On Asymptotically Optimal Routing in Large Wireless Networks and Geometrical Optics Analogy

Roberto Catanuto, Stavros Toumpis, Giacomo Morabito

Abstract—This work is based on the observation that, as the number of nodes in a wireless network approaches infinity, minimum-cost routes become smooth curves that observe the same laws followed by rays of light in properly defined optical media. Accordingly, in this paper an analogy between optimal routing in large wireless networks and geometrical optics is first formally defined. The analogy is based on the concept of the cost function, which plays the role of the refractive index in the networking context. Then, the relevance of the principle of Fermat and the eikonal equation in routing problems is shown, and a methodology for calculating the cost function is proposed and applied in two cases of special interest, i.e., bandwidth-limited and energy-limited networks. The applicability of the Optics-Networking analogy is also discussed in the case of networks with large but finite numbers of nodes. Finally, novel, distributed route discovery protocols that make use of the analogy are outlined.

Index Terms—Geometrical Optics, Massively dense networks, Multihop, Routing, Trajectory Based Forwarding, Wireless networks.

I. INTRODUCTION

A. Routing in Large Wireless Networks

We study routing in wireless networks, focusing on networks with a very large number of nodes. We are motivated by the fact that many currently deployed and envisioned wireless sensor networks [4], mesh networks [5], and vehicular networks [6] have thousands of nodes. (For example, the Athens Metropolitan Wireless Network currently has close to 3000 nodes, spread over the whole Athens metropolitan area [7].)

Routing protocols designed for use in this regime must be scalable, i.e., as the number of nodes increases, the routing overhead must increase gracefully [8]. A routing paradigm that is well-suited for use in large wireless networks is position-centric (or geographic) routing, e.g., [9], [10], and [11]. The key idea is to use knowledge of the location of the destination, and route packets toward its general direction. With this approach, the delivery of packets is not affected by changes in the topology of the network between the source and the destination, which become more frequent as the network size increases. Therefore, this routing paradigm is scalable. However, nodes need to know the physical location of their destination nodes.

A recent refinement to position-centric routing, which is of particular relevance to our work, is Trajectory Based Forwarding (TBF) [12]. Packets are routed along a trajectory, which is defined as a curve connecting the source and the destination. The trajectory does not have to be a straight line, but can have any arbitrary shape, depending on the conditions in the network and the needs of the packet. Whenever the packet arrives at an intermediate node, the next node is chosen so that the packet progresses as much as possible along the trajectory and it remains as close to the trajectory as possible. Under these twin objectives, various rules for selecting the next hop have been proposed [13].

An important issue in this line of research is choosing the trajectory. In particular, the nodes should avoid specifying trajectories in a greedy manner that leads to low costs in the first hops, but to inescapably much higher costs as the trajectory reaches the destination.

B. Contributions

When the number of nodes is very large, the complexity of performing route optimization while maintaining a complete description of the network becomes high, due to the number of parameters involved (such as node locations, channel conditions, etc.) that must be taken into account. Therefore, we make a drastic compromise and adopt a simplified, macroscopic view of the network. The idea for the use of a macroscopic view was recently introduced in a number of independent works [14], [15], and has since been used to yield a wealth of results (see the overview work [16] and the numerous references within).

The macroscopic view is not as detailed as the usual microscopic one, that includes all the details of the network, but still retains enough details to permit a meaningful optimization, that captures the salient features of the optimal routes when the number of nodes is very large. Technically, the suppression of details is achieved by assuming that the network consists of an infinite number of nodes. Therefore, routes are not treated as sequences of nodes used for routing, but as smooth curves connecting the location of the source with the location of the destination. Key to the macroscopic view is the cost function $c(r)$, defined in detail in Section II-A, which captures the cost of transporting a packet through the location $r$.

Within this framework, we study three problems. The first one is the Unicast problem: we are given two locations $A$ and $B$ and we want to find the route that connects the two and incurs the minimum cost possible. The second one is the Anycast problem: we are given a region $\Omega$ and a location $B$, with
and we are asked to determine the minimum cost route that connects $B$ with any point in $\Omega$. The third is the Broadcast problem, in which we are given a region $\Omega$ and we want to determine the routes that must be used to broadcast a given data packet that is already available to the nodes in the region $\Omega$ to all other locations in the network with the minimum total cost.

To solve these problems in the macroscopic regime, we note that a useful analogy exists between the optimal propagation of packets in the network and the propagation of light in Geometrical Optics. For the Unicast problem, we show in Section II-B that the optimal route coincides with a ray of light connecting points $A$ and $B$ if we replace the network with an optical medium of refractive index $n(r) = c(r)$. For the Anycast and Broadcast problems, we show in Sections II-C and II-D that we can determine the optimal routes again by replacing the network with an optical medium of refractive index $n(r) = c(r)$ and solving there the central equation of Geometrical Optics, i.e., the eikonal equation.

To calculate the cost function, we develop in Section III a methodology that uses microscopic aspects of the network, such as the models for the node placement and link costs, and hinges on the fact that the number of nodes in the network is very large. We employ this methodology in two cases of particular interest: bandwidth limited networks and energy limited networks. However, the methodology is general enough to be adaptable to many other cases.

The routes specified by this analogy are asymptotically optimal, i.e., they represent the limit of the optimal routes (as these are determined, for example, by dynamic programming or other traditional approaches) as the number of nodes goes to infinity, and routes start to resemble smooth curves. In networks with a finite number of nodes, optimal routes will not coincide with their optical limit. As we will show in Section IV, however, they are approximated well by the optical limit, when the number of nodes is large, but reasonable so.

Despite its conceptual and practical significance, the analogy does not readily yield a simple method for discovering the asymptotically optimal route connecting two locations. This motivates Section V, where we delineate practical route discovery protocols.

Concluding remarks and directions for future work are offered in Section VI. In the Appendix, there is a short review of Geometrical Optics, focusing on its aspects that are pertinent in our case.

We stress that an Optics-Networking analogy appeared first in the work in [15], which has inspired this work. There, the author considers a network that consists of infinite nodes (termed there massively dense) and calculates the optimal unicast route between two points by analogy to Optics. With respect to [15], our work innovates in the following: firstly, we explicitly define the cost function, we develop a methodology for calculating it, and we apply this methodology in two important network types, bandwidth-limited and energy-limited. Secondly, we investigate the analogy to a greater extent, notably showing how the eikonal equation can be applied. Thirdly, we briefly study the convergence of optimal routes to the Optical limit, and finally we outline route discovery protocols inspired by the analogy.

II. THE OPTICS/ROUTING ANALOGY

A. Cost Function

Our network consists of nodes placed on a 2-dimensional domain $D$. We assume that there are so many nodes, that the typical distance between nearest neighbors is incremental with respect to the dimensions of $D$, and a route connecting two disparate locations $A$ and $B$ in the network resembles a smooth curve $R$ starting at $A$ and ending at $B$.

We associate with the movement of packets over the network a consumption of resources which is cumulative with the distance over which the packet is transported. The resources could be the bandwidth used, the energy wasted, the time needed for the transport, or some other composite of the above. Specific cases will be considered in Section III.

We model the use of resources in terms of a scalar cost function $c(r)$, which is unit-less, and is defined as follows: Let $\epsilon$ be the incremental length of a small line segment centered at $r$. Let $dc(\epsilon, r)$ be the incremental cost associated with transporting a packet along this incremental line segment. Let $c_{\text{nominal}} \times \epsilon$ be the cost incurred by transporting the packet in some given uniform nominal network by a distance $\epsilon$. The cost function is defined as follows:

$$c(r) = \lim_{\epsilon \to 0} \frac{dc(\epsilon, r)}{c_{\text{nominal}} \times \epsilon}. \quad (1)$$

For example, a cost function $c(r) = 2$ means that it is twice as costly to move packets across the location $r$ of our network with respect to the nominal network.

We interpret our assumption that the network consists of a very large number of nodes to mean that the limit is achieved for lengths $\epsilon$ that, though small, correspond to a large number of hops. We expound on this point in Section III, where we use it to calculate the cost function in two settings.

Using (1), it follows that the cost to move a packet from point $A$ to point $B$, along a curve $L$ of non-incremental length, is given by the line integral of the cost function along $L$:

$$[AB]_L = \int_A^B c(r) \, dr. \quad (2)$$

Defining the cost to be unit-less may seem counter-intuitive. Indeed, it would make more sense to simply define the cost function as the incremental cost over the incremental distance, without dividing by $c_{\text{nominal}}$. The advantages of our approach are two-fold: firstly, in this manner the cost function is unit-less, and so it is equally applicable in a wide variety of settings. Secondly, we will soon show an analogy of our setting with Geometrical Optics, where the nominal network is mapped to free space and the cost function is mapped to the refractive index, which is also unit-less.

B. Unicast Routing

Let $A$ and $B$ be the locations of two nodes that need to exchange data packets. The nodes belong to a network with a known cost function $c(r)$. We would like to determine the optimal route $L$, which connects the two nodes with
the smallest possible routing cost. This problem involves
the minimization of the line integral (2) with respect to all
continuous curves $L$ connecting the two points $A$ and $B,$
and can be solved from first principles, using Calculus of
Variations [17]. However, a close analogy exists between our
problem and Geometrical Optics, that we can readily use to
to fully describe optimal routes.

In particular, optical media are described in terms of the
refractive index $n(r)$: in the time it takes light to cover a line
segment of incremental length $\epsilon$, centered at $r$, in the
medium, light can cover a distance of $c\epsilon n(r)$ in free space. Networks, on
the other hand, are described in terms of the cost function $c(r)$:
with the cost it takes the network to transport the packet across
an incremental distance $\epsilon$, the packet could be transported
across a distance $c\epsilon$ in the nominal network. Therefore,
going from Optics to Networking, the optical medium becomes
the free space, the nominal network becomes the nominal network,
the refractive index becomes the cost function, and the optical
length of a curve (given by (9) in the Appendix) becomes the
cost to transport a packet along that curve, given by (2).

We can then solve the Unicast problem by framing it into
the context of Optics: Finding the route with the minimum cost
is equivalent to finding the curve with the minimum optical
length connecting two points $A$ and $B$ in an optical medium
with refractive index $n(r) = c(r)$. However, as discussed in
the Appendix, Fermat’s principle specifies that this curve is
actually a ray of light that starts from $A$ and goes to $B$.
Therefore, our optimal route resembles a light ray, and due
to this analogy we will henceforth refer to it as an optical
route.

The analogy permits the use of intuition and mathematical
tools that are already available, for the study of light rays, in
our context. In particular, it immediately shows that:

1) The optimal route satisfies the same differential equation
that rays of light must satisfy (equation (7) of the Appendix),
if we substitute $n(r)$ with $c(r)$. In particular, optimal routes
are described in terms of the cost function $c(r)$: a ray associated with it.

2) At network interfaces where the network cost changes
abruptly, the optimal route either bends or is totally reflected,
in a manner predicted by Snell’s law (equation (8) of the Appendix),
taking $n_1$ and $n_2$ to be the two costs on either side of the interface).

3) Our minimization problem may have many local minima,
each of which corresponds to a route that is locally optimal
(according to Fermat’s principle), but not necessarily globally
optimal.

C. Anycast Routing

We now move to the case where we want to route traffic
between any node $A$ within a subset of nodes $\Omega$ and a single
node $B$ in the network. We are free to choose $A$ among the
rest of the nodes in $\Omega$ so that the cost is minimized. Therefore,
both the shape of the route and one of its endpoints are subject
to minimization. For example, the set $\Omega$ might be the set of
the locations of all cluster heads of a wireless network using
clustering. In such a case, each node will want to communicate
with the cluster head for which the cost is minimum.

This problem can also be solved by invoking the Optics-
Networking analogy. In particular, we need the minimum costs,
and the routes that achieve them, between the region
$\Omega$ and all other points in the network. In the context of
Optics, we need to find the minimum optical lengths, and
their corresponding rays, between a distributed light source
occupying $\Omega$ and all other points, in an optical medium whose
refractive index is $n(r) = c(r)$. By definition, these minimum
optical lengths are given by the eikonal function which is a
scalar function of the position $r$ and represents the minimum
distance between the source and $r$. (A definition
and brief description of the eikonal function appears in the
Appendix.) Knowing the eikonal function, we can construct
the rays that connect the source with any part of the network.
As an example, the rays plotted in Fig. 13 of the Appendix can
be interpreted to be the optimal routes that connect any point
in the network with the distributed, circular set $\Omega$ centered
at the origin, in a network where the cost function equals
$c(x, y) = \sqrt{x^2 + y^2} + 1$.

A complication arises from the fact that the eikonal function
can have multiple values at a particular location $r$. In
the context of Optics, all of the values carry equal physical signif-
icance: each of them represents the time at which a distinct
ray arrives at the location $r$. In the context of Networking,
each of the values represents the cost of transmitting a packet
along a route that is locally optimal, but not necessarily
globally optimal. Therefore, the analogy between Optics and
Networking breaks down; whereas in Optics we are equally
interested in all values that the eikonal achieves in a location,
in Networking we are interested in the smallest value, and the
ray associated with it.

For example, let us consider a network with a cost function
$c(x, y) = 1 + \exp[-0.1 \sqrt{x^2 + (y - 10)^2}]$. (Such a cost function
might be appropriate if, for example, there is a local traffic
bottleneck in the region centered at the location (0 m, 10 m),
and we would like our packets to steer clear of that region.)
As shown in Fig. 14, in part of the network the eikonal function
has three values. The largest two correspond to the cost
accumulated if the packets follow routes that are locally
optimal but globally suboptimal. In Fig. 15 we have redrawn
the eikonal, keeping only the smallest value. This figure shows
the cost of the source sending a packet to any point in the
network using the globally optimal route.

D. Broadcast Routing

Finally, we consider the problem of determining the optimal
routes with which a data packet can be broadcast to the whole
network. More formally, let $\Omega \subset \mathcal{D}$ contain a subset of the
nodes in the network. If these nodes hold a data packet that
must be received by all other nodes in the network, what is the

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Optics & Networking & refractive index & cost over that curve & nominal network \\
\hline
medium & wireless network & cost function & cost over that curve & free space \\
\hline
\end{tabular}
\caption{Optics Quantities and their Networking Counterparts.}
\end{table}
optimal manner with which the packet must propagate through the network, in order to minimize the total routing cost?

Note that this problem is different from the Anycast Problem, as in that case the set $\Omega$ has a different data packet for each location $B$. Now, there is a single data packet, common to all nodes in $\Omega$, which we want to deliver throughout the network. As an example, this problem is pertinent in a wireless sensor network where a distributed set of sources must issue a network-wide request for data of a particular type.

To show the analogy of this case with Optics, we reason as follows: let $S(\cdot)$ be the function that specifies the cumulative cost $S(r)$ to send the data packet to location $r$ under the optimal broadcasting scheme. It is intuitively clear that $S(\cdot)$ is continuous. Indeed, if it were discontinuous at a location $r_0$, then there would be two routes $R_1$ and $R_2$ intersecting at $r_0$ with cumulative costs $c_1$, $c_2$, and $c_1 < c_2$. By changing the routing so that some of the nodes on $R_2$ receive the data packet through $R_1$, we can reduce the total cost and hence arrive at a contradiction. Therefore, $S(\cdot)$ must be continuous.

Let us focus on an arbitrary point $r_1$, where the cost function $c(\cdot)$ is also continuous\(^1\), and consider the loci $S(r) = S(r_1) \triangleq k$ and $S(r) = k + \phi$. If we set $\phi$ to be very small, and focus on a very small neighborhood around $r_1$, then, by the continuity of $S(\cdot)$ and $c(\cdot)$, as shown in Fig. 1, the two loci become straight lines, and the medium becomes uniform with constant cost $c(r_1)$. In this infinitesimal setting, the optimal propagation of packets clearly involves the transmission of packets from the locus $S(r) = S(r_1) = k$ to the locus $S(r) = k + \phi$, along straight lines normal to the loci. The accumulated cost along these infinitesimal lines is $\phi = c(r_1) \times ds$, where $ds$ is the distance separating the two loci. Therefore, $|\nabla S(r_1)| = \frac{\phi}{ds} = c(r_1)$. In other words, $S(\cdot)$ satisfies the eikonal equation, the same equation satisfied by light as it propagates away from its source.

To conclude, a data packet initially existing in a subset $\Omega$ and optimally propagating over the whole network resembles the light propagating from a distributed source $\Omega$, if the cost function is taken to be the refractive index of the medium. Therefore, the propagation of broadcast packets is identical to the propagation of packets in the Anycast Routing case. In practical terms, the source should launch packet copies toward all directions, and the copies should propagate through the network following the laws of optics. As an example, Fig. 13 shows how a packet originating in a circular source should be broadcast through the whole network.

As in the Anycast Routing case, a complication arises from the fact that the eikonal function can attain multiple values in a particular location. However, it is intuitively clear that in order to minimize the total routing cost, the packet should arrive at each point $r$ following the route that gives the minimum value of the eikonal function at that point. As an example, Fig. 15 could be interpreted as showing the optimal cumulative cost, in a broadcast setting, at each point of a network with cost function $c(x, y) = 1 + \exp[-0.1\sqrt{x^2 + (y - 10)^2}]$.

\(^1\)The non-continuous $c(\cdot)$ case is straightforward but tedious, and so is omitted.

---

**Fig. 1.** Setting for the derivation of the eikonal equation in the broadcast case.

### III. Cost Models

In the previous section we specified a macroscopic network model, built around the notion of an abstract cost function, which we left unspecified. In this section, we construct two microscopic models, specifying for each of them, among others, the model for node placement, the resources consumed by each transmission, and the rules of forwarding packets. We then calculate the macroscopic cost function that corresponds to each of these microscopic models.

#### A. Bandwidth Limited Networks

**Node Deployment:** We assume that the nodes are placed randomly, according to a spatial Poisson distribution [18], [19] with density $\lambda(r)$ measured in nodes per square meter.

**Transmission cost:** We assume that the nodes communicate with each other through a common wireless channel of finite bandwidth. When a node $T$ transmits to a node $R$, nodes in the vicinity must refrain from transmitting, in order to avoid collisions. (If the bandwidth was infinite, no such restriction would exist, as each transmitter-receiver pair could use its own dedicated bandwidth.)

In this setting, it is reasonable to assume that each transmission over a distance $d$ consumes an area proportional to $d^2$. An intuitive justification for this is as follows: If a node $T$ transmits to a node $R$ over some distance $d$, then it is necessary for all other potential transmitters that are in a neighborhood of the receiver to refrain from transmitting, in order not to overpower the signal. Intuitively, the radius of this neighborhood is proportional to the distance $d$. Therefore, the area of the neighborhood is $kd^2$, where $k$ is a unit-less constant that depends on the sensitivity of the receiver, the transmission rate used, etc.

Our heuristic argument clearly applies to a wide range of settings. There are also specific settings where it can be made rigorous. For example, let us assume that in particular the protocol model of [20] holds: A node $X_i$ can successfully send a packet to node $X_j$ as long as $|X_i - X_j|/(1 + \Delta) \leq |X_k - X_j|$, for all other nodes $X_k$ transmitting simultaneously. The parameter $\Delta > 0$ models the robustness of transmissions with respect to competing transmissions. Then, it is straightforward to show by use of the triangle inequality [20] that if a node $A$ is receiving a packet over a given distance $d$, then there can be no other receiver in the disk centered at the node $A$ and of radius $\frac{\Delta}{4}d$ and area $kd^2$ where $k = \frac{\pi d^2}{4}$. (Note that this bound is conservative, however it is sufficiently tight so that it leads to a tight upper bound in the transport capacity [20].)
resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in the Computer Networks Journal, vol. 53, no. 11, pp. 1939-1955, July 2009.

http://dx.doi.org/10.1016/j.comnet.2009.02.021

Clearly, as the areas consumed by communicating pairs in a network increase in size, the number of simultaneous transmissions decreases, as the total consumed area cannot exceed the total area of the network. In turn, this drives the aggregate throughput down. It would be reasonable for network designers to keep the size of the consumed areas as small as possible, in order to maximize throughput, and so it is reasonable to define the transmission cost of a transmission over a distance $d$ to be

$$c_t(d) = d^2,$$

Forming Rule: On the microscopic level, nodes use

Trajectory Based Forwarding (TBF) to select the next hop for a packet, using the macroscopic route of the packet as the trajectory. Therefore, TBF provides the link between the microscopic and the macroscopic sides of the network.

To complete the routing model, we must specify the rule according to which nodes forward packets along the route. Microscopic parts of routes can be viewed as straight lines, so it suffices to have a rule for forwarding packets along a given direction. With no loss of generality, let this direction be that of the positive $x$-axis.

Given that we have associated each transmission over a distance $d$ with a cost $c_t(d) = d^2$, a reasonable first choice for the node is to forward the packet to that of its neighbors closer to the destination (along the trajectory) for which the ratio

$$r \triangleq \frac{c_t(d)}{x_2 - x_1} = \frac{d^2}{x_2 - x_1}$$

is minimized. In the above, $d$ is the distance between the node and its neighbor, $x_1$ is the $x$-coordinate of the node, and $x_2$ is the $x$-coordinate of the neighbor. The difference $x_2 - x_1 > 0$ represents the progress along the trajectory toward the destination. Therefore, we would like to strike a balance between incurring a small transmission cost, but also achieving a substantial progress toward the destination [13]. We will refer to this forwarding rule as Rule A. In Fig. 2 we plot a sample route (route (a)) that results from its use.

As shown in the figure, a problem with Rule A is that it may cause the packet to deviate significantly from its trajectory, over the course of many hops. (In fact, it is intuitively clear that

$^2$We note that the proof for the upper bound of the capacity of a wireless network in [20] was based on exactly this idea.

the deviation from the trajectory can be modeled by a random walk where the steps take continuous values.) Therefore, we are motivated to consider the following modification: if the node forwarding the packet is on the left (right) of the trajectory and at some distance $v$ from it, then the next node to receive the packet cannot also be on the left (right) of the trajectory and at some distance $v'$ from it. The rest of the nodes are acceptable as next hops, and the one minimizing the cost-progress ratio should be chosen. An example route appears in Fig. 2 (route (b)). As the figure shows, the new rule, which we call Rule B, leads to the packets following the trajectory much more closely.

A disadvantage of Rule B with respect to Rule A is that we are excluding from our consideration neighbors with a very small associated cost-progress ratio, but who happen to increase the deviation from the ideal trajectory even very modestly. The observant reader will notice a few such examples in Fig. 2. Obviously, there is significant room for improving both Rule A and Rule B. As the question of choosing the next hop has already attracted significant research interest within the context of TBF and geographic routing (see for example [21] and the references therein), we do not pursue this issue further. In the following, we will assume that Rule B is used.

Forming Rule Performance: Next, we evaluate the performance of Rule B in terms of the expected progress and the expected cost per hop. We assume that the node having the packet is placed at the origin and that the trajectory is described by the equation $y = \text{const} > 0$. The progress is then simply the $x$-coordinate of the next node, and in order to have a positive progress toward the destination, the next hop must have a positive $x$-coordinate. In addition, in order to satisfy Rule B, it must also have a positive $y$-coordinate. Also, as we focus on a microscopic part of the network, we assume that the node density is constant, and simply denoted by $\lambda$.

As the placement of nodes is random, the location of the next node will also be random, and so its associated progress, $y$-coordinate, distance from the origin, cost, and cost-progress ratio are random variables, which we denote by $X, Y, D, C,$ and $R$ respectively. The setting is shown in Fig. 3.

The first step is to find the loci of constant cost-progress ratio. With straightforward algebra, it follows that:

$$\frac{d^2}{x} = r \iff (x - \frac{r}{2})^2 + y^2 = \left(\frac{r}{2}\right)^2.$$ 

Therefore, as shown in Fig. 4, the loci of constant ratio $r$ are semi-circles centered at point $\left(\frac{r}{2}, 0\right)$ with radius $\frac{r}{2}$. Note that as $r \to 0$, the loci implode toward the origin. Therefore, to
minimize the cost-progress ratio, the node would like to avoid transmitting to nodes further away. This is consistent with the long-known fact that many short hops are better than a few long ones, if the bandwidth is limited [20], [22].

Knowing the shape of the loci, we can calculate the distribution of the ratio $R$. The ratio will be larger than $r$ if and only if there is no node within a semi-circle of radius $\frac{r}{2}$. As the nodes follow a Poisson spatial distribution with density $\lambda$, this happens with probability $\exp\left[-\frac{\lambda\pi r^2}{8}\right]$. Therefore,

$$F_R(r) = 1 - e^{-\frac{\lambda\pi r^2}{8}} \iff f_R(r) = \frac{r\lambda\pi}{4} \exp\left[-\frac{\lambda\pi r^2}{8}\right].$$

So the ratio $R$ follows the Rayleigh distribution, with expectation $E[R] = \sqrt{\frac{2}{\pi}}$.

Next, we calculate the expectation $E[X]$ of the progress, using conditioning on $R$. In particular, let us assume that $R = r$. It then follows that the next node is on the circumference of the semi-circle centered at $(\frac{r}{2},0)$ and with radius $\frac{r}{2}$. By symmetry, it follows that the conditional expectation of its $x$-coordinate will be $E[X|R = r] = \frac{r}{2}$. Therefore:

$$E[X] = \int_0^\infty E[X|R = r] f_R(r) dr = \int_0^\infty \frac{r^2\lambda\pi}{8} \exp\left[-\frac{\lambda\pi r^2}{8}\right] dr = \frac{1}{\sqrt{2\lambda}}.$$ 

Finally, we calculate the expectation $E[C] = E[D^2]$ of the cost, again by conditioning on the value of the ratio $R$. For this, we need to calculate $E[D^2|R = r]$. We note that $D^2 = \frac{R^2}{2}(1 + \cos \Phi)$ where, as shown in Fig. 3, $\Phi$ is the angle formed by the positive $x$-axis and the line interval connecting the points $(X, Y)$ and $(\frac{R}{2}, 0)$. Given that $R = r$, by symmetry considerations the angle $\Phi$ is uniformly distributed in $(0, \pi)$. It follows that

$$E[D^2|R = r] = \frac{1}{\pi} \int_0^\pi \frac{r^2}{2} (1 + \cos \phi) d\phi = \frac{r^2}{2}.$$ 

Therefore,

$$E[D^2] = \int_0^\infty E[D^2|R = r] f_R(r) dr = \int_0^\infty \frac{r^3\lambda\pi}{8} \exp\left[-\frac{\lambda\pi r^2}{8}\right] dr = \frac{4}{\lambda\pi}.$$ 

**Cost Function:** Having determined the performance of the forwarding rule, we are now ready to specify the cost function. For this, let us assume that a packet following the forwarding rule gets transported across an incremental distance $\epsilon$. By our assumption that the number of nodes is very large, this distance is covered with a large number of hops. To calculate the associated incremental cost, we note that the sequence of progresses $\{X_i\}$ together with their associated costs $\{C_i\}$ form a reward renewal process, where $C_i$ is the reward for taking a step of size $X_i$. By the strong law of large numbers for reward renewal processes [18], it follows that

$$\frac{\sum_{i=1}^n C_i}{\sum_{i=1}^n X_i} \rightarrow E[C]/E[X] \text{ as } n \rightarrow \infty.$$ 

By our assumption that the number of nodes is very large, the limit is achieved for incremental $\sum_{i=1}^n C_i/\sum_{i=1}^n X_i = \epsilon$, therefore the incremental cost $\sum_{i=1}^n C_i = \epsilon E[C]/E[X]$. It follows that

$$c(r) = \left[\frac{4\sqrt{2}}{\pi c_{\text{nominal}}} \right] \times \frac{1}{\sqrt{\lambda(r)}}.$$ 

The result is intuitive. Indeed, let us assume that the density is increased by a factor of 4. Then the expected length of a transmission decreases by a factor of 2, hence the area occupied by a transmission decreases by a factor of 4. However, now we need twice as many transmissions to cover the same distance. Therefore, if we increase the node density by a factor of 4, the overall area consumed decreases by a factor of 2. This back-of-the-envelope calculation is consistent with the cost function we arrived at.

**Examples:** Let us consider some examples of how this cost function affects the shape of optimal routes. We focus on a network where nodes are placed according to a spatial Poisson process with spatial node density $\lambda(r) = [3 \times 10^{-7} m^2 + 0.01] / m^2$. In this, as in all other examples of bandwidth limited networks, we assume a nominal network cost $c_{\text{nominal}} = 1 / m^2$.

In Fig. 5 line R4 represents the optimal route that connects points A and B, assuming the cost function (4). As expected, the optimal route goes through the denser part of the network, in order to have as short hops as possible. In the figure, we also plot (line R2) the optimal route assuming a constant cost function. Clearly, when the cost is constant the best strategy is to route along the shortest possible route, i.e., the straight line. Even though R2 is shorter than R4, the aggregate cost of routing the packet through R2, calculated by using (4), is 80% larger than the cost of routing through R4.

In Fig. 6(b) we plot the wavefronts $S(r) = 7 x + i$, $i = 1, 2, \ldots$ of the eikonal in the case of a distributed circular source of radius 10 m placed at the location $(10 m, 0 m)$ of the network of Fig. 5. As discussed in Sections II-C and II-D, the value of the eikonal at each location gives the optimal routing cost from the source $\Omega$ at that location in both the anycast and broadcast cases. As expected, the routing cost...
Fig. 6. Constant contours of the eikonal in a network with node density 
\( \lambda(r) = [3 \times 10^{-5} r^2 + 0.01] \) nodes/m², a circular source of radius 10 m centered at the location (20 m, 0 m), and with four cost models: (a) the constant cost model, (b) the bandwidth-limited cost model of Section III-A, (c) the energy-limited cost model of Section III-B, and (d) the model of [15]. Regions of the network depicted with a darker shade of gray have higher node densities.

B. Energy Limited Networks

Let us now consider the same network model, but with the difference that nodes are communicating over a common wireless channel with very large, practically infinite bandwidth. This is the case, for example, with nodes using Ultra-wideband transceivers. In such a case, each transmitter can use its own dedicated bandwidth, and communication is not hampered by interference [23]. On the other hand, we now assume that nodes have a limited energy supply, and so we would like to transport data in the most energy efficient manner. Nodes are assumed to transmit with a fixed data rate.

Transmission Cost: In order to reach a receiver at a distance \( d \), the transmission incurs an energy cost 
\[
\begin{align*}
\text{cost}(d) &= ad^b + f,
\end{align*}
\] measured in Joules. This is a standard model used in the literature [24], [25]. The constant term \( f \) captures the energy consumed in the electronics of the transmitter and receiver, which is independent of the transmission distance. The term \( ad^b \) captures the energy consumed by the transmitter power amplifier, which depends on the distance. The exponent \( b \) typically has a value in the range (2, 6), depending on the environment [26].

Under this energy consumption model, transporting a packet across the network using a large number of very short hops incurs a very large aggregate energy cost, due to the constant term \( f \). Likewise, transporting a packet with a few long hops also incurs a large energy cost, due to the term \( ad^b \) which typically increases with distance much faster than linearly.
In fact, it is intuitively clear that if the nodes could decide on the hop distance, they would select the optimal value \( d_{\text{opt}} \) for which the ratio \( \frac{ad^2 + f}{x} \) is minimized. Straightforward calculation shows that \( d_{\text{opt}} = \left( \frac{f}{b - 1} \right)^{\frac{1}{2}} \). (The calculation first appeared in [27].)

**Cost Function:** The forwarding Rule B used in the bandwidth limited model continues to make sense in this case, with the modification that the transmission cost is given by (5). As in the previous case, we would like to determine the expected cost and progress per hop incurred by this forwarding rule, and use them to establish the cost function.

Unfortunately, the constant cost-progress loci \( \frac{ad^2 + f}{x} = r \) are no longer circles. In Fig. 4 we plot (with dashed lines) the loci for a typical choice of parameter set \((a = 1, b = 4, f = 8 \times 10^9)\). As shown in the figure, the loci are oval shaped, roughly centered at the point \((d_{\text{opt}}, 0)\). As a result, the method of the previous section for finding the expected cost and progress cannot be replicated. Nevertheless, we can easily calculate the expected cost and progress numerically, by Monte Carlo simulation, and divide them to arrive at the cost function. In Fig. 7 we plot the resulting cost function versus the node density.

As in the previous case, the resulting cost function is a decreasing function of the node density. Contrary to the previous case, however, as the node density goes to infinity, the cost function converges to a strictly positive limit. This is expected: after the node density achieves a sufficiently high value, each node is guaranteed to find a node further down the trajectory and at a distance very close at the optimal \( d_{\text{opt}} \). After that point, increasing the density further does not affect the cost function perceptibly, as the optimal point of operation has essentially been achieved.

**Examples:** To gain some intuition about the shape of the optimal routes in this case, we consider the example network of Section III-A, but where the transmission cost is given by (5) with \( a = 1, b = 4, f = 2000 \).

In Fig. 5 we plot the optimal route (line R3) that connects the points A and B. Again, the optimal route deviates through the denser regions of the network. Contrary to the bandwidth limited case, however, the deviation is smaller. The intuition is that, once the route enters a region of sufficiently high density, it has practically achieved the smallest possible cost it can hope for. Deviating further will not diminish the cost function along the route further, but will increase the length of the route.

In Fig. 6 we plot the wavefronts \( S(r) = 12 \times i, i = 1, 2, \ldots \) of the eikonal in the case of the circular distributed source of Section III-A, and assuming \( \epsilon_{\text{nominal}} = 1000 \) Joules/m. The eikonal is similar to the eikonal of the bandwidth limited case, with the exception that now once the wavefronts hit the high density parts of the network, the eikonal increases linearly, as the cost function becomes practically constant.

**C. Other Models**

Our bandwidth-limited and energy-limited models are only two of the many possibilities for modeling routing costs in large wireless networks. Using the same methodology, many other choices are possible. For example, the cost function could be specified in an adaptive manner, depending on the fluctuating needs and capabilities of the network. In other words, we could perform adaptive routing by using a cost function that captures the current congestion at the various parts of the network. Congested areas could temporarily select a high value for their cost function, and so discourage the use of routes through them. On the other hand, areas of the network with spare resources could select low values for their cost function, thereby inviting routes through them. Therefore, we can implicitly perform load balancing, by appropriately modifying, on-line, the cost function.

In fact, alternative cost functions have already been studied in contexts related to our own. For example, the authors of [28] propose, independently from our line of work, a cost function that captures the energy needed for routing in a large-scale static interference field.

As already mentioned, the analogy with Optics was first proposed, in a preliminary form, in [15]. There, the cost function \( c(r) = \sqrt{\lambda(r)} \) was implicitly used. That cost function corresponds to the case where we want to minimize the number of hops, while restraining the communication between nearest neighbors. Line R1 of Fig. 5 represents a route that is optimal with respect to that cost function. Contrary to routes R3 and R4, this route actually prefers the sparsely populated areas of the network, as, in these areas, each hop is along a relatively large distance, and the total number of hops in minimized. In Fig. 6(d) we plot the wavefronts \( S(r) = 4 \times i, i = 1, 2, \ldots \) of the eikonal that corresponds to this cost function (assuming \( \epsilon_{\text{nominal}} = 1 \)). In sharp contrast to the eikonals under the bandwidth limited and energy limited cost functions, this eikonal increases fastest in the densely populated regions of the network.
Fig. 8. A realization of a network with Poisson spatial node density $\lambda(r) = [0.5 \times 10^{-4} r^2 + 0.025]$ nodes $m^{-2}$ and roughly 5000 nodes, with a source placed at $(0 \text{ m}, 0 \text{ m})$ and a destination placed at $(0 \text{ m}, 200 \text{ m})$. (a) Straight line connecting the source and destination, and associate TBF route. (b) Optical route using quantized node density information and associated OTBF route. (c) Curve determined by the Optics analogy and associated OTBF route. (d) Optimal route determined by Bellman-Ford algorithm.

IV. OPTICAL TRAJECTORY BASED FORWARDING

In Sections II and III, the analogy between Geometrical Optics and Networking was developed in the context of the macroscopic limit where the number of nodes is very large, technically infinite, so that routes are modeled as smooth curves. As we show in this section, however, our work is very useful in the standard context where the number of nodes is large but finite.

Indeed, as the number of nodes in a network increases, calculating optimal routes becomes harder, or even impossible, particularly in a wireless setting where the communication and computational capabilities of nodes are limited and nodes may be mobile. Therefore, the optical limit can be used as an approximation of the optimal routes. This optical limit can be calculated much faster, by simulating the propagation of light rays. Furthermore, if we want to approximate the optimal route by a real route (and not a smooth curve), we can use the optical route to perform Trajectory Based Forwarding, which is also much simpler to perform than finding the optimal route. We refer to this approach to routing as Optical Trajectory Based Forwarding (OTBF).

Using the optical limit as the trajectory within the TBF framework is a very natural choice, given that this limit becomes optimal as the number of nodes approaches infinity. It excludes, among others, the common pitfall of TBF of using...
a trajectory that initially passes through a low cost region of the network, but later on becomes strongly suboptimal.

As a simple example, in Fig. 8 we consider a realization of a bandwidth-limited network occupying the region \(-50 \, \text{m} \leq x \leq 100 \, \text{m}, 0 \, \text{m} \leq y \leq 100 \, \text{m}\), with node density

\[
\lambda(r) = [0.5 \times 10^{-4} x^2 + 0.025] \text{ nodes/m}^2.
\]

Within the region, approximately 5000 nodes are placed. We focus on routes connecting two nodes placed at the locations \((0 \, \text{m}, 0 \, \text{m})\) and \((0 \, \text{m}, 200 \, \text{m})\). In the figure, we have plotted (line (d)) the optimal route connecting the source and the destination, as discovered using the Bellman-Ford algorithm. We have also plotted (with the smooth dashed line denoted by (c)) the optical limit, on which the optimal routes converge as the number of nodes goes to infinity, and the route derived by performing trajectory based forwarding on that limit (the continuous line denoted by (c)). Note that this route closely approximates the truly optimal. For comparison, we have also plotted the straight line connecting the source and the destination and the route discovered by performing TBF on that line (both are denoted by (a)). This route is strongly suboptimal, as it passes through a thinly populated region of the network.

A disadvantage of OTBF is the fact that each node must have an accurate estimate of the cost function. For example, if the cost function depends on the node density, then all nodes must know the density in their neighborhood. This information may not be available with high precision.

To evaluate the effect of imperfect information on the cost function, we have also plotted the optical route (the dashed line denoted by (b)), assuming that nodes do not know the actual value of the node density, but only a quantized value, with a total of four quantization levels used throughout the network. The nodes are now forced to calculate the optical limit, but using inaccurate (i.e., grossly quantized) information about the refractive index. In particular, whenever there is an abrupt change in the value of the refractive index, the nodes introduce a bend on the trajectory by executing Snell’s law. In all other regions, the route is a straight line. The resulting OTBF route is the continuous line denoted by (b). Note that, even though we have only four quantization levels, the OTBF route is still quite close to the optimal one.

As a second example, in Fig. 9 we consider a bandwidth-limited network occupying the region \(0 \, \text{m} \leq x, y \leq 100 \, \text{m}\), with node density \(\lambda(r) = [0.5 \times 10^{-4} x^2 + 0.025] \text{ nodes/m}^2\). A source node is placed at the location \((50 \, \text{m}, 100 \, \text{m})\). There are approximately 8000 nodes in the network. In Fig. 9(a) we plot the wavefronts and rays emanating from the source in the optical equivalent of the network. As discussed in Sections II-C and II-D, the rays represent the optimal routes with which the source can send packets to the rest of the network in both the Anycast and the Broadcast Routing cases. Note that the routes tend to bend so that they use regions of the network where the node density is larger and the cost function is smaller. (These regions are plotted with a darker shade of gray.)

In Fig. 9(b) we plot (with dotted lines) 10 rays emanating from the source, and arriving at 10 destinations located throughout the borders of the network. We also plot (with continuous lines) the associated routes determined by OTBF, and finally, with dashed lines, the optimal routes, as determined by Bellman Ford’s algorithm. Note that there is good agreement between the three routes of each of the 10 destinations, particularly for these destinations that are placed on the right, where the node density is larger and we are closer to achieving the optical limit.

In Fig. 9(c) we plot the optimal routing tree that connects, with the minimum incurred cost, each node in the network with the source. Note the close resemblance between the truly optimal routes of the optimal tree of this figure and the asymptotically optimal routes that are determined by solving the eikonal equation and appear in Fig. 9. In particular, the truly optimal routes display the same bending toward the high density regions of the network.

To develop a sense about how fast is the convergence of the optimal routes to the optical limit, we consider in Fig. 10 an inhomogeneous network with a node density 

\[
d(x, y) = a \times (3 \times 10^{-5} + 0.005) \text{ nodes per m}^2,
\]

and for three values of

\[
\lambda(r) = [0.5 \times 10^{-4} x^2 + 0.025] \text{ nodes/m}^2.
\]

Within the region, approximately 5000 nodes are placed. We focus on routes connecting two nodes placed at the locations \((0 \, \text{m}, 0 \, \text{m})\) and \((0 \, \text{m}, 200 \, \text{m})\). In the figure, we have plotted (line (d)) the optimal route connecting the source and the destination, as discovered using the Bellman-Ford algorithm. We have also plotted (with the smooth dashed line denoted by (c)) the optical limit, on which the optimal routes converge as the number of nodes goes to infinity, and the route derived by performing trajectory based forwarding on that limit (the continuous line denoted by (c)). Note that this route closely approximates the truly optimal. For comparison, we have also plotted the straight line connecting the source and the destination and the route discovered by performing TBF on that line (both are denoted by (a)). This route is strongly suboptimal, as it passes through a thinly populated region of the network.

A disadvantage of OTBF is the fact that each node must have an accurate estimate of the cost function. For example, if the cost function depends on the node density, then all nodes must know the density in their neighborhood. This information may not be available with high precision.

To evaluate the effect of imperfect information on the cost function, we have also plotted the optical route (the dashed line denoted by (b)), assuming that nodes do not know the actual value of the node density, but only a quantized value, with a total of four quantization levels used throughout the network. The nodes are now forced to calculate the optical limit, but using inaccurate (i.e., grossly quantized) information about the refractive index. In particular, whenever there is an abrupt change in the value of the refractive index, the nodes introduce a bend on the trajectory by executing Snell’s law. In all other regions, the route is a straight line. The resulting OTBF route is the continuous line denoted by (b). Note that, even though we have only four quantization levels, the OTBF route is still quite close to the optimal one.
a: $a = 0.07$, which leads to networks with approximately 400 nodes, $a = 0.35$, which leads to approximately 2000 nodes, and $a = 1.7$, which leads to approximately 10000 nodes.

From the figure it is clear that, even for the smallest of networks, the optimal routes resemble their optical limit. On the other hand, convergence is slow. Intuitively, the slow convergence is expected, for two reasons: Firstly, each location has its own optical route to the destination. Therefore if a packet has moved away from the original optical route, its move from then on is approximated by another optical route. Secondly, the environment is two dimensional, with a wealth of alternative paths, and it is easy for one of them, due to a fortuitous placement of nodes, to offer a better overall cost. Both these effects are being suppressed slowly as the number of nodes increases.

To develop a sense about how fast is the convergence of the OTBF routes to the optical limit, we are considering in Fig. 11 the inhomogeneous network of Fig. 10, for the same choice of densities, but we now plot, in each case, 20 OTBF route realizations under Rule B. Clearly, convergence is much faster. This is expected, as Rule B strongly enforces the route to follow the trajectory. Under rules that permit larger deviations, aiming at lower overall costs, the convergence is expected to be slower.

Our examples show that the Networking/Optics analogy, in addition to the theoretical interest it has in characterizing the limiting behavior of optimal routes, can also be of use in the context of networks with a number of nodes in the order of a few thousands. As already discussed, such networks have already been deployed or will be deployed, in a variety of applications. Note that in many such settings, calculating the optimal route in a traditional manner, for example by executing the Bellman-Ford algorithm, might be impossible, due to the nodes being mobile and/or having limited computing capabilities. In such settings, OTBF can be very useful, as, on the one hand, it inherits the scalability properties of traditional TBF, but on the other hand it is using the trajectory that is asymptotically optimal.

As a final comment, the examples in this section provide us with a sense of how fast the convergence of optimal routes is to the optical limit. Nevertheless, they do not prove that there is convergence, nor do they provide us with rates of convergence. The theory we have developed so far does not address these very important issues. Rather, we have implicitly assumed that the convergence takes place, and then we showed that the limiting routes must satisfy the rules of Geometrical Optics. To the best of the authors’ knowledge, this compromise was adopted in all other works of a similar flavor (for example all the works referenced in the review [16]).

The process resembles similar methods in calculus, where a sequence is assumed to converge, and this assumption is used to determine its limit. In these cases, if the sequence does not actually converge, the result is meaningless and misleading. The formal proof of convergence and the calculation of rates of convergence, despite their obvious theoretical importance, are relegated to future work, due to their apparent complexity.

V. OUTLINE OF DISTRIBUTED TRAJECTORY DISCOVERY PROTOCOLS

Despite its conceptual and practical usefulness, the analogy between Networking and Geometrical Optics has an important shortcoming: it does not explicitly specify the shape of optical trajectories. Mathematically speaking, we still need to find the angle with which the trajectory will be launched from a source node $A$. Then, we must construct it, using (7).

The second step, i.e., the construction of the trajectory, can be performed by the nodes in a decentralized fashion. In particular, $A$ will launch the trajectory, with every node finding itself on it introducing an incremental bend according to (7) and propagating it further. However, the first step, i.e., the calculation of the initial angle with which the trajectory will leave $A$, cannot be executed in a decentralized manner.
Indeed, even if we know the locations of $A$ and $B$, we cannot use them to determine the initial angle, as the trajectory will bend as it travels away from the source $A$.

In order to deal with this shortcoming, let us consider the following protocol for trajectory discovery (we note that somewhat similar protocols have appeared in the context of position based routing [29]): any node $A$ interested in communicating with a specific node $B$, this information will be included in the RRQ packets. Suppose that $A$ wants to communicate with $B$. Provided $N$ is chosen sufficiently large, the exploratory trajectories from the two nodes will intersect each other (possibly in multiple places). The nodes in the intersections will play the role of middlemen, informing $A$ that packets to $B$ can be routed through them, and node $B$ that it should be expecting packets from $A$.

An example of the process appears in Fig. 12(a), where 2 nodes, placed at the locations $(-2, -2)$ and $(2, 2)$ launch 11 exploratory trajectories each. The network is described by a cost function $c(r) = 1 + 2 \exp[-y^2]$. Therefore, there is a horizontal strip of higher cost, which is depicted with a darker shade of gray. Note that the exploratory trajectories bend so that they cross the high cost strip as fast as possible.

After receiving notification from all locations where the exploratory trajectories intersect, each node will have an idea about the physical location of the other node, and so may choose to launch additional, better aimed rays, with the aim of determining more accurately the launching angle that can reach the other party.

After launching a sufficient number of exploratory trajectories, $A$ and $B$ will adopt the route that consists of two segments of exploratory trajectories and incurs the minimum cost. Therefore, packets will not use an optimal route, but rather two segments of optimal routes. The concatenated route will not be optimal, but, if the number of exploratory trajectories is sufficient, it will be very close to the optimal.

This protocol has two important advantages. First, the nodes $A$ and $B$ do not use any kind of information regarding their own location, or the location of the other party. Secondly, the exploratory trajectories can be created in a distributed manner: when a packet reaches a location of the network $r$, the node in that location will impose an incremental change in the direction of the ray, according to (7), and then will propagate the ray further. (Unless the cost function is not continuous at that location, in which case the change in the direction will be non-incremental, and predicted by Snell’s law (8).) By the Optics-Networking analogy, this property is intuitively clear — light arriving at a location bends according to the properties of the environment only at that location. Therefore, and in contrast to similar schemes, such as Trajectory Based Forwarding [12], there is no need for a node to either find or store a whole route. The only routing overhead that the packet must carry, as it progresses along a route, is the direction of the route at the particular location of the packet. Note that this
overhead does not increase at all with the number of nodes in the network. Therefore, the protocol can be executed in a distributed manner, with participating nodes only requiring information they can obtain locally.

The discussion until now has implicitly assumed that rays emanating from nodes do not intersect with each other. However, this might not be the case, particularly in strongly non-homogeneous environments. As an example, in Fig. 12(b) we repeat the route discovery procedure of Fig. 12(a), but where the first node is moved to the location (−2 m, 0.5 m). As the rays are now launched from within the high cost strip, most of them are trapped within it, and they intersect. The situation is similar to the launching of light within optical fibers, which contain a central core of high refractive index, used precisely with the aim of trapping light.

As a result of the rays intersecting, it is clear that the protocol we have developed until now requires modification. It continues to be true that, for any two points $A$ and $B$, there is a ray connecting them that incurs the minimum cost of communication. If, however, we are restricted to use a given set of exploratory trajectories emanating from the points $A$ and $B$, then it is possible that the optimal route will consist of multiple segments of these rays. For example, the optimal route between the points in Fig. 12(b) consists of three ray segments, which are shown in the figure. Therefore, in order to find the minimum cost route the source node will need to consider all possible combinations of routes, for example by executing the Bellman-Ford algorithm on the graph whose vertices are all the intersection points and edges are the sections of exploratory trajectories.

Another complication arises from the fact that rays emanating from a point uniformly will not cover the network uniformly. One such example appears in Fig 12(c). In the figure, a node placed at the origin launches 12 rays. As shown in the figure, the rays tend to concentrate on the high cost strip, and only two rays manage to escape its pull. Clearly, the network is not covered uniformly. A potential solution is to expand the set of original emanating rays by an additional set of rays, that emanate from positions along those of the original emanating rays that managed to penetrate areas of the network with low cost. This is easy to perform in a distributed manner: When a ray enters areas where the cost is much smaller than the cost was where it started, it is in a part of the network that is not covered well, and so additional rays must be launched. Examples of these secondary sets of exploratory trajectories appear in Fig 12(c).

The discussion until now has focused on the Unicast case. However, the routing protocol we introduced can be readily applied in the anycast case, with only a minor modification: the distributed source $\Omega$ will launch a number $N$ of rays, starting from $N$ positions distributed uniformly along its periphery, and emanating vertically from the boundary of the source.

The routing protocol is also relevant in the broadcast case. In this case, the source can launch its broadcast packet to the rest of the network along $N$ exploratory trajectories, designed to cover the network uniformly, as is shown in Fig. 13. One problem is that the exploratory trajectories will inevitably thin out. Therefore, it will be necessary to introduce frequent bifurcations. The details of a protocol based on this idea are subject for future work.

**VI. Conclusions**

In this work, we investigate an analogy that exists between the optimal routing of packets in networks and the propagation of light according to the laws of Geometrical Optics. The analogy is based on the adoption of a novel macroscopic view of the network, which emerges if we assume the number of nodes to be infinite. Key to the macroscopic view is the cost function, which can be used to describe the cost of transporting packets across the network. Three routing problems are considered: Unicast Routing, Anycast Routing, and Broadcast Routing. We develop and apply a methodology for determining the cost function, we evaluate the applicability of the analogy in networks with finite number of nodes using Trajectory Based Forwarding, and we also outline trajectory based protocols.

The aim of the work was to formulate the analogy and make only an initial investigation of its theoretical foundations and applicability. Clearly, important questions remain unanswered, both on the theoretical and on the practical side.

On the theoretical side, as already discussed in the text, we have not formally determined the rate with which the optimal routes converge to the optical routes as the number of nodes goes to infinity. In fact, we have not even proved that convergence takes place, although our examples and intuition strongly suggest that it does. Also on the theoretical side, we note that the laws of Geometrical Optics can be determined by use of Dynamic Programming. As the Bellman-Ford and related algorithms are also based on Dynamic Programming, there appears to be a strong linkage between Routing and Optics, through Dynamic Programming, which can be leveraged to improve our understanding and arrive at new results. A preliminary investigation of this linkage was undertaken in [2]. The issue of explicitly determining the globally optimal route among many locally routes must also be addressed. Although we limited the analogy in the context of three routing problems, it could be extended to cover other routing problems as well.

Also, we note that our optical routes can be mapped to the geodesics of a properly defined two-dimensional manifold. In more detail, it follows by the Nash isometric embedding theorem that we can map our two-dimensional space, where distances are calculated by use of the cost function, to a two-dimensional set of the three dimensional space (i.e., a manifold) that is not flat, and such that our optical rays map to curves of that set that locally look like straight lines (i.e., they are geodesics of that manifold) [30]. Therefore, a rigorous approach of the problem through appropriate theorems from Differential Geometry might be highly illuminating. This is the subject of future work.

On the practical side, the trajectory discovery protocols of Section V are clearly not fully delineated, and a large number of issues must be specified before they become complete. Some of these issues are: the effects of having a finite number of nodes on the trajectory discovery process, the distributed
calculation of the cost function, the node density, and their gradient by the nodes, the accurate modification of the route direction according to (7), the effects of packets drifting from their trajectory, and the development of alternative cost functions capturing the various limitations of networks. These issues go beyond the scope of this work, and so are also the subject of future work.

As a final comment, note that we have used the Optics-Networking analogy by performing Trajectory Based Forwarding using the optical limit as a trajectory. The analogy could be of use also outside the context of TBF. Consider for example the following approach: for each node pair in the network, the two nodes could establish the angles with which they should launch rays to reach each other. Then, they could use these angles to decide on the next hop of each packet that they receive from other nodes and is destined for the other party. This is equivalent to permitting packets to jump to different trajectories, depending on the results of each hop. This is also a subject of future work.

APPENDIX: REVIEW OF GEOMETRICAL OPTICS

Geometrical Optics studies the propagation of light when the features of the environment have dimensions that are much larger than the wavelength, or, more technically, in the limit when the wavelength goes to zero. Note that the wavelengths of the visible light are on the order of a few micrometers, and the dimensions of almost all the features of our everyday environment, even knife edges, are much larger than that. Therefore, Geometrical Optics describes very well the propagation of the visible light, and its main results follow closely our intuition and everyday experience [31].

In this section, we briefly review Geometrical Optics, focusing on those aspects that are relevant to our work. In particular, we consider only nonhomogeneous, isotropic, and non-conducting materials. The properties of such materials, as far as the propagation of light is concerned, can be completely captured by the refractive index \( n(r) \), a scalar function of the position \( r \). In all media \( n(r) \geq 1 \), and in free space \( n(r) = 1 \). As our networks are two-dimensional, we present Geometrical Optics in a two-dimensional setting.

A. The Eikonal Equation

The emanation of light from a light source is described in terms of the eikonal function \( S(r) \), a scalar function of the position \( r \), measured in meters. At the location \( r \), \( S(r) \) equals the time it takes light starting from the source to reach \( r \), multiplied by the speed of light in free space \( c \). Therefore, \( S(r) \) equals the distance that light would cover in free space in the amount of time it needs to travel from the source to \( r \). Surfaces of the form \( S(r) = k > 0 \) are called wavefronts, and represent loci of points at which light, leaving the source at time 0, will arrive at time \( \frac{k}{c} \).

It follows from Maxwell’s equations that the eikonal function satisfies the following eikonal equation:

\[
\left( \frac{\partial S}{\partial x} \right)^2 + \left( \frac{\partial S}{\partial y} \right)^2 = n^2 \Leftrightarrow (\nabla S)^2 = n^2 \Leftrightarrow |\nabla S| = n. \tag{6}
\]

In the above, \( n(r) \) is the refractive index at \( r = (x, y) \). The vector \( \nabla S = \frac{\partial S}{\partial x} \hat{x} + \frac{\partial S}{\partial y} \hat{y} \) is the gradient of \( S \). The boundary condition that accompanies the eikonal function is that \( S(r) = 0 \) on all points \( r \) that belong to the boundary \( \partial \Omega \) of the source set \( \Omega \).

To understand the meaning of (6), note that at a point \( r \), \( |\nabla S| = \frac{ds}{dr} = \frac{dx}{dt} \) where \( ds \) is the length of an incremental vector starting at \( r \) and pointing to the direction of maximum increase of \( S \), \( dS \) is its increase along \( ds \), and \( dt \) is the time it takes light to cover the distance \( ds \). Applying (6), it follows that \( \frac{dx}{dt} = \frac{c}{n(r)} \), i.e., (6) specifies that at the location \( r \) the wavefront travels with speed \( \frac{c}{n(r)} \).

As an example, in Fig. 13 we plot the wavefronts \( S(r) = 15 \times i, i = 1, 2, \ldots \), of the eikonal function in the case of a circular light source of radius 2 m, placed at the origin, and assuming a medium with refractive index \( n(r) = n(x, y) = \sqrt{6 \times 10^{-3} x^2 + 1} \). In the figure, darker shades of gray denote regions with larger refractive index.

Contrary to Maxwell’s equations, the eikonal equation (6) is nonlinear, and so can have multiple branches. One example appears in Fig. 14. There, we plot the wavefronts \( S(r) = 4 \times i, i = 1, 2, \ldots \), of the eikonal function created by a circular source of radius 2 m located at the origin and embedded in a medium of refractive index \( n(r) = 1 + \exp[-0.1 \sqrt{x^2 + (y - 10)^2}] \). Therefore, the medium has a region with high refractive index (i.e., a lens) centered at the location \( (x, y) = (0, 10) \). As they leave the source, the wavefronts are initially circular. However, as light...
approaches the lens, it progressively slows down, and as a result the wavefronts start to fold. Consequently, in a part of the environment the eikonal function has three branches, all of which satisfy the eikonal equation. Locations within that part have three values for the eikonal function. The physical interpretation of the multiple values is that, in these locations, light arrives along three different wavefronts, each of which has its own time and angle of arrival. In Fig. 15 we plot the loci of the eikonal $S(r) = 4 \times i, i = 1, 2, \ldots$, for the medium with refractive index $n(x, y) = 1 + \exp[-0.1\sqrt{x^2 + (y - 10)^2}]$. Where the eikonal has multiple values, only the smallest is retained.

**B. Light Rays**

Light rays are curves that originate at the light source and move away from it, while staying orthogonal to wavefronts. Examples appear in Figs. 13 and 14. Intuitively, light rays are the paths that light takes as it emanates away from the source.

By applying the eikonal equation (6), it can easily be seen [31] that light rays satisfy the following nonlinear second-order differential equation:

$$\frac{d}{ds}(n \frac{dr}{ds}) = \nabla n(r),$$

(7)

where $r = r(s)$ is a point on the ray, $s$ is the length of the ray measured from the source to the point $r$, $\frac{dr}{ds}$ is the direction of the ray, $n(r)$ is the refractive index at $r$, and $\nabla n(r) = \frac{\partial n}{\partial x}(r)\hat{x} + \frac{\partial n}{\partial y}(r)\hat{y}$ is its gradient at $r$. To determine the propagation of a light ray across the medium, we must apply (7) together with initial conditions on the origin of the ray, i.e., $r|_{s=0} = r_0$, and the angle with which it is launched, i.e., $\frac{dr}{ds}|_{s=0} = \theta_0$.

To gain intuition on (7), note that the gradient $\nabla n(r)$ points toward the direction where the refractive index increases fastest. Therefore, (7) expresses the fact that light rays tend to bend toward the direction where the medium is optically denser, i.e., the refractive index has a higher value. A quick inspection of Figs. 13 and 14 confirms this.

Note that the bending of the light rays toward regions of high refractive index does not imply that light rays ‘prefer’ to travel through these regions. On the contrary, light rays ‘prefer’ to travel through areas of low refractive index, as there their speed is large, and so they arrive to their destination faster, as required by the Principle of Fermat (see Section C of this Appendix). Indeed, the bending of the rays toward the regions of high refractive index helps them to avoid these regions while traveling to their final destination. This is shown best in Fig. 14.

Equation (7) holds wherever the refractive index is a continuous function of the position. Locations with discontinuities can be handled by applying (6) together with limiting arguments, leading to Snell’s law of refraction. In particular, let us consider an interface between two media $M_1$ and $M_2$, crossed by a ray of light which travels from $M_1$ into $M_2$. Observing the interface from sufficiently close, we can assume the interface and the two sections of the ray to be straight lines. We can also assume that the refractive index remains constant on either side of the interface, and equal to $n_1$ and $n_2$ respectively. The setting appears in Fig. 16. Let $\theta_1$ be the angle between the part of the ray in $M_1$ and the normal to the interface, and $\theta_2$ the angle between the part of the ray in $M_2$ and the normal.
Snell’s law of refraction specifies that
\[ n_1 \sin \theta_1 = n_2 \sin \theta_2. \]  
(8)

Therefore, as in the case of a smoothly varying refractive index, the rays bend toward optically denser material. Note that when \( n_1 > n_2 \) and \( \theta_1 \) is sufficiently large, there is no \( \theta_2 \) that can satisfy (8). In this case, we have total internal reflection, and the boundary acts as a perfect mirror, reflecting the ray back to \( M_1 \) [31] with an angle of reflection equal to the angle of incidence.

As shown in Fig. 14, in a location where the eikonal function has multiple values, there are multiple rays arriving at the location, all from different directions, and each associated with a distinct wavefront. The physical interpretation is that a person sitting at that location will be seeing multiple images of the source, arriving at different times, and coming from different angles. Wearers of spectacles are (consciously or unconsciously) familiar with this situation.

### C. The Principle of Fermat

We define the optical length \([AB]_L\) of a curve \(L\) (not necessarily a ray) connecting two points \(A\) and \(B\) to be the line integral
\[ [AB]_L \triangleq \int_A^B n(r) \, ds \]  
(9)
along the curve \(L\). Since the speed with which light moves at a point \(r\) is \( v = \frac{c}{n(r)} \), the time needed by light to go from \(A\) to \(B\) or vice versa along the curve \(L\) equals
\[ \int_A^B dt = \int_A^B \frac{1}{v} \, ds = \frac{1}{c} \int_A^B n(r) \, ds = \frac{[AB]_L}{c}. \]

Therefore, the optical length of a curve equals the time it would take light to travel across the curve, multiplied by the speed of light in free space \(c\).

Let \(R\) be a curve starting at \(A\) and ending at some other point \(B\), as shown in Fig. 17. The principle of Fermat states that the curve \(R\) will be a ray of light if and only if its optical length is a local minimum, i.e., it is not larger than the optical length of any other curve \(L\) that passes through \(A\) and \(B\) and exists within a sufficiently small neighborhood of the ray \(R\). In other words, if we introduce a sufficiently small perturbation in a ray \(R\), resulting in a new curve \(L\), then the new curve cannot have a strictly smaller optical length. As the optical length is proportional to the time needed to go from \(A\) to \(B\), the principle of Fermat asserts that light takes the locally fastest route.

We stress that the optical lengths of rays of light are local, and not global minima. Therefore the principle of Fermat does not exclude the existence of multiple rays with unequal optical lengths connecting two points, as happens, for example, in Fig. 14. Such rays, however, cannot be in the local neighborhood of each other.

### References