a routing matrix $\mathbf{T}$ and a vector of arrival rates $\mathbf{r}$, (18) and (19) can be used to check stability. For any candidate routing matrix $\mathbf{T}$, we can define the stability region $\mathcal{S}$ of arrival rate vectors leading to stable queues as

$$\mathcal{S} = \{\mathbf{r} \in \mathbb{R}^J : \mathbf{r} = (\mathbf{I} - \mathbf{K}_D)\boldsymbol{\lambda},\ \text{with } \mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}\}. \tag{20}$$

The interpretation of $\mathcal{S}$ in (20) is that rates strictly inside $\mathcal{S}$ lead to stable queues, while points strictly outside $\mathcal{S}$ lead to unstable queues. Rates in the boundary of $\mathcal{S}$ may or may not be stable.

A. Physical/medium-access/network layer interaction

Because the framework so far as well as the SRPs of the ensuing section rely on knowledge of $\mathbf{R}$, we delineate here how this matrix is determined depending on the access scheme (orthogonal or non-orthogonal) used at the physical layer.

If terminals transmit over orthogonal channels as when frequency (F-), time (T-), or code (C-) division multiple access (DMA) is utilized at the physical layer, $\mathbf{R}$ clearly depends on the power transmitted by individual users. Furthermore, lack of contention implies that $\mathbf{R}$ does not depend on the transmission probabilities $\boldsymbol{\mu}$. Let P_j denote the power transmitted by U_j, $\mathbf{x}_j$ its position and $L(\mathbf{x}_j - \mathbf{x}_i)$ a distance-dependent path loss coefficient – e.g., an exponential path loss law for which $L(d) = d^{-\alpha}$. The signal to noise ratio (SNR) for the $U_j \rightarrow U_i$ link is thus

$$\text{SNR}_{ij} = \frac{P_j L(\mathbf{x}_j - \mathbf{x}_i)}{N_0} \tag{21}$$

For the SNR in (21) and a given modulation and error control code pair one can readily obtain a certain packet success probability $R_{ij}(\mathrm{SNR}_{ij}) = R_{ij}(P_j)$ for the link $U_j \to U_i$. Depending on how fast fading varies with respect to packet lengths channels are classified as fast, slow, or block fading. If fading is invariant over several packet transmissions, $R_{ij}(P_j)$ is given by the instantaneous packet success probabilities for the given fading state. If fading is fast, so that any packet experiences a sufficiently large number of independent channel realizations, the receiver can collect the available time diversity and $R_{ij}(P_j)$ can be approximately obtained from the error probability for additive white Gaussian noise channels. In a block fading model, the channel changes from packet to packet, and the transmitter is confronted with an unknown fading state. In this case $R_{ij}(P_j)$ can be computed from the average of the instantaneous error probabilities over all fading states. In all three cases, $R_{ij}(P_j)$ is expressible as a function of P_j.

For contention- or interference-limited networks as is respectively the case for random access and CDMA with pseudo-noise spreading sequences $\mathbf{R}$ and $\boldsymbol{\mu}$ are coupled in the sense that $\mathbf{R}$ is a function of the transmitted powers $\mathbf{p} := [P_1, \ldots, P_J]^T$ and the transmission probabilities $\boldsymbol{\mu}$. A usual practice to further simplify the problem, is to assume that for the purposes of accounting for interference $\boldsymbol{\mu} = \mathbf{1}$, which implies that the signal to noise plus interference ratio SINR is given by

$$\mathrm{SINR}_{ij} = \frac{P_j L(\mathbf{x}_j - \mathbf{x}_i)}{N_0 + \gamma \sum_{k \neq i,j}^{J} P_k L(\mathbf{x}_k - \mathbf{x}_i)} \tag{22}$$

The SINR in (22) can be mapped to a packet success probability $R_{ij}(\mathrm{SINR}_{ij})$. This eliminates $\boldsymbol{\mu}$, as a variable determining $\mathbf{R}(\mathbf{p}, \mathbf{1})$ that now depends only on $\mathbf{p}$. The approximation can be justified by noting that any rate $\boldsymbol{\rho}$ achievable in a network with reliability matrix $\mathbf{R}(\mathbf{p}, \boldsymbol{\mu})$ is also achievable in a network with reliability $\mathbf{R}(\mathbf{p}, \mathbf{1})$ and in that sense the latter represents an upper bound on the stability region of the former.

The important observation here is that regardless of the physical and medium access control (sub-) physical layers from a networking layer perspective the matrix $\mathbf{R}$ can be measured by counting acknowledgements of correctly decoded packets. Thus, a network layer protocol can be envisioned to select the routing matrix $\mathbf{T}$ that maximizes $\mathbf{r}$ in some sense for a given matrix $\mathbf{R}$. We pursue this problem in the next section.

Remark 2: The stability condition in Theorem 4 was derived for a reliability matrix $\mathbf{R} := \mathbf{R}(\boldsymbol{\mu})$. The reader familiar with contention limited networks, e.g., Aloha, may realize that this is not a completely accurate model. In fact, $\mathbf{R} = \mathbf{R}(\boldsymbol{\nu})$ is a function of the transmission probabilities $\boldsymbol{\nu} := [\nu_1, \ldots, \nu_J]^T$ with ν_j denoting the probability that U_j transmits a packet in a given slot. The vector $\boldsymbol{\nu}$ is not necessarily equal to $\boldsymbol{\mu}$ due to the probability of having empty queues but in general $\boldsymbol{\nu} \preceq \boldsymbol{\mu}$. Under this model finding the stability region becomes intractable for the same reasons it is intractable to find Aloha's stability region [19]. A dominant system approach as in e.g., [8] can be used to prove that $\mathcal{S}$ in (20) is an achievable stability region as we show in [20].

B. Maximum arrival rate routing

With $\mathbf{R}$ available at a central location, we look for routing matrices $\mathbf{T} \in \mathcal{T}$ for which we have $\mathbf{K} \in \mathcal{K}$ [cf. (3) and (4)] that maximize a measure of the arrival rate vector $\mathbf{r}$. Different optimization criteria can be devised to obtain routing algorithms maximizing the arrival rate vector. A first approach is to maximize a weighted sum of rates $\sum_{j=1}^{J} r_j = \mathbf{w}^T \mathbf{r}$ with $\mathbf{w} \succeq \mathbf{0}$. The sum-rate optimal matrix can be obtained as the solution of the optimization problem

$$\mathbf{K}^* = \arg \max_{\mathbf{K} \in \mathcal{K}, \mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}} \mathbf{w}^T \mathbf{r} = \max_{\mathbf{K} \in \mathcal{K}, \mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}} \mathbf{w}^T (\mathbf{I} - \mathbf{K}_D) \boldsymbol{\lambda}. \tag{23}$$

A concern with the formulation in (23) is that it tends to favor terminals close to the destination. An alternative approach is to maximize $\min_j r_j$, the rate of the least favored user. We refer to this as max-min optimal routing; the corresponding routing matrix can be obtained as the solution to

$$\mathbf{K}^* = \arg \max_{\mathbf{K} \in \mathcal{K}, \mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}} \min_j r_j = \arg \max_{\mathbf{K} \in \mathcal{K}, \mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}} \min_j \left[(\mathbf{I} - \mathbf{K}_D) \boldsymbol{\lambda}\right]_j. \tag{24}$$

The optimization problems in (23) and (24) are bilinear in $\mathbf{K}_D$ and $\boldsymbol{\lambda}$, and as such, notoriously difficult to solve in general. Enticingly, we can capitalize on the structure of the problem to reduce them to simple linear programs. The main result allowing this reduction is stated in the following theorem.

Theorem 5: Consider a maximization problem of the form

$$v^* := \max_{\mathbf{K} \in \mathcal{K}, \mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}} g[(\mathbf{I} - \mathbf{K}_D) \boldsymbol{\lambda}], \tag{25}$$

where $g : \mathbb{R}^J \to \mathbb{R}$ is a function monotonically non-decreasing in each component, i.e., for vectors $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}$ with $v_j^{(1)} \leq v_j^{(2)}$ and $v_i^{(1)} = v_i^{(2)}$ for $i \neq j$, we have that $g[\mathbf{v}^{(1)}] \leq g[\mathbf{v}^{(2)}]$. Then, there exists a matrix $\mathbf{K} \in \mathcal{K}$ such that

$$v^* = \max_{\mathbf{K} \in \mathcal{K}} g[(\mathbf{I} - \mathbf{K}_D) \boldsymbol{\mu}]. \tag{26}$$

Proof: See [20]. ■

Theorem 5 establishes that routing algorithms involving component-wise non-decreasing objective functions can be solved by setting $\boldsymbol{\lambda} = \boldsymbol{\mu}$ in the argument function to be optimized. Clearly, this is the case for max-min rate optimal and sum-rate optimal routing in which the functions are $g(\mathbf{v}) = \min_i(v_i)$ and $g(\mathbf{v}) = \mathbf{1}^T \mathbf{v}$, respectively. Furthermore, with $\boldsymbol{\lambda} = \boldsymbol{\mu}$, the bilinear arguments in (27) and (28) become linear functions of $\mathbf{K}_D$ implying the following corollary.

Corollary 1: Max-min optimal routing and sum-rate optimal routing can be obtained as solutions of linear programs (LP) in $\mathbf{K}$:

(i) For max-min optimal routing

$$\mathbf{K}^* = \arg \max_{\mathbf{K} \in \mathcal{K}} \min_i \left[(\mathbf{I} - \mathbf{K}_D) \boldsymbol{\mu}\right]_i. \tag{27}$$

(ii) For sum-rate optimal routing

$$\mathbf{K}^* = \arg \max_{\mathbf{K} \in \mathcal{K}} \mathbf{w}^T (\mathbf{I} - \mathbf{K}_D) \boldsymbol{\mu}. \tag{28}$$

Proof: The functions $g(\mathbf{v}) = \min_i(v_i)$ and $g(\mathbf{v}) = \mathbf{1}^T \mathbf{v}$ are component-wise non-decreasing in the sense considered in Theorem 5. This proves the equivalence of (27) with (24) and (28) with (23), respectively. That (28) is an LP follows after noting that the argument to be maximized is linear and recalling that the set $\mathcal{K}$ is a convex polyhedron. To prove that (27) is an LP introduce the auxiliary variable $t \geq \left[(\mathbf{I} - \mathbf{K}_D) \boldsymbol{\mu}\right]_i$, for all i and rewrite the maximization as

$$\begin{aligned} &\max \quad t \\ &\text{s.t.} \quad \mathbf{K} \in \mathcal{K}, \quad t\mathbf{1} \leq (\mathbf{I} - \mathbf{K}_D) \boldsymbol{\mu}. \end{aligned} \tag{29}$$

In (29), the argument and the constraints are linear entailing, by definition, an LP. ■

Corollary 1 demonstrates that sum-rate and max-min optimal routing can be efficiently solved by convex optimization techniques, e.g., interior point methods [4]. Solving an LP incurs roughly cubic complexity – $O(J^{3.5})$ to be precise – in the number of nodes J and in that sense it is only moderately more complex than finding a traditional shortest path route whose complexity is quadratic in J.

The applicability of Theorem 5 is fairly broad, implying that we can propose different routing algorithms and expect them to yield tractable optimization problems. As an example, consider maximum product-rate for which the optimal routing matrix is obtained as

$$\begin{aligned}\mathbf{K}^* &= \arg\max_{\mathbf{K}\in\mathcal{K},\mathbf{0}\preceq\boldsymbol{\lambda}\preceq\boldsymbol{\mu}} \prod_{j=1}^{J} r_j \\ &= \arg\max_{\mathbf{K}\in\mathcal{K},\mathbf{0}\preceq\boldsymbol{\lambda}\preceq\boldsymbol{\mu}} \sum_{j=1}^{J} \log\left[(\mathbf{I}-\mathbf{K}_D)\boldsymbol{\lambda}\right]_j, \end{aligned} \tag{30}$$

with the second equality following because the logarithm function is monotonically increasing. The max-product rate in (30) constitutes a more fair alternative to the max-sum rate in (23) since it prevents solutions in which some users receive a very small packet delivery rate. Notice that the argument in (30) is componentwise non-decreasing, and we thus can use Theorem 5 to obtain

$$\begin{aligned}\mathbf{K}^* &= \arg\max_{\mathbf{K}\in\mathcal{K},\mathbf{0}\preceq\boldsymbol{\lambda}\preceq\boldsymbol{\mu}} \prod_{j=1}^{J} r_j \\ &= \arg\max_{\mathbf{K}\in\mathcal{K},\mathbf{0}\preceq\boldsymbol{\lambda}\preceq\boldsymbol{\mu}} \sum_{j=1}^{J} \log\left[(\mathbf{I}-\mathbf{K}_D)\boldsymbol{\mu}\right]_j, \end{aligned} \tag{31}$$

which, since the logarithm is a concave function, is a convex optimization problem in $\mathbf{K}_D$.

C. An overall constraint in the total traffic

Imposing individual traffic constraints, the requirement $\mathbf{0} \preceq \boldsymbol{\lambda} \preceq \boldsymbol{\mu}$ does not impose an overall traffic constraint, something that is sometimes reasonable and sometimes not. In certain cases we may want to limit the total traffic in the network, e.g., to leave room for critical traffic, or, to ensure a fixed power consumption per time unit. In any event, the total traffic constraint can be written as $\boldsymbol{\lambda}^T\mathbf{1} = \mu_0 \leq \min_{j\in[1,J]} \mu_j$. In this context, we can consider different optimization criteria as in Section III-B yielding routing algorithms of the form

$$\begin{aligned}\mathbf{K}^* = \arg\max_{\mathbf{K},\boldsymbol{\lambda}} \;& g\left[(\mathbf{I}-\mathbf{K}_D)\boldsymbol{\lambda}\right] \\ \text{s.t.} \;& \mathbf{K}\in\mathcal{K},\ \mathbf{0}\preceq\boldsymbol{\lambda},\ \boldsymbol{\lambda}^T\mathbf{1}=\mu_0. \end{aligned} \tag{32}$$

The added constraint $\boldsymbol{\lambda}^T\mathbf{1} = \mu_0$ prevents application of Theorem 5 and, in general, problems of the form (32) will be difficult to solve. However, for the specific case of max-min optimal routing with an overall traffic constraint, i.e., $g(\mathbf{v}) = \min_i(v_i)$ in (32), we can establish a quite surprising connection with shortest path routing.

To study this connection note that since the constraints in $\boldsymbol{\lambda}$ and $\mathbf{K}_D$ are decoupled we can solve the optimization in two separate steps

$$\mathbf{K}^* = \arg\max_{\mathbf{K}\in\mathcal{K}} \left\{ \begin{array}{ll} \max_{\boldsymbol{\lambda}} & \min_i\left[(\mathbf{I}-\mathbf{K}_D)\boldsymbol{\lambda}\right]_i \\ \text{s.t.} & \mathbf{0}\preceq\boldsymbol{\lambda},\ \boldsymbol{\lambda}^T\mathbf{1}=\mu_0 \end{array} \right\}. \tag{33}$$

If $\mathbf{K}_D$ is fixed, then the innermost optimization is a simple linear max-min problem widely studied in a variety of contexts, e.g., game theory. The important point here is that the solution to this problem is well known, and in some cases computable in closed-form. This allows us to obtain the following theorem.

Theorem 6: For consistent routing matrices, max-min optimal routing with a global traffic constraint as defined by (33) is equivalent to

$$\mathbf{K}^* = \arg\min_{\mathbf{K}\in\mathcal{K}} \mathbf{1}^T(\mathbf{I}-\mathbf{K}_D)^{-1}\mathbf{1}. \tag{34}$$

Moreover, the optimal traffic vector is given by

$$\boldsymbol{\lambda}^* = \frac{(\mathbf{I}-\mathbf{K}_D^*)^{-1}\mathbf{1}}{\mathbf{1}^T(\mathbf{I}-\mathbf{K}_D^*)^{-1}\mathbf{1}}\mu_0. \tag{35}$$

Proof: See [20]. ■

Even though Theorem 6 transforms the problem in (33) into a conceptually simpler form, it is not yet clear how (34) might be solved. However, recalling (15) we see that quite surprisingly, max-min rate routing with a global traffic constraint as per (33) is equivalent to minimum delay routing as defined in (13). Since the solution to the latter, as we have already seen, is given by the shortest path in a fully connected graph with arc weights $1/R_{ij}$, so is the solution to (33); a fact that we summarize in the following corollary.

Corollary 2: The matrix $\mathbf{K}^\dagger \in \mathcal{K}$ satisfying Bellman's principle of optimality in (14) solves the max-min routing problem with a global traffic constraint defined by (33).

Proof: If $\mathbf{K}^\dagger \in \mathcal{K}$ satisfies (14), it solves (15) [cf. Theorem 3]. But (15) is identical to (34) which we know solves (33) [cf. Theorem 6]. Thus, if $\mathbf{K}^\dagger \in \mathcal{K}$ satisfies (14), it solves (33). ■

Corollary 2 implies that in order to find the matrix optimizing (33) it suffices to find the SPR in a fully connected graph with the arc between U_i and U_j having weight $1/R_{ij}$. On the other hand, Theorem 6 provides interesting insights on the optimal solution that we discuss in the following remarks.

Remark 3: The proof establishes that for any $\mathbf{K}_D$, the optimal $\boldsymbol{\lambda} = \boldsymbol{\lambda}^\dagger$ is given by (35). The corresponding rate offered to each user is subsequently given by $r_j = 1/\mathbf{1}^T(\mathbf{I}-\mathbf{K}_D)^{-1}\mathbf{1}$ showing that every user gets the same rate. The vector of optimal offered rates is

$$\mathbf{r}^* = \mu_0\left[\mathbf{1}^T(\mathbf{I}-\mathbf{K}_D^*)^{-1}\mathbf{1}\right]^{-1}\mathbf{1}. \tag{36}$$

Eq. (36) reveals that max-min routing is fair in the sense that it evenly divides the traffic resources available.

Remark 4: Strong duality applied to the innermost optimization over $\boldsymbol{\lambda}$ in (33) proves the equivalence of the latter with

$$\mathbf{K}^* = \arg\max_{\mathbf{K}\in\mathcal{K}} \left\{ \begin{array}{ll} \min_{\mathbf{x}} & \max_i\left[(\mathbf{I}-\mathbf{K}_D^T)\mathbf{x}\right]_i \\ \text{s.t.} & \mathbf{0}\preceq\mathbf{x},\ \mathbf{x}^T\mathbf{1}=\mu_0 \end{array} \right\} \tag{37}$$

where we used that $(\mathbf{I}-\mathbf{K}_D)^T = \mathbf{I}-\mathbf{K}_D^T$; see also [20]. The formulation in (37) corresponds to min-max optimal routes for a multihop cooperative downlink subject to a constraint in the total traffic delivered by $D \equiv U_{J+1}$. The interpretation is that of a group of terminals competing to receive information from $D \equiv U_{J+1}$ that can transmit at a rate of μ_0 packets per packet slot [cf. $\mathbf{x}^T\mathbf{1} \leq \mu_0$]. The access point (D) is interested in a fair formulation that minimizes the rate of the greediest user node while still using its own resources to a prescribed extent [cf. $\mathbf{x}^T\mathbf{1} = \mu_0$]. This problem turns out equivalent to max-min optimal routing for a multihop cooperative uplink [cf. (33) and (37)]. In particular, we deduce that every node is served with the same rate given by (36).

D. Simulations and numerical examples

Let us illustrate the distinctions between the different routing protocols. In Fig. 6 we represent the connectivity matrix $\mathbf{R}$ for a network with 50 nodes randomly deployed in a circle of radius 2 km. The results of sum-rate optimal routing as defined in (23) and

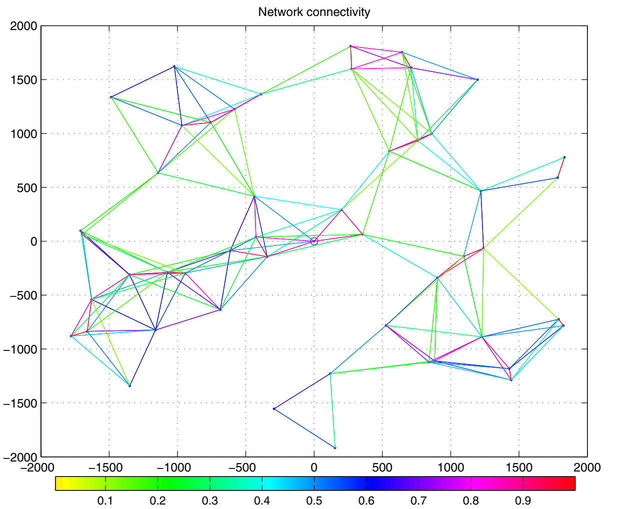

Fig. 6. Connectivity graph for a network with 50 nodes uniformly deployed in a circle with radius 2 Km. at whose center is the common AP. The color index represents the value of R_{ij} that is generated according to the empirical distribution in [7].

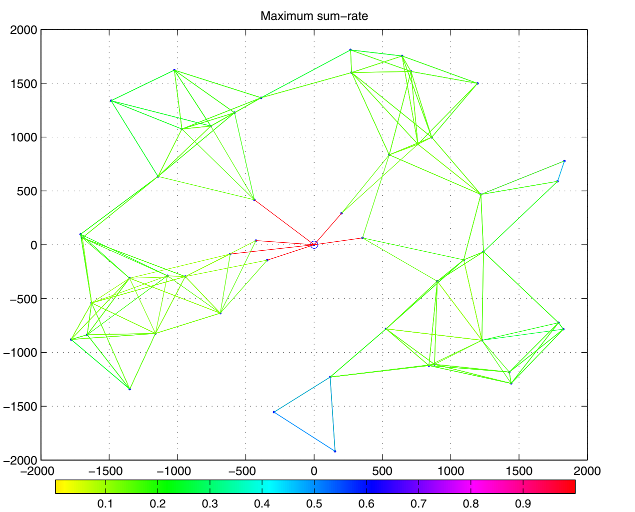

Fig. 7. Sum-rate optimal routes as given by (23). The nodes with good connections to the destination get most of the total rate available.

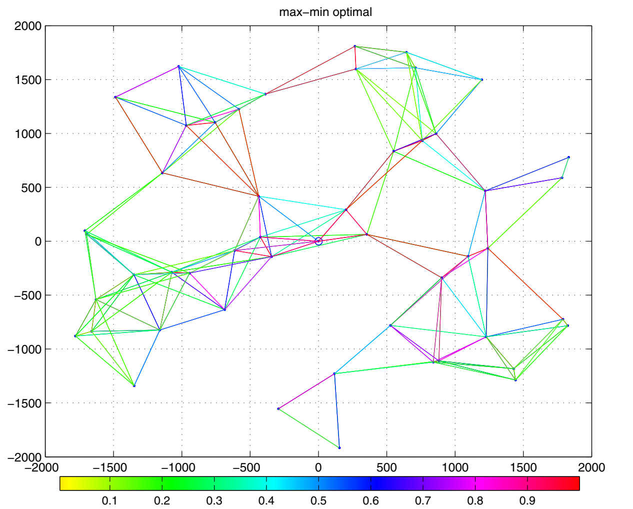

Fig. 8. Max-Min routes obtained as the solution of (24) for the network in Fig.6. Most nodes divide their traffic between many different neighbors to avoid the formation of bottlenecks.

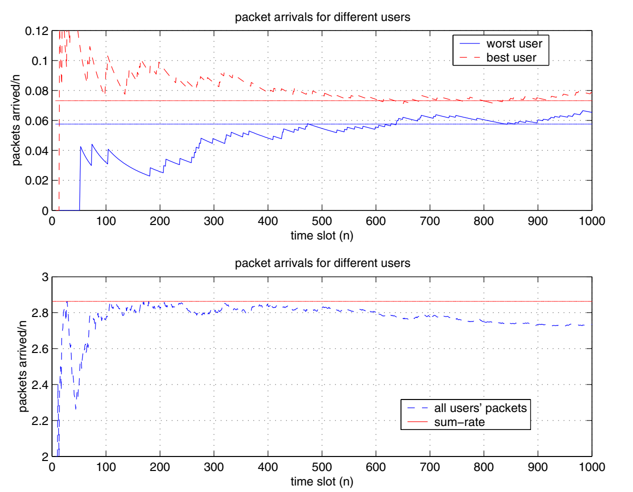

Fig. 9. Instances of the arrival rate processes for the max-min optimal routes in Fig. 8. The fairness of the protocol is manifested in the not so different rates offered to the best and worst nodes.

max-min optimal routing as per (24) are shown in Figs. 7 and 8, respectively. Fig. 10 depicts the shortest path routes solving max-min optimal routing with a global traffic constraint as given by (32). We recall that all these optimal routing algorithms can be efficiently solved as guaranteed by Corollaries 1 and 2.

Sum-rate optimal routing yields a matrix $\mathbf{K}$ in which the nodes with reliable links to the destination get most of the rate. Actually, a possible solution maximizing the sum-rate is for all the U_j's with $R_{(J+1)j} \neq 0$ to send their traffic to the destination $D \equiv U_{J+1}$ without forwarding any traffic belonging to other users. To this end we can add the constraint $\mathbf{r} \geq r_0\mathbf{1}$, which ensures that every user has a guaranteed rate $r_j = r_0$ with the excess traffic assigned to the most favored users. The result of this approach is shown in Fig. 7 ($r_0 = 0.01$). Note that the constraint $\mathbf{r} = (\mathbf{I}-\mathbf{K}_D)\boldsymbol{\mu} \geq r_0\mathbf{1}$ is linear in $\mathbf{K}_D$ maintaining the convexity of the problem in (28).

A perhaps better approach is max-min routing whose corresponding optimal routes are depicted in Fig. 8. We see that most users divide their traffic between many different neighbors to avoid the formation of bottlenecks. The fairness of this approach is illustrated in Fig. 9 where we show instances of the arrival processes of the best and worst users. For the network considered, the offered rates were 0.056 and 0.074, respectively. We see that the simulated arrival processes are accurately modelled by (16). As time progresses the number of packets delivered for terminal U_j approaches its rate r_j (times the elapsed time). We also plot the sum-rate for this case which is to be compared with 5.36 achieved by sum-rate optimal routing.

We finally depict in Fig. 10 the max-min optimal routes when we enforce a global traffic constraint set to $\boldsymbol{\lambda}^T\mathbf{1} = 1$. We note that among all the approaches considered in this subsection, this is the only one resulting in deterministic routes.

IV. Conclusions

We introduced a general framework for stochastic routing in wireless multi-hop networks. Deviating from the traditional graph models, we considered a general framework based on the packet delivery

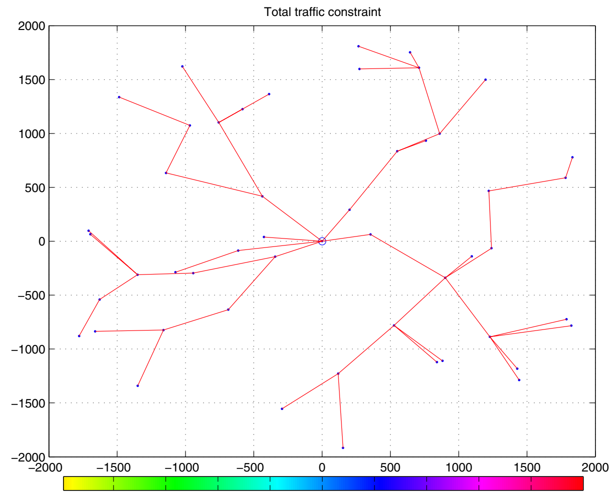

Fig. 10. Max-min routing with a global traffic limit as per (32). The routes are obtained as the shortest path in a fully connected graph with arc weights equal to $1/R_{ij}$.

probability matrix and showed that different routing algorithms can be either described by the evolution of a properly defined Markov chain – per session model of routing –, or by a network of backlogged queues – saturated system. These connections permit characterization of properly defined deliverability and stability conditions in terms of the spectral radius of a stochastic routing matrix.

For the per-session model of routing we introduced stochastic routing algorithms that maximize the convergence rate of the Markov chain, entailing a maximization of the packet delivery probability for a fixed, sufficiently large delay n. This routing approach is meaningful in the context of delay sensitive traffic involved in, e.g, voice and/or video conferencing. We further found an expression for the average packet delay measured by the number of hops and identified the corresponding optimal routing scheme that minimizes it. Interestingly, we proved that the optimum routing matrix in this case can be obtained as the shortest path route in a fully connected graph with the arc between users having a weight inversely proportional to the corresponding delivery ratio.

For the saturated system we defined different routing algorithms corresponding to different maximization criteria of the arrival rate vector. These approaches include maximization of the rate of the least favored user (max-min), maximization of the sum of rates (max-sum) and of the product of rates (max-prod). We showed that all these problems can be efficiently solved using convex optimization techniques. Rather unexpectedly, we also established equivalence between max-min routing with a global traffic constraint and minimum average delay routing.

In future research we plan to address additional topics related with the optimization framework introduced here. One direction is to consider different optimality criteria that were omitted due to space considerations. These include lower and upper bounds in user rates, robust formulations in the presence of fading, and different network topologies. Another direction is to develop algorithms for distributed implementations of optimal routing. These can be obtained along the lines of primal and dual decomposition for network optimization as in, e.g., [5], [16], [23]. A distributed implementation of algorithms proposed here has been reported in [21]. We will also work on generalizing the approach in Sections III and IV to ad-hoc networks.

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, Oct. 2004.

[2] D. Bertsekas and R. Gallager, *Data networks*. Prentice-Hall, Inc., 2nd ed., 1992.

[3] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia, “Routing with guaranteed delivery in ad hoc wireless networks,” *ACM Wireless Networks*, vol. 7, pp. 609–616, Nov. 2001.

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

[5] 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.

[6] K.-W. Chin, J. Judge, A. Williams, and R. Kermode, “Implementation experience with MANET routing protocols,” *ACM SIGCOMM Computer Commun. Review*, vol. 32, pp. 49–59, November 2002.

[7] 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.

[8] G. Dimic, N. D. Sidiropoulos, and R. Zhang, “Signal processing and queuing tools for MAC-PHY cross-layer design,” *IEEE Signal Process. Mag.*, vol. 21, pp. 40–50, September 2004.

[9] R. Dube, C. D. Rais, K.-Y. Wang, and S. K. Tripathi, “Signal stability-based adaptive routing (SSA) for ad hoc mobile networks,” *IEEE Pers. Commun.*, vol. 4, pp. 36–45, February 1997.

[10] A. Ephremides, “Energy concerns in wireless networks,” *IEEE Wireless Commun.*, vol. 9, pp. 48–59, August 2002.

[11] A. J. Goldsmith and S. B. Wicker, “Design challenges for energy-constrained ad hoc wireless networks,” *IEEE Wireless Commun.*, vol. 9, pp. 8–27, August 2002.

[12] P. Gupta and P. Kumar, “The capacity of wireless networks,” *IEEE Trans. Inf. Theory*, vol. 46, pp. 388–404, March 2002.

[13] M. Haenggi, “On routing in random Rayleigh fading networks,” *IEEE Trans. Wireless Commun.*, vol. 4, pp. 1553–1562, July 2005.

[14] M. Haenggi and D. Puccinelli, “Routing in ad hoc networks: a case for long hops,” *IEEE Commun. Mag.*, vol. 43, pp. 93–101, October 2005.

[15] J. Kuruvila, A. Nayak, and I. Stojmenovic, “Hop count optimal position based packet routing algorithms for ad hoc wireless networks with a realistic physical layer,” *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 1267–1275, June 2005.

[16] S. H. Low, F. Paganini, and J. C. Doyle, “Internet congestion control,” *IEEE Control Syst. Mag.*, vol. 22, pp. 28–43, February 2002.

[17] 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.

[18] T. Nadeem and A. Agrawala, “IEEE 802.11 fragmentation-aware energy-efficient ad-hoc routing protocols,” in *Proc. of the 1st IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, pp. 90–103, Fort Lauderdale, FL, Oct. 2004.

[19] R. Rao and A. Ephremides, “On the stability of interacting queues in a multi-access system,” *IEEE Trans. Inf. Theory*, vol. 34, pp. 918–930, September 1988.

[20] A. Ribeiro, Z. Q. Luo, N. D. Sidiropoulos, and G. B. Giannakis, “A General Optimization Framework for Stochastic Routing in Wireless Multi-hop Networks,” *IEEE Trans. Signal Process.*, July 2006 (submitted).

[21] 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.

[22] 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.

[23] G. Scutari, D. P. Palomar, and S. Barbarossa, “Optimal multiplexing strategies for wideband meshed networks based on game theory part II: algorithms,” *IEEE Trans. Signal Process.*, June 2006 (submitted).

[24] E. S. Souza and J. A. Silvester, “Optimum transmission ranges in a direct sequence spread-spectrum multihop packet radio network,” *IEEE J. Sel. Areas Commun.*, vol. 8, pp. 762–771, June 1990.

[25] I. Stojmenovic, A. Nayak, and J. Kuruvila, “Design guidelines for routing protocols in ad hoc and sensor networks with a realistic physical layer,” *IEEE Commun. Mag.*, vol. 43, pp. 101–106, March 2005.