

Research Article

On Traffic Load Distribution and Load Balancing in Dense Wireless Multihop Networks

Esa Hyytiä¹ and Jorma Virtamo²

¹The Telecommunications Research Center Vienna (ftw.), Donau-City Strasse 1, 1220 Vienna, Austria

²Networking Laboratory, Helsinki University of Technology, P.O. Box 3000, 02015 TKK, Finland

Received 29 September 2006; Accepted 13 March 2007

Recommended by Stavros Toumpis

We study the load balancing problem in a dense wireless multihop network, where a typical path consists of a large number of hops, that is, the spatial scales of a typical distance between source and destination and mean distance between the neighboring nodes are strongly separated. In this limit, we present a general framework for analyzing the traffic load resulting from a given set of paths and traffic demands. We formulate the load balancing problem as a minmax problem and give two lower bounds for the achievable minimal maximum traffic load. The framework is illustrated by considering the load balancing problem of uniformly distributed traffic demands in a unit disk. For this special case, we derive efficient expressions for computing the resulting traffic load for a given set of paths. By using these expressions, we are able to optimize a parameterized set of paths yielding a particularly flat traffic load distribution which decreases the maximum traffic load in the network by 40% in comparison with the shortest-path routing.

Copyright © 2007 E. Hyytiä and J. Virtamo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

In a wireless multihop network, a typical path consists of several hops and the intermediate nodes along a path act as relays. Thus, in general, each node has two functions. First, they can act as a source or a destination for some flow, that is, the nodes can communicate with each other. Second, when necessary, nodes have to relay packets belonging to the flows between other nodes.

Several types of wireless multihop networks exist with different unique characteristics. For example, wireless sensor networks are networks designed to collect some information from a given area and to deliver the information to one or more sinks. Thus, for example, the traffic distribution in sensor networks is typically highly asymmetric. Another example of wireless multihop network is a wireless mesh network consisting of both mobile and fixed wireless nodes and one or more gateway nodes through which the users have access to the Internet.

In this paper, we focus on studying a wireless multihop network at the limit when the number of nodes is large. At this limit, the network is often referred to as a massively dense network [1–3], or simply a dense network [4, 5]. In particular, we assume a strong separation in spatial scales

between the macroscopic level, corresponding to a distance between the source and destination nodes, and the microscopic level, corresponding to a typical distance between the neighboring nodes. This assumption justifies modeling the routes on the macroscopic scale as smooth geometric curves as if the underlying network fabric formed a homogeneous and isotropic (homogeneity and isotropicity are not crucial but are assumed here to simplify the discussion) continuous medium.

The microscopic scale corresponds to a single node and its immediate neighbors. At this scale, the above assumptions imply that only the direction in which a particular packet is traversing is significant. In particular, considering one direction at a time, there exists a certain maximum flow of packets a given MAC protocol can support (packets per unit time per unit length, “density of progress”). Generally, this maximal sustainable directed packet flow depends on the particular MAC protocol defining the scheduling rules and possible coordination between the nodes. Determining the value of this maximum is not a topic of this paper but is assumed to be given (known characteristic constant of the medium). By a simple time-sharing mechanism, this maximal value can be shared between flows propagating in different directions. As a result, the scalar or total flux (to be defined in Section 3)

of packets is bounded by the given maximum, and the load balancing task is to determine the paths in such a way that the maximum flux is minimized.

Under the assumption of a dense multihop network, the shortest paths (SPs) are at macroscopic-level straight line segments [6]. Straight paths yield an optimal solution in terms of mean delay when the traffic demands are low and there are no queueing delays. However, they typically concentrate significantly more traffic in the centre of network than elsewhere, and as the traffic load increases the packets going through the centre of the network start to experience queueing delays and eventually the system becomes unstable when the maximal sustainable scalar flux is exceeded. Hence, the use of shortest paths limits the capacity of the multihop network unnecessarily and our task is to minimize the maximum packet flux in the network by a proper choice of paths on the macroscopic scale. Note that in this paper, we are not addressing details of any routing protocol. The idea is, however, that when the destination of the packet is known, also the optimal macroscopic path to the destination is known. This path determines the direction to which the packet should be forwarded, and this information is used at the node level to make the actual forwarding decisions.

The main contributions of this paper are the formulation of the traffic load and the corresponding load balancing problem in general case, and the derivation of a computationally efficient expression for traffic load in a symmetric case of a unit disk, which then allows us to optimize a parameterized family of paths. By traffic load we mean, roughly speaking, the rate at which packets are transmitted in the proximity of a given node, and the objective of load balancing is to find such paths that minimize the maximum traffic load in the network. Formally, the spatial traffic load distribution is defined as a scalar packet flux.

The organization of the paper is as follows. First, in Section 2 related earlier work is briefly reviewed. Then, in Section 3 we present the necessary mathematical framework, that is, give a formal definition for different quantities at the limit of (massively) dense network. In Section 4 we concentrate on deriving some bounds for the load balancing problem. The load balancing problem in wired networks is well known and provides some insight into this problem. In particular, we give two lower bounds for the load balancing problem, where both bounds have a similar counterpart in wired networks. Then, in Section 5 we return to the original problem and derive general expressions for the traffic load with curvilinear paths. In Section 6 we demonstrate the framework by considering a unit disk with uniform traffic demands. First, we evaluate two heuristically chosen path sets and compare their performance to the one of shortest paths and to the lower bounds. Then we derive a simple computationally efficient expression for evaluating the traffic load for a general family of paths, making full use of the symmetry of the problem. By using these expressions, we finally optimize a parameterized set of paths which yields about 40% reduction of the maximum traffic load. Section 7 contains our conclusions. Even though the results presented in this work are valid only in the limit of a dense network (i.e., a

large number of nodes and a small transmission range), they give insight to the problem and can serve as useful approximations for more realistic scenarios.

2. RELATED WORK

A lot of earlier work has been devoted to different aspects of large-scale wireless multihop networks. In [6], Pham and perreau, and later in [7] Ganjali and Keshavarzian have studied the load balancing using multipath routes instead of shortest paths. The analysis is done assuming a disk area and a high node density so that the shortest paths correspond to straight line segments. In multipath situation, the straight line segments are replaced by rectangular areas where the width of the rectangle is related to the number of multiple paths between a given pair of nodes. In particular, multiple paths are fixed on both sides of the shortest path.

In [8], Dousse et al. study the impact of interference on the connectivity of large ad hoc networks. They assume an infinite area and the behavior of each node to be independent of other nodes, which, together with interference assumptions, define the stochastic properties for the existence of links. With these assumptions, the authors study the existence of a gigantic component, which is related to the network connectivity.

In [5], Sirkeci-Mergen and Scaglione study a dense wireless network with cooperative relaying, where several nodes transmit the same packet simultaneously in order to achieve a better signal-to-noise ratio. In the analysis, an infinitely long strip is studied and the authors are able to identify a so-called critical decoding threshold for the decoder, above which the message is practically transmitted to any distance (along the strip). The analysis assumes a dense network similarly as in the present paper.

In [1], Jacquet studies also the problem of optimal routes in (massively) dense wireless network. The problem is approached by studying a so-called *traffic density* denoted by $\lambda(\mathbf{r})$ and expressed in bit/s/m². Relying on the famous result by Gupta and Kumar [9], it is assumed that the mean hop length in the vicinity of \mathbf{r} is $\beta/\lambda(\mathbf{r})$, where β is some constant depending on, for example, MAC protocol and environment. Consequently, at the limit of dense network, the mean number of hops along route \mathcal{C} is given by $\int_{\mathcal{C}} n(\mathbf{r}) ds$, where $n(\mathbf{r}) = \lambda(\mathbf{r})/\beta$. The optimization problem is then formulated as finding such a route for a given source-destination pair $(\mathbf{r}_1, \mathbf{r}_2)$ that minimizes the mean number of hops. In particular, it is assumed that the traffic belonging to the given path does not have significant effect on the traffic density. In this case, quantity $n(\mathbf{r})$ can be interpreted as a nonlinear optical density and finding the optimal path is equivalent to finding the path light traverses in a medium with optical index of refraction $\lambda(\mathbf{r})$. It is further pointed out that the general problem of determining the optimal paths for all possible pairs of locations may be a hard problem as the distribution of paths affects the traffic density.

In a similar fashion, Kalantari and Shayman [10] and Toumpis and Tassiulas [2] have studied dense wireless multihop networks by leaning to theory of electrostatics. In

particular, Kalantari and Shayman consider the routing problem where a large number of nodes are sending data to a single destination. In this case, the optimal paths are obtained by solving a set of partial differential equations similar to Maxwell's equations in the theory of electrostatics. Toupis and Tassiulas [2], on the other hand, have studied a related problem of optimal placement of the nodes in a dense sensor network. The approach is also based on the analogy with electrostatics. It seems, however, essential for the used approach that at any point of the network, the information flows exactly to one direction only, which can be argued to be a reasonable assumption for a sensor network. However, in general case there will be "crossing traffic" at each point of the network.

In a dense network with shortest-path routing, the transmission of each packet corresponds to a line segment in the area of the network. This line segment process with uniformly distributed endpoints is similar to the so-called random waypoint (RWP) mobility model commonly used in studies of wireless ad hoc networks [11–14]. In the RWP model the nodes move along straight line segments from one waypoint to the next and the waypoints are assumed to be uniformly distributed in some convex domain. The similarity between the RWP process and the packet transport with the shortest path routes is striking and we can utilize the readily available results from [15] in this case. For curvilinear paths, the situation, however, is more complicated and the new results derived in the present paper allow us to compute the resulting scalar packet flux (i.e., traffic load).

3. PRELIMINARIES

In this section, we introduce the necessary notation and definitions for analyzing the transport of the packets and the resulting traffic load in the network. Let \mathcal{A} denote a two-dimensional region where the network is located and A is the area of \mathcal{A} . The packet generation rate corresponding to traffic demand density is defined as follows.

Definition 1 (traffic demand density). The rate of flow of packets from a differential area element dA about \mathbf{r}_1 to a differential area element dA about \mathbf{r}_2 is $\lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot dA^2$, where $\lambda(\mathbf{r}_1, \mathbf{r}_2)$ is called the traffic demand density and is measured in units $1/s/m^4$.

Remark 1. The total packet generation rate measured in $1/s$ is given by

$$\Lambda = \int_{\mathcal{A}} d^2\mathbf{r}_1 \int_{\mathcal{A}} d^2\mathbf{r}_2 \lambda(\mathbf{r}_1, \mathbf{r}_2). \quad (1)$$

Each generated packet is forwarded along some multihop path.

Definition 2 (paths). Set of paths, denoted by \mathcal{P} , defines directed continuous loop free paths in \mathcal{A} . In case of single-path routes, set \mathcal{P} consists of exactly one path for each

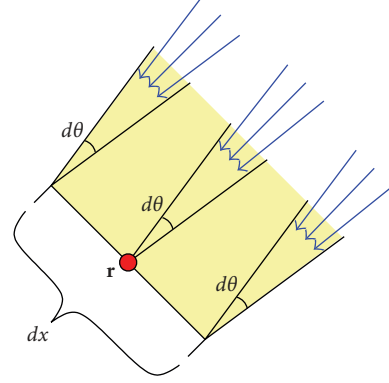


FIGURE 1: Angular flux $\varphi(\mathbf{r}, \theta)$ is the rate of packets crossing a small perpendicular line segment dx from angle $(\theta, \theta + d\theta)$ divided by $d\theta \cdot dx$ at the limit $d\theta, dx \rightarrow 0$.

source-destination pair. For multipath routes, it is further assumed that the corresponding proportions are well defined in \mathcal{P} .

In this paper, we are mainly concerned with single-path routing, but in Section 6.3 also multipath routing is considered.

Remark 2. The mean path length, that is, the mean distance a packet travels measured in m, is given by

$$\bar{\ell} = \frac{1}{\Lambda} \int_{\mathcal{A}} d^2\mathbf{r}_1 \int_{\mathcal{A}} d^2\mathbf{r}_2 \lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot s(\mathcal{P}, \mathbf{r}_1, \mathbf{r}_2), \quad (2)$$

where $s(\mathcal{P}, \mathbf{r}_1, \mathbf{r}_2)$ denotes the (mean) distance from \mathbf{r}_1 to \mathbf{r}_2 with path set \mathcal{P} .

Example 1. For the shortest paths, we have

$$\bar{\ell}_{\text{sp}} = \frac{1}{\Lambda} \int_{\mathcal{A}} d^2\mathbf{r}_1 \int_{\mathcal{A}} d^2\mathbf{r}_2 \lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot |\mathbf{r}_2 - \mathbf{r}_1|. \quad (3)$$

Note that in our setting at each point the information can flow to any direction (depending on the destination of each packet) in contrast to the sensor networks where it can be assumed that at any given location the information flows to exactly one direction [2].

Probably the most important quantity for our purposes is the packet arrival rate into the proximity of a given node. This is described by the notion of scalar flux, which in turn is defined in terms of the angular flux. These are similar to corresponding concepts of particle fluxes in physics, for example, in neutron transport theory [16]. In our case, the packet fluxes depend on the traffic demand density $\lambda(\mathbf{r}_1, \mathbf{r}_2)$ and the chosen paths \mathcal{P} , and are defined as follows (see also Figure 1).

Definition 3 (angular flux). Angular flux of packets at \mathbf{r} in direction θ , denoted by $\varphi(\mathbf{r}, \theta) = \varphi(\mathcal{P}, \mathbf{r}, \theta)$, is equal to the rate ($1/s/m/\text{rad}$) at which packets flow in the angle interval $(\theta, \theta + d\theta)$ across a small line segment of the length dx perpendicular to direction θ at point \mathbf{r} divided by $dx \cdot d\theta$ in the limit $dx \rightarrow 0$ and $d\theta \rightarrow 0$.

Definition 4 (scalar flux). Scalar flux of packets (1/s/m) at \mathbf{r} is given by

$$\Phi(\mathbf{r}) = \Phi(\mathcal{P}, \mathbf{r}) = \int_0^{2\pi} \varphi(\mathcal{P}, \mathbf{r}, \theta) d\theta. \quad (4)$$

With the above notation, we can formulate the optimization problem.

Definition 5 (load balancing problem). Find such a set of paths, \mathcal{P}_{opt} , that minimizes the maximum scalar flux,

$$\mathcal{P}_{\text{opt}} = \arg \min_{\mathcal{P}} \max_{\mathbf{r}} \Phi(\mathcal{P}, \mathbf{r}). \quad (5)$$

Remark 3 (optimal maximum traffic load). With the load balanced paths, the maximum load is

$$\Phi_{\text{opt}} = \max_{\mathbf{r}} \Phi(\mathcal{P}_{\text{opt}}, \mathbf{r}) = \min_{\mathcal{P}} \max_{\mathbf{r}} \Phi(\mathcal{P}, \mathbf{r}). \quad (6)$$

In Definition 5, one needs the scalar flux $\Phi(\mathcal{P}, \mathbf{r})$. In Section 5, we will show how this can be calculated for a given set of paths \mathcal{P} , and in Section 6 we present a particularly simple and efficient formula for calculating the flux in a circularly symmetrical system. The remaining problem of finding the optimal paths is a difficult problem of calculus of variation. In this paper, we do not search for a general solution but rather study three heuristically chosen families of paths and compare their performance with that of the shortest paths and with the bounds introduced in the next section.

4. LOWER BOUNDS FOR SCALAR PACKET FLUX

Our next goal is to derive two lower bounds for achievable load balancing, that is, for a given traffic demand density $\lambda(\mathbf{r}_1, \mathbf{r}_2)$, we want to find bounds for the minimum of the maximal traffic load that can be obtained by a proper choice of paths. These lower bounds are valid for both single and multipath routes. Let us start with two preparatory remarks that give additional characterizations of the scalar flux.

Remark 4. Scalar flux of packets is equal to the rate at which packets enter a disk with diameter d at point \mathbf{r} divided by d in the limit when $d \rightarrow 0$.

The proof follows trivially from the definitions. Note that Remark 4 justifies the interpretation of the scalar packet flux as a measure of spatial traffic load.

Remark 5 (density of cumulative progress rate). Scalar flux $\Phi(\mathbf{r})$ can also be interpreted as the cumulative progress (m) of packets per unit time (s) per unit area (m^2) about point \mathbf{r} (rendering 1/s/m as its dimension). By progress we mean the advance a packet has made in a given time interval in the direction of its path.

Proof. Consider the packet flux within small angle interval $d\theta$ entering a small square with side h from left as shown in Figure 2, ultimately letting $d\theta \rightarrow 0$ and $h \rightarrow 0$. According to Definition 3, the rate of such packets is $\varphi(\mathbf{r}, \theta) \cdot h \cdot d\theta$. The

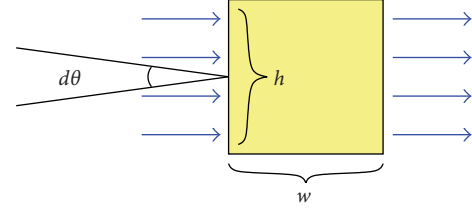


FIGURE 2: Cumulative progress in a small square.

same flow departs the square from the right side. Thus, inside the square the cumulative progress per unit time (for packets moving within the angle interval $d\theta$) is $\varphi(\mathbf{r}, \theta) \cdot h \cdot d\theta \cdot w$. Per unit area, the above yields $\varphi(\mathbf{r}, \theta) d\theta$. Integrating over θ then gives that $\Phi(\mathbf{r})$ corresponds to the cumulative progress per unit time and unit area. \square

Proposition 1 (distance bound).

$$\max_{\mathbf{r}} \Phi(\mathcal{P}, \mathbf{r}) \geq \frac{\Lambda \cdot \bar{\ell}}{A}. \quad (7)$$

Proof. The cumulative progress rate in the whole area is obviously $\Lambda \cdot \bar{\ell}$. Thus, the right-hand side equals the average density of progress rate, that is, the average scalar flux. \square

Remark 6. Accordingly, we have identity

$$\Lambda \cdot \bar{\ell} = A \cdot \text{mean } \Phi(\mathbf{r}). \quad (8)$$

For example, in the absence of congestion there are no queueing delays and the (mean) sojourn time of a packet is proportional to the (mean) path length. Then (8) is similar to Little's result for the mean number of customers in a single server queue.

Remark 7. Combining (6) and (7), we have

$$\Phi_{\text{opt}} \geq \frac{\Lambda}{A} \min_{\mathcal{P}} \bar{\ell}. \quad (9)$$

It is obvious that the minimum of $\bar{\ell}$ is obtained when \mathcal{P} consists of the shortest paths. Denoting the corresponding mean path length by $\bar{\ell}_{\text{sp}}$, (cf. (3)), we get

$$\Phi_{\text{opt}} \geq \frac{\Lambda \cdot \bar{\ell}_{\text{sp}}}{A}. \quad (10)$$

Another bound is obtained by considering traffic flows crossing an arbitrary boundary (cf., cut bound in wired networks).

Proposition 2 (cut bound). For any curve \mathcal{C} which separates the domain \mathcal{A} into two disjoint subdomains \mathcal{A}_1 and \mathcal{A}_2 , it holds that

$$\Phi_{\text{opt}} \geq \frac{1}{L} \int_{\mathcal{A}_1} d^2 \mathbf{r}_1 \int_{\mathcal{A}_2} d^2 \mathbf{r}_2 (\lambda(\mathbf{r}_1, \mathbf{r}_2) + \lambda(\mathbf{r}_2, \mathbf{r}_1)), \quad (11)$$

where L is the length of the curve \mathcal{C} and the double integral gives the total rate of packets between \mathcal{A}_1 and \mathcal{A}_2 (both directions included).

Proof. Consider first a short line segment dx at \mathbf{r} at some point along the curve \mathcal{C} . Let γ denote a direction perpendicular to the curve at \mathbf{r} such that the packets arriving from the angles $(\gamma - \pi/2, \gamma + \pi/2)$ cross dx from side 2 to side 1, and packets arriving from $(\gamma + \pi/2, \gamma + 3\pi/2)$ cross dx from side 1 to side 2. The rate $\lambda(\mathbf{r})dx$ at which packets move across dx is given by

$$\lambda(\mathbf{r})dx = \int_{-\pi/2}^{\pi/2} \cos \alpha (\Phi(\mathbf{r}, \gamma + \alpha) + \Phi(\mathbf{r}, \gamma + \alpha + \pi)) d\alpha dx, \quad (12)$$

which yields

$$\begin{aligned} \lambda(\mathbf{r})dx &\leq \int_{-\pi/2}^{\pi/2} \Phi(\mathbf{r}, \gamma + \alpha) + \Phi(\mathbf{r}, \gamma + \alpha + \pi) d\alpha dx \\ &= \Phi(\mathbf{r})dx \leq \max_{x \in \mathcal{A}} \Phi(x)dx. \end{aligned} \quad (13)$$

Integrating over the curve \mathcal{C} completes the proof. \square

5. SCALAR PACKET FLUX WITH CURVILINEAR PATHS

In this section, unless stated otherwise, we assume uniform traffic demand density. We make the assumption of uniformity mainly for notational simplicity. It is easy to generalize the results for any distribution. Also single-path routes are implicitly assumed throughout the section.

Definition 6 (single path). Packets from \mathbf{r}_1 to \mathbf{r}_2 are forwarded along a unique loop free path denoted by $\tilde{p}(\mathbf{r}_1, \mathbf{r}_2)$.

Next, we give some additional properties that characterize the single-path routes considered in this study.

Definition 7 (bidirectionality). The paths are bidirectional if $\tilde{p}(\mathbf{r}_2, \mathbf{r}_1)$ is $\tilde{p}(\mathbf{r}_1, \mathbf{r}_2)$ in reverse direction.

Note that a flow on a given path contributes to the scalar flux at any point on the path by an amount equal to the absolute size of the flow, no matter what the direction of the flow is. Thus, allowing a different return path is, from the load balancing point of view, essentially equivalent to allowing two paths for each pair of locations.

Definition 8 (destination-based forwarding). The paths adhere to a destination-based forwarding rule if

$$\mathbf{r} \in \tilde{p}(\mathbf{r}_1, \mathbf{r}_2) \implies \tilde{p}(\mathbf{r}, \mathbf{r}_2) \subset \tilde{p}(\mathbf{r}_1, \mathbf{r}_2). \quad (14)$$

The above definition means that the routing decision made at each point depends on the destination of the packet only, not on the source. Fixing destination \mathbf{x} induces a set of curves along which the packets are routed towards \mathbf{x} (see Figure 9 for illustration). Together with bidirectional paths (Definition 7), the same curves also describe how the packets from \mathbf{x} are forwarded to all possible destinations.

Definition 9 (path continuity). Path continuity is satisfied if

$$\mathbf{r} \in \tilde{p}(\mathbf{r}_1, \mathbf{r}_2) \implies \tilde{p}(\mathbf{r}_1, \mathbf{r}_2) = \tilde{p}(\mathbf{r}_1, \mathbf{r}) \cup \tilde{p}(\mathbf{r}, \mathbf{r}_2). \quad (15)$$

Note that (i) Definitions 7 and 8 \implies Definition 9, and (ii) Definition 9 \implies Definition 8. In this section we, however, assume that the set of paths is defined by a family of continuous curves.

Definition 10 (paths defined by curves). Paths are defined by a family of curves \mathcal{C} for which it holds that

- (i) the curves are continuous, piecewise smooth, and loop-free;
- (ii) given two points \mathbf{r}_1 and \mathbf{r}_2 , there exists a unique curve $c \in \mathcal{C}$ to which both points belong. This curve then defines the path $\tilde{p}(\mathbf{r}_1, \mathbf{r}_2)$.

From Definition 10, it follows that also Definitions 6–9 are satisfied. Moreover, unambiguity of curves in condition (ii) implies that the curves may not cross each other except at \mathbf{x} (and possibly at the endpoints, which can be neglected). In particular, Definition 10 allows one to characterize the curves going through \mathbf{x} according to their direction at \mathbf{x} . To this end, consider a small ϵ -circle at \mathbf{x} and an arbitrary point \mathbf{x} outside the circle. According to condition (ii), there is a unique continuous curve c connecting \mathbf{r} to \mathbf{x} , which defines the path from \mathbf{r} to \mathbf{x} . This path cuts the circumference of ϵ -circle at a certain point \mathbf{r}_ϵ . Furthermore, unambiguity of the curves ensures that c is the only curve to which \mathbf{x} and \mathbf{r}_ϵ belong, thus defining the direction θ in the limit $\epsilon \rightarrow 0$. Hence, we let $p(\mathbf{x}, \theta)$ denote a curve going through point \mathbf{x} in direction θ . The points along the curve are denoted by

$$\mathbf{p}(\mathbf{x}, \theta, s), \quad s \in [-a_1, a_2], \quad a_1, a_2 > 0, \quad (16)$$

where $\mathbf{p}(\mathbf{x}, \theta, 0) = \mathbf{x}$, and a_1 and a_2 denote the distances to the boundary along the curve in opposite directions.

For simplicity of notation, we furthermore assume that the curves defining the paths towards (and from) \mathbf{x} start from the boundary. Then, $a_1 = a_1(\mathbf{x}, \theta)$ and $a_2 = a_2(\mathbf{x}, \theta)$. In general, we can also allow closed curves and curves with endpoints inside the domain. For the closed curves, one must explicitly define which direction is to be taken. Thus, in this case, $a_1 = a_1(\mathbf{x}, \theta)$ defines the maximum distance from \mathbf{x} along path $p(\mathbf{x}, \theta)$ in “negative direction” from where a packet is forwarded across point \mathbf{x} to the “positive side.” Similarly, $a_2 = a_2(\mathbf{x}, \theta, s)$ defines the maximum distance on the “positive side,” measured from \mathbf{x} , to where nodes about $\mathbf{p}(\mathbf{x}, \theta, -s)$, $0 < s < a_1$, communicate to using the path $p(\mathbf{x}, \theta)$. This complicates the notation unnecessarily, and thus in the following we assume that the curves start and end at the boundary. However, it is straightforward to show that essentially the same results hold also in the general case where some of the curves may be closed or have the endpoints inside the domain.

Definition 11 (curve divergence). Let $h(\mathbf{x}, \theta, s)$ denote the rate with respect to the angle θ at which curves going through \mathbf{x} diverge at the distance of s ,

$$h(\mathbf{x}, \theta, s) = \left| \frac{\partial}{\partial \theta} \mathbf{p}(\mathbf{x}, \theta, s) \right|. \quad (17)$$

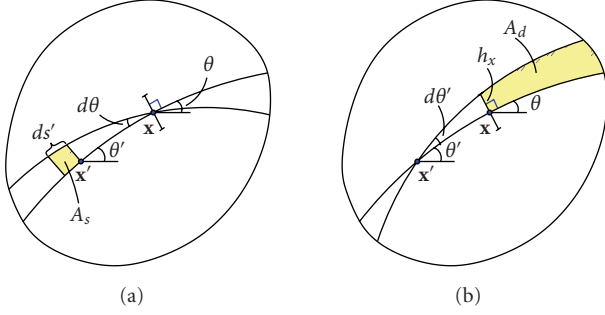


FIGURE 3: Derivation of expression (18) for the scalar flux.

The curve divergence is assumed to be (piecewise) well defined and finite with a given set of curves.

Proposition 3 (angular flux with curvilinear paths). *For uniform traffic demand density, $\lambda(\mathbf{r}_1, \mathbf{r}_2) = \Lambda/A^2$, the angular flux at point \mathbf{x} in direction θ is given by*

$$\varphi(\mathbf{x}, \theta) = \frac{\Lambda}{A^2} \int_0^{a_1} \frac{h(\mathbf{x}, \theta, -s')}{h(\mathbf{x}', \theta', s')} \int_0^{a_2} h(\mathbf{x}', \theta', s+s') ds ds', \quad (18)$$

where $\mathbf{x}' = \mathbf{p}(\mathbf{x}, \theta, -s')$ and θ' is the direction of the path at \mathbf{x}' (see Figure 3).

Proof. Without loss of generality, we may assume that $\Lambda = 1$. The aim is to determine the angular flux at \mathbf{x} in direction θ . To this end, consider path $\mathbf{p}(\mathbf{x}, \theta, s)$, where s denotes the position on path relative to \mathbf{x} (positive in one direction, negative in other). Assume that a particular source contributing the angular flux is located in a differential area element about point \mathbf{x}' (see Figure 3(a)),

$$\mathbf{x}' = \mathbf{p}(\mathbf{x}, \theta, s'), \quad s' \leq 0, \quad (19)$$

for which it clearly holds that (the same curve)

$$\mathbf{p}(\mathbf{x}', \theta', s - s') = \mathbf{p}(\mathbf{x}, \theta, s). \quad (20)$$

Let $d\theta$ denote a differential angle at \mathbf{x} as illustrated in Figure 3(a). According to (17), the differential source area about \mathbf{x}' is given by

$$A_s = h(\mathbf{x}, \theta, s') \cdot d\theta \cdot ds'. \quad (21)$$

Similarly, let $d\theta'$ denote a small angle at point \mathbf{x}' , which yields a destination area of

$$A_d = \int_0^{a_2} h(\mathbf{x}', \theta', s - s') ds d\theta', \quad (22)$$

as illustrated in Figure 3(b). The curve divergence at \mathbf{x}' tells us the perpendicular distance of two paths passing \mathbf{x}' in directions θ' and $\theta' + d\theta'$ as a function of the distance s' along the path. Thus, the height of the “target line segment” perpendicular to the path at point \mathbf{x} is $h_x = h(\mathbf{x}', \theta', -s') \cdot d\theta'$,

and the contribution to the angular flux from the differential source area A_s about \mathbf{x}' is

$$\begin{aligned} d\varphi &= \frac{A_s \cdot A_d}{A^2 \cdot d\theta \cdot h_x} \\ &= \frac{1}{A^2} \cdot \frac{1}{d\theta} \cdot \left(\frac{1}{h(\mathbf{x}', \theta', -s') \cdot d\theta'} \right) \\ &\quad \cdot (h(\mathbf{x}, \theta, s') \cdot d\theta \cdot ds') \cdot \int_0^{a_2} h(\mathbf{x}', \theta', s - s') ds d\theta' \\ &= \frac{1}{A^2} \cdot \frac{h(\mathbf{x}, \theta, s')}{h(\mathbf{x}', \theta', -s')} \cdot \int_0^{a_2} h(\mathbf{x}', \theta', s - s') ds ds'. \end{aligned} \quad (23)$$

Consequently, the angular flux at \mathbf{x} in direction θ is given by

$$\varphi(\mathbf{x}, \theta) = \frac{1}{A^2} \int_{-a_1}^0 \frac{h(\mathbf{x}, \theta, s')}{h(\mathbf{x}', \theta', -s')} \int_0^{a_2} h(\mathbf{x}', \theta', s - s') ds ds'. \quad (24)$$

The proposition follows upon substitution $s' \leftarrow -s'$. \square

Remark 8 (angular flux with nonuniform $\lambda(\mathbf{r}_1, \mathbf{r}_2)$). It is straightforward to generalize (18) to the case of nonuniform traffic demand density $\lambda(\mathbf{r}_1, \mathbf{r}_2)$. In this case, the angular flux at \mathbf{x} in direction θ is given by

$$\begin{aligned} \varphi(\mathbf{x}, \theta) &= \int_0^{a_1} \frac{h(\mathbf{x}, \theta, -s')}{h(\mathbf{x}', \theta', s')} \\ &\quad \cdot \int_0^{a_2} \lambda(\mathbf{x}', \mathbf{p}(\mathbf{x}', \theta', s+s')) \cdot h(\mathbf{x}', \theta', s+s') ds ds'. \end{aligned} \quad (25)$$

Example 2 (shortest paths). For the shortest paths, that is, straight lines,

$$h(\mathbf{x}, \theta, s) = |s|, \quad (26)$$

and the angular flux is given by

$$\varphi(\mathbf{x}, \theta) = \int_0^{a_1} \int_0^{a_2} \lambda(\mathbf{r}_1, \mathbf{r}_2) \cdot (s + s') ds ds', \quad (27)$$

where $\mathbf{r}_1 = \mathbf{x} - s' \mathbf{e}_\theta$, and $\mathbf{r}_2 = \mathbf{x} + s \mathbf{e}_\theta$, with \mathbf{e}_θ denoting the unit vector in direction θ . Consequently, for uniform traffic demand density,

$$\begin{aligned} \varphi(\mathbf{x}, \theta) &= \frac{\Lambda}{A^2} \int_0^{a_1} \int_0^{a_2} (s+s') ds ds' \\ &= \frac{\Lambda}{2A^2} a_1 a_2 (a_1 + a_2), \end{aligned} \quad (28)$$

in accordance with the result on RWP model in [17].

Remark 9 (optical paths). A family of paths can be defined in terms of paths of light rays in an optical medium with index of refraction $n(\mathbf{x})$. For optical paths, it can be shown with the aid of Snell's law that

$$\frac{h(\mathbf{x}, \theta, -s')}{h(\mathbf{x}', \theta', s')} = \frac{n(\mathbf{x})}{n(\mathbf{x}')}. \quad (29)$$

Substituting (29) into (18) yields

$$\varphi(\mathbf{x}, \theta) = \frac{n(\mathbf{x})}{A^2} \int_0^{a_1} \int_0^{a_2} \frac{h(\mathbf{x}', \theta', s + s')}{n(\mathbf{x}')} ds ds'. \quad (30)$$

It is worth noting that the optical paths minimize the mean travelling time assuming that the velocity of the packet is inversely proportional to the index of refraction,

$$\min_{\mathbf{p}: \mathbf{p}(0)=\mathbf{r}_1, \mathbf{p}(\ell)=\mathbf{r}_2} \int_0^\ell n(\mathbf{p}(s)) ds. \quad (31)$$

6. UNIT DISK WITH UNIFORM TRAFFIC DEMANDS

In this section, we will demonstrate how the proposed framework can be applied. To this end, we consider a special case of a unit disk with uniform load,

$$\mathcal{A} = \{\mathbf{r} \in \mathbb{R}^2 : |\mathbf{r}| < 1\}, \quad \lambda(\mathbf{r}_1, \mathbf{r}_2) = \frac{\Lambda}{\pi^2}. \quad (32)$$

First, we study the performance of two simple families of paths: outer and inner radial ring paths. The performance of these path sets is compared with that of the shortest paths, and with the appropriate lower bounds for the minimal maximum traffic load. Then we focus on a general family of paths and derive computationally efficient expression for calculating the packet flux distribution in this special case of unit. Using these expressions we further evaluate the so-called circular and modified circular path sets, where the parameters of the latter form are optimized.

Example 3 (shortest paths in unit disk). For transport according to the straight line segments, we can either use (28) or rely on the results for the RWP model (see [15]). Accordingly, the scalar flux at the distance of r from the origin is given by

$$\Phi_{\text{sp}}(r) = \frac{2(1-r^2) \cdot \Lambda}{\pi^2} \int_0^\pi \sqrt{1-r^2 \cos^2 \phi} d\phi. \quad (33)$$

The function $\Phi_{\text{sp}}(r)$ is depicted in Figure 5 (denoted by SP). In particular, the maximum flux is obtained at the centre,

$$\Phi_{\text{sp}}(0) = \frac{2}{\pi} \cdot \Lambda \approx 0.637 \cdot \Lambda. \quad (34)$$

Example 4 (distance bound for unit disk). The distance bound gives a relationship between the obtainable maximum load and the mean path length. With shortest paths, we have $\bar{\ell}_{\text{sp}} = 128/45\pi$ which upon substitution in (10) yields

$$\Phi_{\text{opt}} \geq \frac{\Lambda \cdot 128}{45\pi^2} \approx 0.288 \cdot \Lambda. \quad (35)$$

Example 5 (greatest sensible mean path length). With the aid of (34), we can write the distance bound (7) in terms of Φ_{sp} ,

$$\max_{\mathbf{r}} \Phi(\mathcal{P}, \mathbf{r}) \geq \Phi_{\text{sp}} \cdot \frac{\bar{\ell}}{2}. \quad (36)$$

Shortest paths are not optimal for uniform traffic demand density. But the above relation says that in searching for a

better set of paths (which necessarily has $\bar{\ell} \geq \bar{\ell}_{\text{sp}}$), one can outright reject such path sets for which $\bar{\ell} > 2$ since for them, the maximal scalar flux surely is greater than that for the shortest paths. That is, in order to lower the maximal flux, one has to bend the paths away from the loaded region but without increasing the mean length of the paths too much at the same time.

Example 6 (cut bounds for unit disk). Let us consider two curves, a diameter \mathcal{C}_1 separating the unit disk into two semicircles, and a concentric circle \mathcal{C}_2 with radius r , $0 < r < 1$. For the packet rate λ_1 across \mathcal{C}_1 , it holds that $\lambda_1 \geq \Lambda/2$, and

$$\Phi_{\text{opt}} \geq \frac{\Lambda}{4} = 0.25 \cdot \Lambda. \quad (37)$$

Similarly, the packet rate across \mathcal{C}_2 is bounded by $\lambda_2(r) \geq 2r^2(1-r^2) \cdot \Lambda$, which corresponds to radial flux

$$\Phi_r(r) = \frac{2r^2(1-r^2)}{2\pi r} \cdot \Lambda = \frac{r-r^3}{\pi} \cdot \Lambda. \quad (38)$$

By the cut bound we have $\Phi_{\text{opt}} \geq \Phi_r(r)$. The tightest lower bound is obtained by maximizing $\Phi_r(r)$ with respect to r ,

$$\Phi_{\text{opt}} \geq \Phi_r\left(\frac{1}{\sqrt{3}}\right) = \frac{2}{3\sqrt{3} \cdot \pi} \cdot \Lambda \approx 0.123 \cdot \Lambda. \quad (39)$$

We see that in the case of unit disk with uniform traffic demand density, the distance bound provides the tightest lower bound for the solution of the minmax problem (6).

6.1. Radial ring paths

Let us consider next the three actual path sets illustrated in Figure 4. The shortest paths (SPs) are equivalent to RWP model as has been already mentioned. The two radial path sets, referred to as ‘‘Rin’’ and ‘‘Rout,’’ are similar in the sense that each path consists of two sections. One section is a radial path towards (or away from) the origin, and the other section is an angular path along a ring with a given radius. The difference between the two sets is the order of sections, ‘‘Rin’’ uses the inner angular rings and ‘‘Rout’’ the outer ones, as the names suggest. Note that locally, at any point, the packets are transmitted only in 4 possible directions (2 radial and 2 angular), which may simplify the possible implementation of the time-division multiplexing. It is easy to see that the radial ring paths satisfy Definitions 6–9, but not condition (ii) of Definition 10. Thus, (18) cannot be used to calculate the scalar packet flux. However, given their simple form, the scalar packet flux can be easily obtained by other means.

In particular, when considering the arrival rate into a small area at the distance of r from the origin, one needs to consider only two components: (1) the radial component and (2) the angular component. The radial component of the flux is the same for both path sets, that is,

$$\Phi_r(r) = \frac{r-r^3}{\pi} \cdot \Lambda. \quad (40)$$

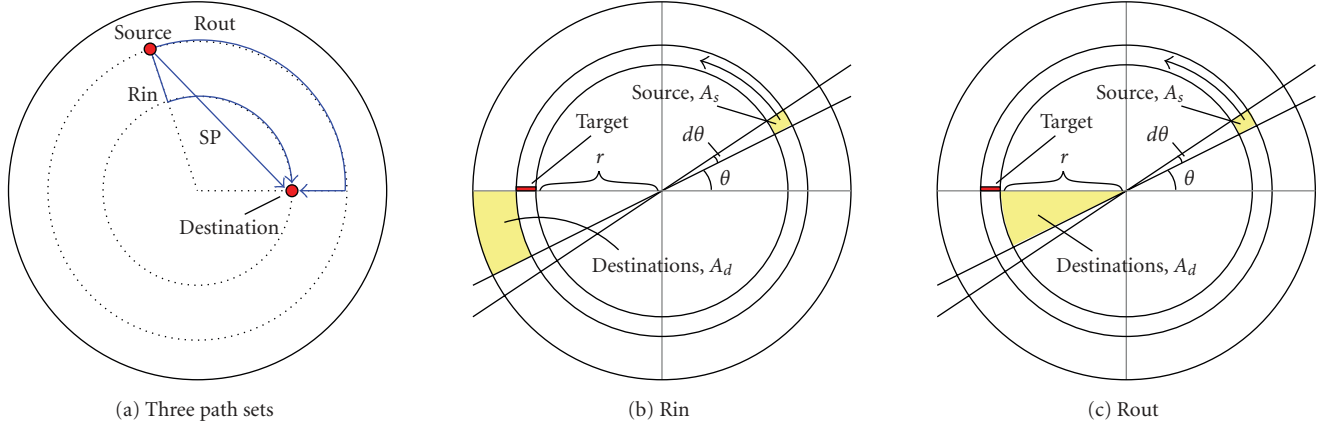


FIGURE 4: Radial ring paths. (a) illustrates the three path sets considered: straight line segments (SP), radial paths with outer (Rout) and inner (Rin) angular ring transitions. (b) illustrates the derivation of the angular ring flux at the distance r from the origin for Rin paths, and (c) for Rout paths.

6.1.1. Inner radial ring paths

Let us next consider inner radial ring paths. We want to determine the flux along the ring at the distance of r . To this end, consider a small line segment from $(-r, 0)$ to $(-r - \Delta, 0)$ as the target line segment, as illustrated in Figure 4(b). Packets originating from a small source area A_s at the distance of r in direction θ travel through the target line segment if their destination is in the destination area A_d . The size of the source area is

$$A_s = r \cdot \Delta \cdot d\theta, \quad (41)$$

while the possible destination area is

$$A_d = \frac{1 - r^2}{2} \cdot \theta. \quad (42)$$

Combining the above with $\lambda = \Lambda/\pi^2$, and taking into account the symmetries (factor of 4), gives the angular component of the flux at the distance of r ,

$$\begin{aligned} \Phi_\theta(r) &= \frac{4\Lambda}{\Delta\pi^2} \int_0^\pi \frac{1 - r^2}{2} \theta r \Delta d\theta \\ &= (r - r^3)\Lambda. \end{aligned} \quad (43)$$

Hence, the total flux at the distance r for the outer path set is given by

$$\Phi_{\text{Rin}}(r) = \Phi_r(r) + \Phi_\theta(r) = \frac{(\pi + 1)(r - r^3)}{\pi} \cdot \Lambda. \quad (44)$$

The maximum is obtained at $r = 1/\sqrt{3}$,

$$\Phi_{\text{Rin}}\left(\frac{1}{\sqrt{3}}\right) \approx 0.507 \cdot \Lambda. \quad (45)$$

6.1.2. Outer radial ring paths

For outer radial ring paths, we find by similar considerations (see Figure 4) that destination area of the packet going through the target line segment is $r^2/2 \cdot \theta$. Thus we have

$$\Phi_\theta(r) = \frac{4\Lambda}{\Delta\pi^2} \int_0^\pi \frac{r^2}{2} \cdot \theta \cdot r \cdot \Delta d\theta = r^3 \cdot \Lambda. \quad (46)$$

Combining the above with (40) gives

$$\Phi_{\text{Rout}}(r) = \frac{(\pi - 1)r^3 + r}{\pi} \cdot \Lambda. \quad (47)$$

The maximum flux is obtained at $r = 1$,

$$\Phi_{\text{Rout}}(1) = \Lambda. \quad (48)$$

6.1.3. Comparison of radial ring and shortest paths

The resulting scalar packet fluxes for these three path sets are illustrated in Figure 5 as a function of the distance r from the centre. It can be seen that each of them exhibits a rather distinctive form, none of which is flat. The key performance quantities are given in Table 1. Thus, the outer version leads to a clearly higher maximum load than the shortest paths while the inner version yields a slightly better solution.

According to (8), there is a direct relationship between the mean path length and the average scalar packet flux, that is, in unit disk with $\Lambda = 1$,

$$\text{mean } \Phi(\mathbf{r}) = \pi \cdot \bar{\ell}. \quad (49)$$

Consequently, by definition, the shortest-path routes yield always the minimum average scalar flux, and in order to decrease the maximum scalar flux one must at the same time increase the average scalar flux.

As mentioned, the shortest paths tend to concentrate too much traffic in the center of the area. The main shortcoming with the outer radial ring paths is easy to illustrate by

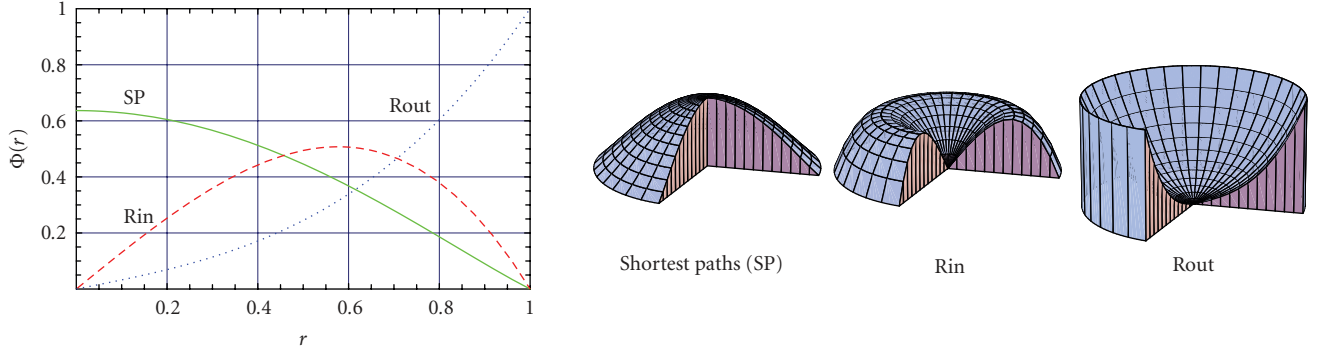


FIGURE 5: In the graph on left the resulting flux is plotted as a function of distance r from the center for the three path sets (SP, Rin, and Rout) in unit disk ($\Lambda = 1$). The 3D graphs on the right illustrate the same situation.

TABLE 1: Results with shortest and radial ring paths ($\Lambda = 1$).

Path set	Max. flux, max $\Phi(\mathbf{r})$	Average flux, mean $\Phi(\mathbf{r})$	Mean path length $\bar{\ell}$
Shortest paths (SP)	$\frac{2}{\pi} \approx 0.637$	$\frac{128}{45\pi^2} \approx 0.288$	$\frac{128}{45\pi} \approx 0.905$
Inner radial ring (Rin)	$\frac{2 + 2\pi}{3\sqrt{3}\pi} \approx 0.507$	$\frac{4 + 4\pi}{15\pi} \approx 0.352$	$\frac{4 + 4\pi}{15} \approx 1.104$
Outer radial ring (Rout)	1	$\frac{4 + 6\pi}{15\pi} \approx 0.485$	$\frac{4 + 6\pi}{15} \approx 1.523$

an example. Consider a situation where a source node is located near the origin, for example, about $(\epsilon, 0)$, and the destination is near the circumference about $(1 - \epsilon, 0)$. In such cases, the packet is first forwarded to a totally opposite direction until it reaches the perimeter and then along a half-circle to the destination, that is, the chosen route is clearly unefficient and contributes unnecessarily to the traffic load near the perimeter. Also the inner radial ring paths evade the center area too much. In the next section, we consider better smooth curvilinear paths which yield better performance in terms of a lower maximum scalar flux.

6.2. General paths in unit disk

While (18) provides a general formula for calculating the angular flux in the general case, and the scalar flux is then obtained by integration over angles (4), in the special case of circularly symmetric system the calculation of the scalar flux can be done in a simpler way by making full use of the symmetry. In this way we derive an explicit formula for the scalar flux as a function of the radius for a general family of paths. We then demonstrate the use of this formula for the minimization of the maximum flux with a two-parameter family of paths.

To begin with, we need a few definitions. The *basic set of paths* is given by the set of curves $y = y(x, a)$, where $y(x, a)$ is an even function of x , $y(x, a) = y(-x, a)$, that is, the curves are in a “horizontal position,” meaning for instance that the derivative is zero at $x = 0$. For each curve $y(x, a)$, also its mirror image with respect to the x -axis, $-y(x, a)$, belongs to the basic set. Without loss of generality, we can choose the

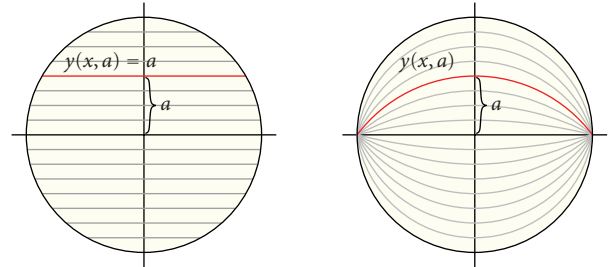


FIGURE 6: Basic set of paths defines a unique path for each value of parameter a . Paths on the left figure correspond to the shortest paths (i.e., straight line segments) and paths on the right correspond to the circular paths (see Example 7).

curve parameter a so that $y(0, a) = a$, $a \in [-1, 1]$. We make also the reasonable assumption of the type of paths that for $a \geq 0$, it holds that $0 \leq y(x, a) \leq y(0, a)$ for all x . Then a is the “height” of the curve. From these definitions, it follows that $y(x, -a) = -y(x, a)$ and also that $y(x, 0) = 0$, that is, the path corresponding to value $a = 0$ is the horizontal diagonal of the disk.

We assume that the curves in the basic set fill the unit disk completely so that each interior point of the disk belongs to one and only one path in the basic set, see Figure 6 for illustration. From the basic set of paths, the full set of paths is obtained by rotations of the whole set around origin by an angle in the range $[0, \pi]$. In the full set of paths, there is a unique path through any given point in any given

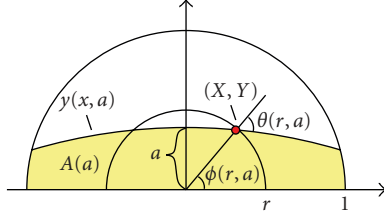


FIGURE 7: Notation for basic paths.

direction (see Figure 9, for an example for a full set of paths going through a given point).

Some additional notation needs to be introduced. Partial derivatives are denoted as

$$y_x(x, a) = \partial_x y(x, a) = \frac{\partial}{\partial x} y(x, a), \quad (50)$$

$$y_a(x, a) = \partial_a y(x, a) = \frac{\partial}{\partial a} y(x, a).$$

$X(r, a)$, $a \leq r$, is defined as the positive x -coordinate of the intersection point of the a -path $y(x, a)$ and the circle with radius r , that is, the positive solution x of the following equation¹:

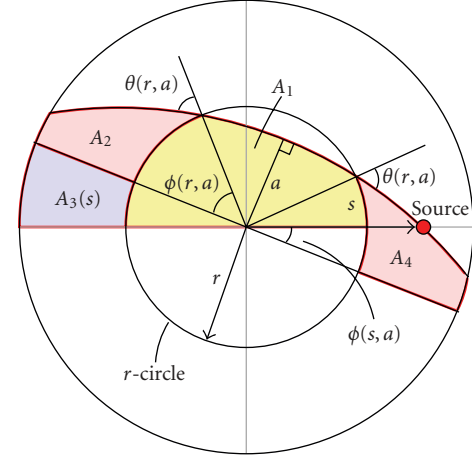
$$x^2 + y(x, a)^2 = r^2. \quad (51)$$

The corresponding y -coordinate of the intersection point is denoted as $Y(r, a) = y(X(r, a), a)$. The angle between the vector to this point and the x -axis is denoted by $\phi(r, a)$,

$$\phi(r, a) = \arctan \frac{Y(r, a)}{X(r, a)}. \quad (52)$$

Finally, the angle of incidence of curve $y(x, a)$ and r -circle is denoted by $\theta(r, a)$, that is, this is the angle between the tangent of the curve and the normal of the circle at the point of intersection. See Figure 7 for the illustration of these definitions.

In order to calculate the scalar flux $\Phi(r)$, we start by considering the contribution from a source point at distance $s \geq r$ from the origin (see Figure 8). Instead of focusing on a given destination point and trying to determine the angular flux at that particular point, we can consider the contribution of the source point to the flux at *any* point on the circle with radius r . So in the first step, we calculate the total flow $I(r, a; s)$ from the source point across the circle along the paths with parameter less than or equal to a . By symmetry, this flow is the same for all source points at distance s and the total contribution from all source points within an annulus with radius in the range $(s, s + \Delta s)$ is $2\pi s \Delta s I(r, a; s)$. Having summed the flows from all the sources within an annulus, the resulting flow across the r -circle is symmetric and

FIGURE 8: Calculating the total traffic flow from a source point at distance s from the origin crossing the r -circle.

the intensity of the flow at any point of the circle is obtained by dividing by the length of the circumference, $2\pi r$, resulting in intensity $I(r, a; s) s \Delta s / r$.

In the above discussion, we considered a partial intensity by restricting ourselves to paths with parameter less than or equal to a . This makes it possible to find the angular flux at distance r . By partial derivation with respect to a , we have that the intensity of flow, from sources in the annulus, across the circle along paths in the parameter range $(a, a + \Delta a)$ is $\partial_a I(r, a; s) s \Delta s \Delta a / r$. All these paths meet the r -circle at the incidence angle $\theta(r, a)$. By dividing the above expression by $\cos \theta(r, a)$, we get the angular flux (times the angle difference $\Delta \theta$ corresponding to the parameter difference Δa). This is so because, conversely, given angular flux $\varphi(\theta)$, the flow across the surface is given by $\int \varphi(\theta) \cos \theta d\theta$. Now, the scalar flux is obtained by integrating over all angles. In addition, we integrate over all source distances $r \leq s \leq 1$, yielding

$$\Phi(r) = \frac{1}{r} \int_0^r da \int_r^1 ds s \frac{\partial_a I(r, a; s)}{\cos \theta(r, a)}. \quad (53)$$

Next we focus on determining $I(r, a; s)$ and at the same time explain why the source point can be restricted to be outside the r -circle. As the total flow of the packets per second in the whole area is Λ , the source-destination density of flow (per unit area at the source and per unit area at the destination) is Λ / π^2 . Then the total flow from the source (per unit area at the source) across the circle along paths with parameter at most a is obtained by considering the “target area,”

$$I(r, a; s) = \frac{4\Lambda}{\pi^2} (A_1 + A_2 + A_3), \quad (54)$$

where A_1 , A_2 , and A_3 are the three shaded areas depicted in Figure 8. The factor 4 comes because, first, we have the same areas below the diagonal and, second, for areas A_2 and A_3 we have to take into account that the flow from the source crosses the circle twice, once in, once out (both times at the same angle of incidence). For area A_1 , we have to take into account that when restricting explicitly the source point to be

¹ It is assumed that there are only two solutions $\pm X(r, a)$ to this equation. This is not true, for instance, for strongly bell-shaped paths, for which the analysis is more complicated.

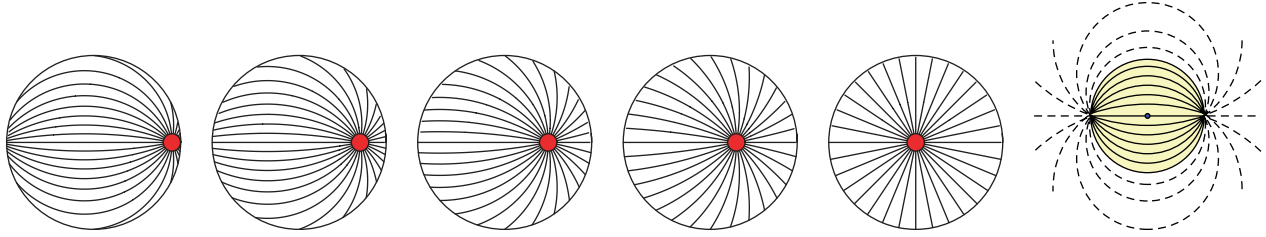


FIGURE 9: Circular paths are paths formed by the circumferences of circles which cross the unit disk at the opposite points.

outside the circle, we have neglected the equal flow from inside sources to outside, and this has to be compensated for by another factor of 2. For areas A_2 and A_3 , this further doubling is not needed, since the source point is let to be located at any point outside the circle, also in these areas.

The areas A_1 and A_2 are independent of s allowing us to make the s -integration in (53), $\int_r^1 s ds = (1/2)(1 - r^2)$. By inspection of Figure 8, the area A_3 is found to be $(1/2)(1 - r^2)\phi(s, a)$. The s -dependent factor $\phi(s, a)$ can also now be integrated: $\int_r^1 s \phi(s, a) ds = A_4$, where A_4 is the rightmost shaded area in Figure 8. From the figure, we further see that total area of A_1 , A_2 , and A_4 equals the area between the a -curve and the corresponding diagonal, denoted by $A(a)$ in Figure 7. Collecting the above pieces together, we finally end up with the simple result

$$\Phi(r) = \frac{2\Lambda}{\pi^2} \frac{1-r^2}{r} \int_0^r \frac{A'(a)}{\cos \theta(r, a)} da, \quad (55)$$

where

$$\begin{aligned} A'(a) &= \int_{-X(1,a)}^{X(1,a)} y_a(x, a) dx, \\ \cos \theta(r, a) &= \frac{X(r, a) + Y(r, a) y_x(X(r, a), a)}{r \sqrt{1 + y_x(X(r, a), a)^2}}. \end{aligned} \quad (56)$$

The former is obvious, and the latter follows, upon applying a trigonometric identity, from the observation that $\theta(r, a)$ is the difference between the angle $\phi(r, a) = \arctan Y(r, a)/X(r, a)$ and the (negative) angle of slope, $\arctan y_x(X(r, a), a)$, of the tangent of the a -path at $x = X(r, a)$, see Figure 7.

Because of the factor $\cos \theta(r, a)$ in the denominator, the integrand of (55) has a singularity at the upper limit of integration $a = r$, where $\cos \theta(r, r) = 0$. This is, however, an unessential singularity meaning that the integral is convergent. It may still cause some problems in numerical integration. The problems can be avoided by a simple change of variable of integration from a to α defined by $a = r \cos \alpha$, $\alpha \in [0, \pi/2]$.

As a check, consider the flux resulting from the use of shortest paths, that is, straight lines. Then we have $\theta(r, a) = \phi(r, a)$ and $r \cos \theta(r, a) = \sqrt{r^2 - a^2}$. It also holds that $A'(a) = 2\sqrt{1 - a^2}$. Using (55) and the above change of variable, $a = r \cos \alpha$, (55) is rederived.

By a limit consideration,² an even simpler expression can be derived from (55) for the flux at the centre,

$$\Phi(0) = \frac{\Lambda}{\pi} A'(0) = \frac{\Lambda}{\pi} \int_{-1}^1 \lim_{a \rightarrow 0} \frac{y(x, a)}{a} dx. \quad (57)$$

The integrand in the latter form is a “very low a ”-curve normalized so that the normalized curve has the height 1.

Example 7 (circular paths). As a first example, we consider a set of curvilinear paths, referred to as *circular paths*, which consist of such sections of circumference of circles (with radius ≥ 1) that cut the unit disk at the opposite points as illustrated in Figure 9 (see also Figure 6). From the figure, it can be seen that these paths smoothly move some portion of the traffic away from the centre of the disk. In passing, we note that the circular paths belong to the family of optical paths, and are obtained with the index of refraction profile

$$n(\mathbf{r}) = \frac{n(0)}{1 + r^2}. \quad (58)$$

Additionally, there is an analogy between the circular paths and electrostatics. The circular paths can be interpreted as electrical field lines of the 2D field between two line charges (perpendicular to the plane of the figure).

The equation for the basic set of circular paths is

$$y_{\text{circ}}(x, a) = \sqrt{(1 - x^2) + \left(\frac{1 - a^2}{2a}\right)^2} - \frac{1 - a^2}{2a}. \quad (59)$$

For $a = 1$, the function is $\sqrt{1 - x^2}$, whereas for small a it is approximately $a(1 - x^2)$.

The scalar flux calculated using (55) is depicted in Figure 10 (the middle curve). It can be seen that the traffic load is fairly well distributed. The maximum flux is obtained at the centre of the disk, where the exact result given by (57) is

$$\Phi_{\text{circ}}(0) = \frac{4}{3\pi} \cdot \Lambda \approx \underline{0.424 \cdot \Lambda}. \quad (60)$$

² When $r \rightarrow 0$ and $a \in [0, r]$, we have $A'(a) \rightarrow A'(0)$. Further, all basic paths inside the r -circle tend to straight horizontal lines and $r \cos \theta(r, a) \rightarrow \sqrt{r^2 - a^2}$. The integral can then be done, $\int_0^r da / \sqrt{r^2 - a^2} = \pi/2$.

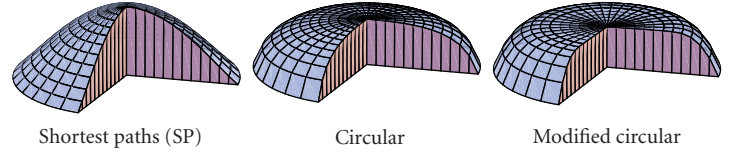
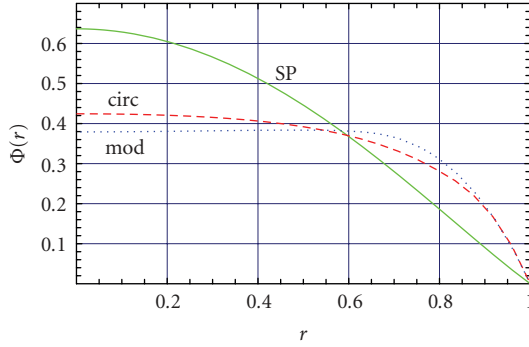


FIGURE 10: Scalar flux as the function of the radius for shortest paths, circular paths, and two-parameter modified circular paths with optimized parameters ($\Lambda = 1$).

This is precisely $2/3$ of the scalar flux with the shortest paths (cf., Example 3) and is also smaller than the maximal scalar fluxes with the ring paths. The factor $2/3$ simply follows from the fact that the area under the parabola $y = 1 - x^2$, $x \in [-1, 1]$, is $2/3$ of the area below the line $y = 1$ (and above the x -axis), that is, the setting of Figure 6 in the limit $a \rightarrow 0$.

Example 8 (modified circular paths). From (57), one sees that the flux at the centre can be made arbitrarily small by changing the shape of small- a paths to a bell shape. The area under such a bell curve (normalized to have the height 1) is the smaller the sharper the bell is. Of course, the flux at the centre can be made very small only at the expense of making it larger somewhere else; this is exactly the tradeoff we are trying to balance. To this end, we modify the basic curves as follows:

$$a \left(\frac{1}{a} y_{\text{circ}}(x, a) \right)^{\beta + (1-\beta)a}, \quad (61)$$

which for small a indeed makes the curve more bell-shaped, $a(1 - x^2)^\beta$ (when $\beta > 1$), while leaving the outer curves $a \approx 1$ untouched (the exponent is close to 1). In order to control in more detail how the exponent changes from β to 1 when a varies from 0 to 1, we change the expression further by introducing another tunable parameter γ as follows:

$$y_{\text{mod}}(x, a | \beta, \gamma) = a \left(\frac{1}{a} y_{\text{circ}}(x, a) \right)^{\beta + (1-\beta)a^{\gamma + (1-\gamma)a}}. \quad (62)$$

In principle, the exponent of a in the exponent could simply be γ but we found the present slightly more complicated form to work better.

With this two-parameter (β, γ)-family of paths, we can again numerically calculate the scalar flux $\Phi(r)$ using (55). The parameters can even be optimized in order to minimize the maximum flux. The lowest maximum flux

$$\min_{\beta, \gamma} \max_r \Phi_{\text{mod}}(r) \approx \underline{0.384} \cdot \Lambda \quad (63)$$

was obtained approximately at $\beta = 1.45$ and $\gamma = 12.2$. The basic path set for these optimal parameters is shown in Figure 11. Visually, the paths are very similar to the circular

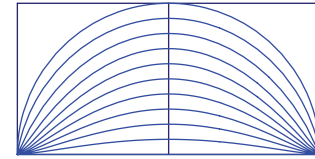


FIGURE 11: Modified circular paths with the optimized parameters $\beta = 1.45$ and $\gamma = 12.2$.

ones but one can distinguish the slightly bell-shaped form of the lowermost curves.

The corresponding flux as a function of radius is shown in Figure 10 (the lowest curve) and is compared with similar curves for shortest paths and (unmodified) circular paths. The flux distribution with the modified circular paths is remarkably flat and probably cannot be much improved with any other family of paths. It can be conjectured that with optimal paths, the flux is constant up to a certain distance and then falls to zero. This kind of conjecture is supported by the well-known behavior of optimal load balancing in finite networks obtained by solving an LP problem: typically the links in the center of the network are constraining, realizing the same maximum utilization, while links at the outer parts are not, and in fact the solution is not unique.

6.3. Randomized path selection approach

One option to achieve a lower maximum load is to allow the use of several paths for each pair of nodes (similarly as in [6, 7]). To this end, let us relax our assumptions and allow a finite number of path sets $\{\mathcal{P}_i\}$, where $i = 1, \dots, n$. Upon transmission of a packet, the source node chooses a path from path set \mathcal{P}_i with probability of p_i , $i = 1, \dots, n$.

Remark 10 (packet flux with randomized path sets). Randomized path selection upon transmission from path sets $\{\mathcal{P}_i\}$ with probabilities p_i , $i = 1, \dots, n$, yields a scalar packet flux of

$$\Phi(\mathbf{r}) = \sum_i p_i \cdot \Phi(\mathcal{P}_i, \mathbf{r}). \quad (64)$$

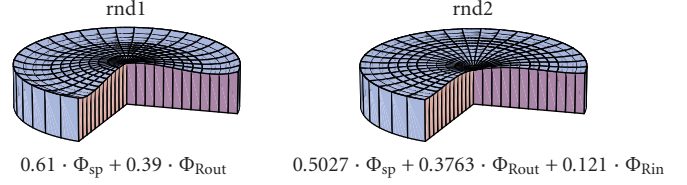
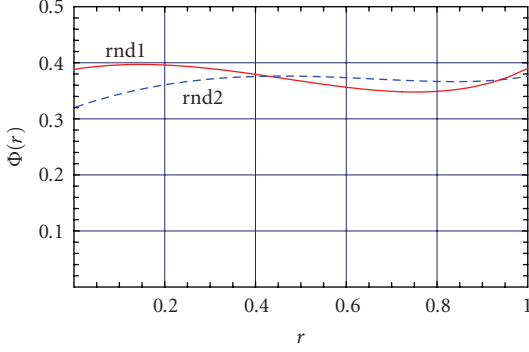


FIGURE 12: Scalar flux as a function of the radius r for the two elementary randomized path sets rnd1 and rnd2, which are obtained by relaxing the assumption of single-path routing ($\Lambda = 1$).

Example 9. Consider uniform traffic demand density in unit disk and two elementary path sets: (1) shortest paths, and (2) the outer radial paths. Weights $p_1 = 0.61$ and $p_2 = 0.39$ give a scalar packet flux of

$$\Phi(r) = 0.61 \cdot \Phi_{\text{sp}}(r) + 0.39 \cdot \Phi_{\text{Rout}}(r). \quad (65)$$

The resulting flux is rather constant as illustrated in Figure 12 with label “rnd1.” The maximum is $0.397 \cdot \Lambda$. The same technique can be taken further, for example, by combining all three elementary path sets as follows:

$$\Phi(r) = 0.5027 \cdot \Phi_{\text{sp}}(r) + 0.3763 \cdot \Phi_{\text{Rout}}(r) + 0.121 \cdot \Phi_{\text{Rin}}(r), \quad (66)$$

which gives a maximum flux of $0.3763 \cdot \Lambda$ corresponding to $\Phi_{\text{Rout}}(r)$ at the circumference (see curve with label “rnd2” in Figure 12). Similarly, the results with circular and modified circular paths can be slightly improved by moving a fraction of traffic to Rout paths.

Remark 10 may have one interesting application. First we note that as a single path between any source-destination pair is a special case of the randomized path selection, the optimal solution to the latter problem can never be worse. For the uniform traffic pattern in unit disk, we made a conjecture that the scalar flux obtained with an optimal (basic) set of paths is a constant up to some distance r^* and then decreases to zero which is achieved at $r = 1$, that is, the scalar flux would be a concave function of r . With Rout paths, the scalar flux is zero at $r = 0$ and then a strictly increasing convex function reaching a value 1 at $r = 1$. Thus, if the distance at which the fluxes of these two path sets are equal is strictly larger than r^* , then the maximum scalar flux can be further lowered by moving a small portion of traffic to Rout paths. In particular, this would mean that by using multiple paths a higher relative increase in traffic demands could be sustained than with a single-path routing.

6.4. Discussion

In general, deciding on the routes involves considering several factors and is not a straightforward task. In fact, often

it may be sufficient to simply use the shortest paths. In this paper, we have focused on the problem of load balancing, where, instead of using shortest paths, part of the traffic is deliberately routed along slightly longer paths in order to reduce the load in the most highly congested links. In our context of dense multihop networks, this translates to minimizing the maximum scalar flux, that is, finding such a set of paths which allows a maximal increase in traffic (with a given traffic pattern) the network can sustain.

This, however, has several unfavorable effects at times when the traffic load is low. Firstly, as the mean number of hops increases, the round-trip times become higher. Secondly, the higher mean number of hops also leads to a higher energy consumption, which can be an important factor, for example, for battery-powered wireless multihop networks.

In other words, there is a tradeoff between the mean path length (corresponding to delay and energy consumption in lightly loaded network) and the maximum sustainable traffic intensity with a given traffic pattern. In particular, the shortest paths represent the optimal set of paths under lightly loaded network and the optimal load balanced paths allow the maximal increase with given traffic pattern.

These two criteria can be combined by giving arbitrary weights for both objectives. The optimal set of paths for each combined objective has some mean path length and maximum scalar flux, which can be represented as a point in $(\bar{\ell}, \max \Phi(\mathbf{r}))$ -space. These points are Pareto optimal and form a concave curve with endpoints corresponding to the shortest paths and to the optimal load balanced paths.

In order to illustrate this, in Figure 13 we have plotted the points corresponding to the different path sets for unit disk considered earlier together with two lower bounds. The x -axis corresponds to the mean path length $\bar{\ell}$, which, according to (8), can be obtained from the scalar flux,

$$\bar{\ell} = \frac{1}{\Lambda} \int_{\mathcal{A}} \Phi(\mathbf{r}) d^2 \mathbf{r} = \frac{A}{\Lambda} \cdot \text{mean } \Phi(\mathbf{r}), \quad (67)$$

and the y -axis to the maximum scalar flux,

$$\max_{\mathbf{r} \in \mathcal{A}} \Phi(\mathbf{r}). \quad (68)$$

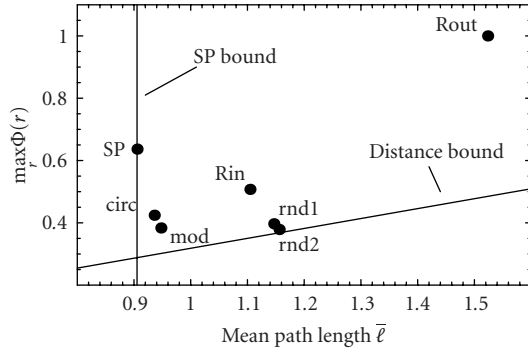


FIGURE 13: Comparison between the mean path length (i.e., overall forwarding load in the network) and the maximum scalar flux (i.e., traffic load) for different path sets in unit disk uniform traffic demands ($\Lambda = 1$).

By definition, no path set yields a lower mean path length than the shortest paths which gives a lower bound for the mean path length denoted by “SP bound.” The distance bound is given by (7).

From the figure, it can be seen that the radial ring paths (Rin and Rout) clearly are not even close to Pareto optimal while the other three path sets (SP, circular, and modified circular) can be justified with different objectives or constraints. Furthermore, the randomized path sets (rnd1 and rnd2) obtained by combining the shortest paths and radial ring path(s) achieve a low maximum scalar packet flux, but at the same time increase unnecessarily the mean path length. This is due to the use of Rout paths to move a portion of traffic away from the center. The fact that they are close to the distance bound is due to a rather constant scalar packet flux as illustrated in Figure 12.

7. CONCLUSIONS

In this paper, we have presented a general framework for analyzing traffic load and routing in a large dense multihop network. The approach relies on strong separation of spatial scales between the microscopic level, corresponding to the node and its immediate neighbors, and the macroscopic level, corresponding to the path from the source to the destination. In a dense wireless network with this property the local traffic load can be assimilated with the so-called scalar (packet) flux. The scalar flux is bounded by a maximal value that the network with a given MAC and packet forwarding protocol can sustain. The scalar flux depends on traffic demand density $\lambda(\mathbf{r}_1, \mathbf{r}_2)$ and the chosen set of routing paths \mathcal{P} . The load balancing problem thus comprises determining the set of routing paths such that the maximal value of the flux in the network is minimized. While the general solution of this difficult problem remains for future work, our main contribution in this paper consists of giving bounds for the scalar flux and giving a general expression for determining the scalar flux at a given point for a given set of curvilinear paths.

A particular attention was given to the special case of unit disk with uniform traffic demands for which we have derived a simple computationally efficient expression for calculating the scalar flux for any family of paths. In this case, we were able to reduce the general three-dimensional integral equation to a two-dimensional one, which is both numerically stable and convenient to evaluate.

These results are illustrated by numerical examples with different heuristically chosen sets of paths, and also by optimizing a parameterized set of paths. In particular, as a result of optimization, we have found a set of paths with a remarkably flat scalar flux distribution and the maximum scalar flux reduced by about 40% when compared to the shortest paths. In this paper, we have limited our attention to specific types of paths satisfying the so-called path continuity condition. This may be an unnecessarily restricting requirement and one may be able to further reduce the maximum scalar flux by relaxing this assumption. This is a topic for further study.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for valuable comments. For the part of E. Hyytiä, this work has been performed partially in the Telecommunications Research Center Vienna (ftw.) in the framework of the Austrian Kplus Competence Centre Programme, and partially in the “Centre for Quantifiable Quality of Service in Communication Systems (Q2S), Centre of Excellence” appointed by The Research Council of Norway and funded by the Research Council, NTNU, and UNINETT. For the part of J. Virtamo, this work was performed in the project Fancy funded by the Academy of Finland (Grant no. 210275).

REFERENCES

- [1] P. Jacquet, “Geometry of information propagation in massively dense ad hoc networks,” in *Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '04)*, pp. 157–162, Roppongi Hills, Tokyo, Japan, May 2004.
- [2] S. Toumpis and L. Tassiulas, “Optimal deployment of large wireless sensor networks,” *IEEE Transactions on Information Theory*, vol. 52, no. 7, pp. 2935–2953, 2006.
- [3] S. Toumpis, “Optimal design and operation of massively dense wireless networks,” in *Proceedings of Workshop on Interdisciplinary Systems Approach in Performance Evaluation and Design of Computer & Communication Systems (INTERPERF '06)*, Pisa, Italy, October 2006.
- [4] E. Hyytiä and J. Virtamo, “On load balancing in a dense wireless multihop network,” in *Proceedings of the 2nd Conference on Next Generation Internet Design and Engineering (NGI '06)*, pp. 72–79, València, Spain, April 2006.
- [5] B. Sirkeci-Mergen and A. Scaglione, “A continuum approach to dense wireless networks with cooperation,” in *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '05)*, vol. 4, pp. 2755–2763, Miami, Fla, USA, March 2005.
- [6] P. P. Pham and S. Perreau, “Performance analysis of reactive shortest path and multi-path routing mechanism with load balance,” in *Proceedings of the 22nd Annual Joint Conference*

- of the *IEEE Computer and Communications Societies (INFOCOM '03)*, vol. 1, pp. 251–259, San Francisco, Calif, USA, March–April 2003.
- [7] Y. Ganjali and A. Keshavarzian, “Load balancing in ad hoc networks: single-path routing vs. multi-path routing,” in *Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '04)*, vol. 2, pp. 1120–1125, Hong Kong, March 2004.
- [8] O. Dousse, F. Baccelli, and P. Thiran, “Impact of interferences on connectivity in ad hoc networks,” *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 425–436, 2005.
- [9] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, 2000.
- [10] M. Kalantari and M. Shayman, “Routing in wireless ad hoc networks by analogy to electrostatic theory,” in *Proceedings of IEEE International Conference on Communications (ICC '04)*, vol. 7, pp. 4028–4033, Paris, France, June 2004.
- [11] D. B. Johnson and D. A. Maltz, “Dynamic source routing in ad hoc wireless networks,” in *Mobile Computing*, vol. 353, chapter 5, pp. 153–181, Kluwer Academic, Dordrecht, The Netherlands, 1996.
- [12] C. Bettstetter and C. Wagner, “The spatial node distribution of the random waypoint mobility model,” in *Proceedings of German Workshop on Mobile Ad Hoc Networks (WMAN '02)*, pp. 41–58, Ulm, Germany, March 2002.
- [13] C. Bettstetter, G. Resta, and P. Santi, “The node distribution of the random waypoint mobility model for wireless ad hoc networks,” *IEEE Transactions on Mobile Computing*, vol. 2, no. 3, pp. 257–269, 2003.
- [14] W. Navidi and T. Camp, “Stationary distributions for the random waypoint mobility model,” *IEEE Transactions on Mobile Computing*, vol. 3, no. 1, pp. 99–108, 2004.
- [15] E. Hyttiä and J. Virtamo, “Random waypoint mobility model in cellular networks,” *Wireless Networks*, vol. 13, no. 2, pp. 177–188, 2007.
- [16] G. I. Bell and S. Glasstone, *Nuclear Reactor Theory*, Van Nostrand Reinhold, New York, NY, USA, 1970.
- [17] E. Hyttiä, P. Lassila, and J. Virtamo, “Spatial node distribution of the random waypoint mobility model with applications,” *IEEE Transactions on Mobile Computing*, vol. 5, no. 6, pp. 680–694, 2006.