Landmark Selection and Greedy Landmark-Descent Routing for Sensor Networks

An Nguyen† Nikola Milosavljević† Qing Fang* Jie Gao‡ Leonidas J. Guibas†

† Department of Computer Science, Stanford University. {anguyen,nikolam,guibas}@cs.stanford.edu
* Department of Electrical Engineering, Stanford University. jqfang@stanford.edu
‡ Department of Computer Science, Stony Brook University. jgao@cs.sunysb.edu

***Abstract*—We study the problem of landmark selection for landmark-based routing in a network of fixed wireless communication nodes. We present a distributed landmark selection algorithm that does not rely on global clock synchronization, and a companion local greedy landmark-based routing scheme. We assume no node location information, and that each node can communicate with some of its geographic neighbors. Each node is named by its hop count distances to a small number of nearby landmarks. Greedy routing at a node is performed to equalize its vector of landmark distances to that of the destination. This is done by following the shortest path to the landmark that maximizes the ratio of its distances to the source and the destination. In addition, we propose a method to alleviate the difficulty in routing to destinations near the boundaries by virtually expanding the network boundaries. The greedy routing, when combined with our landmark selection scheme, has a provable bounded path stretch relative to the best path possible, and guarantees packet delivery in the continuous domain. In the discrete domain, our simulations show that the landmark selection scheme is effective, and the companion routing scheme performs well under realistic settings.**

Both the landmark selection and greedy routing assumes no specific communication model and works with asymmetric links. Although some of the analysis are non-trivial, the algorithms are simple, flexible and cost-effective enough to warrant a real-world deployment.

I. Introduction

The applications emerging from the sensornet community will depend, in part, on good design and implementation of scalable point-to-point routing. In this space, both geographic location-based and location-free methods have been proposed.

In geographic routing, the node currently holding the packet forwards the packet to its neighboring node that is "closest" to the destination. Various meanings of "closest" are possible. The presence of holes in the sensor field can cause these greedy methods to fail at nodes with no neighbor closer to the destination. The failure of greedy forwarding is mainly due to the mismatch between naming/routing rules and the actual network connectivity when the node distribution is sparse, having holes or of complex shape.

A variety of methods have been proposed to overcome this difficulty and to guarantee packet delivery, if at all possible. Probably the best known approach among these, the scheme of greedy routing combined with face routing [1], [2], [3] builds a planar subgraph of the connectivity graph and uses perimeter forwarding when greedy forwarding gets stuck. While greedy forwarding is quite robust to sensor location inaccuracies, perimeter routing fails in practice — a delivery rate of only 68% is observed [4], [5]. The problem comes from the failure of the planarization of the communication graph in real world deployment. The relative neighborhood graph and the Gabriel graph are both connected planar subgraphs of unit disk graphs (UDG) in theory. In reality they turn out to be disconnected due to directional or crossing edges, as the consequence of location inaccuracy and the deviation of the real connectivity model from the UDG model.

To solve this problem despite the fact that accurate geographic location information is difficult and expensive to obtain, several location-free routing schemes have been developed [6], [7], [8], [9]. These schemes build virtual coordinate systems, respecting the true network connectivity, on which greedy message delivery rules are developed.

Several virtual coordinate routing systems are landmark-based [7], [8], [9]. Generally speaking, a subset of the nodes are selected as *landmarks*. Every node records its hop count distances to these landmarks. The landmark distances are then used to generate virtual coordinates for the nodes. Routing is typically guided in a greedy way by a potential function on the landmark-based distances. The major differences among these schemes are mostly in their potential functions.

Take two landmark-based schemes as examples. Gradient Landmark-based Routing (GLIDER) [7] separates global and local routing by partitioning the network into combinatorial Voronoi tiles such that the nodes inside the same tile are closest to the same landmark. The tile adjacency graph is used to guide global routing across tiles. Local routing within a tile is done by greedy descending on a potential function constructed using distances to a set of local landmarks (the landmarks in the adjacent tiles). Such a potential function is local minimum free in the continuous domain. In another scheme, Beacon Vector Routing (BVR) [8] takes a similar but more heuristic approach and designs a potential function that depends on the distances to the k landmarks closest to the destination, where k is a system parameter. Out of the k landmarks, the ones that are closer to the destination than to the source impose a "pulling" force while the rest impose a "pushing" force. The potential function is a combination of the two.

Landmark-based schemes are favored by their simplicity and independence on the dimensions of the network. Unlike face routing in planar graphs that requires a planar deployment of sensor nodes, landmark-based routing schemes can be easily extended to sensors deployed in 3D.

Acknowledgment: The authors wish to acknowledge the support of NSF grants CNS-0435111, CNS-0626151, ARO grant W911NF-06-1-0275, and the support of DoD Multidisciplinary University Research Initiative (MURI) program administered by ONR under Grant N00014-00-1-0637.

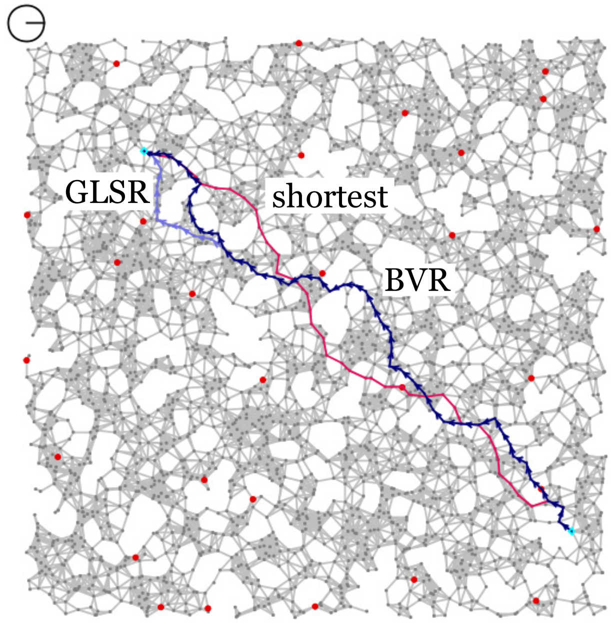

Fig. 1. A typical sensor field and routing paths between 2 nodes in it.

An important issue in all landmark-based schemes that has not been addressed is landmark selection. Obviously, poorly selected landmarks can result in poor routing performance. Simulations often assume that landmarks are randomly selected or manually placed along network boundaries [7], [8]. It remains a challenging open problem, both in theory and in practice, to understand the criteria for landmark selection and their influence on routing performance.

In this paper we propose a *distributed landmark selection protocol* and an accompanying simple routing scheme, *Greedy Landmark Descent Routing (GLDR)*. In the continuous domain, GLDR guarantees packet delivery with provably bounded path stretch relative to the best path possible. It exhibits good routing performance in networks, as verified by our simulations. In our scheme, landmarks are selected by a randomized and distributed process that ensures approximately uniform spacing and good separation of landmarks, for a given separation parameter.

We show that only a constant number of nearest landmarks of a node, denoted by the *addressing landmarks*, is needed to address the node for routing. In practice, as few as 6 nearest landmarks are enough to obtain a high routing success rate. The routing is simple: intuitively, when the source is far away, the addressing landmark that is closest to the destination applies a pulling force until the packet gets close to the destination. At this point, the algorithm routes the packet towards a single landmark selected by *a local rule* until the distance to the landmark equals the distance from the destination to that landmark. Then a new landmark is selected and the above process repeats. We evaluated the performance on low-degree graphs on randomly placed nodes and observed favorable performance in both routing stretch and delivery rate compared to BVR.

The proposed landmark selection protocol can be combined with other landmark-based routing methods. We show by simulation that our landmark selection protocol improves the performance of BVR consistently and significantly compared to that using a random landmark selection.

II. Landmark Selection

The success of GLDR depends on making good choices of the set of landmarks. We build a shortest path tree rooted at each landmark via flooding in the initialization phase. Intuitively, GLDR allows the landmarks near the destination to impose a "pulling force" on the packet following the shortest path. A packet en route is therefore under the forces from a set of nearby landmarks. By choosing only one pulling force at a time according to a *local rule*, and with a set of well selected landmarks, we can navigate the packet in the network. Therefore, a desirable landmark selection algorithm should give a set of landmarks that are evenly spread out, and each node in the network, if not near a field boundary, is well "surrounded" by a small set of nearby landmarks.

We show that we can use an r-sampling to select a 'good' set of landmarks. We show, in the next section, how the r-sampling properties are essential to guarantee the greedy routing success, and to yield a theoretically bounded per-hop stretch factor (of 5 for realistic settings). In fact, by incurring higher cost in node address composition, even lower stretch could be achieved. For details, see section III-B. We propose a distributed protocol to implement an approximate r-sampling in a practical setting.

A. r-sampling in the Continuous Domain

Let $B(p, r)$ denote the ball (disk) of radius r centered at p.

Definition 1: Given a continuous domain S, an *r-sampling* of S is a set of points $R \subset S$ such that each point in S is within distance r of some point in R, i.e. $S \subseteq \cup_{p \in R} B(p, r)$.

Definition 2: We call a point p an *interior point* if a ball centered at p with radius $4r$ is completely inside S, i.e. $B(p, 4r) \subseteq S$.

We are interested in landmarks that form an r-sampling and would like to reduce the total number of landmarks needed for a given sensor field.

For a 2-D infinite continuous domain, the best result possible is to place landmarks at points such that they form a hexagonal sampling pattern [10]. This could be achieved by a global sequential algorithm.

Another strategy is to use random sampling. How many independent samples (in fact landmarks), selected uniformly at random from S, are sufficient to yield an r-sampling with a confidence level s? We divide the domain into square grid cells of area $\Theta(r^2)$ and solve for the number of landmarks necessary to guarantee at least one landmark in each grid cell. This reduces our question to the *coupon collector's problem*. By a well known result [11], for m equally sized grid cells, the expected number of landmarks needed is $m(\ln m + O(1))$. The logarithmic factor characterizes the expected non-uniformity of a uniform random distribution. In other words, the over-sampling rate of the randomized landmark selection algorithm with respect to the deterministic optimal algorithm is about logarithmic in the number of grid cells.

B. r-sampling in the Discrete Case

Definitions 1 and 2 have their discrete-case analogs, with the distance between a pair of nodes defined as the length of the shortest path between them.

Definition 3: The *r-neighborhood* of node c, denoted by $B(c, r)$, is the set of nodes at most r hops away from c.

We assume in this subsection that the sensor field is nice, in a sense that any neighborhood of radius r in the sensor field contains at least αr^2 nodes for some constant α.

To obtain an r-sampling, we need to have at least one landmark in every r-neighborhood of any interior node. One way to carry out the sampling is to use the following sequential greedy procedure.

Algorithm 1: All nodes are initially unmarked. While there are unmarked nodes, repeatedly pick an unmarked node, make it a landmark, and mark all nodes in its r-neighborhood.

This algorithm produces a set of landmarks with a minimum distance of r-hops apart, and each non-sampled node being in the r-neighborhood of at least one landmark. Since the landmark's $r/2$-neighborhoods are disjoint, there are at most $4n/(\alpha r^2)$ landmarks.

On the other hand, in an ad hoc network, any sequential algorithm is prohibitively expensive. A parallel algorithm is thus much more preferable. This motivates the following randomized algorithm that could be executed in parallel.

Algorithm 2: Each node makes itself a landmark with a probability p where p is some constant.

We have the following result:

Theorem 1: The set of landmarks produced by Algorithm 2 is an r-sampling with probability s if $p \geq 1 - (1 - s^{\frac{\alpha r^2}{16n}})^{\frac{4}{\alpha r^2}}$.

Proof: First select a subset of nodes which we call *centers* by following the procedure described in Algorithm 1, substituting the r-neighborhood with $\frac{r}{2}$-neighborhood. By the previous analysis, there are at most $m = 16n/(\alpha r^2)$ centers.

If we make sure that there is a landmark in each center's $\frac{r}{2}$-neighborhood, the set of landmarks would be an r-sampling. Let $\mathcal{E}_i$ be the event that no node is sampled in center i's $\frac{r}{2}$-neighborhood, then $\Pr[\mathcal{E}_i] \leq (1-p)^{\alpha r^2/4}$. The probability that Algorithm 2 produces an r-sampling is given by $\Pr\left[\cap_{i=1}^{m}(\neg\mathcal{E}_i)\right] \geq (1-(1-p)^{\alpha r^2/4})^{n/(\alpha r^2/16)}$. Setting this probability to be at least s and solve for p we obtain the desired result. ■

Corollary 1: The expected over-sampling rate of the parallel random sampling algorithm to that of the sequential deterministic algorithm is therefore at least:

$$\frac{n \cdot [1 - (1 - s^{\frac{\alpha r^2}{16n}})^{\frac{4}{\alpha r^2}}]}{\frac{4n}{\alpha r^2}} \approx \frac{\alpha r^2}{4} \cdot \left[1 - \left(\frac{1-s}{\frac{16n}{\alpha r^2}}\right)^{\frac{4}{\alpha r^2}}\right]$$
$$= k \cdot \left[1 - \left(\frac{f}{4m}\right)^{1/k}\right]$$

Where $k = \alpha r^2/4$ is the lower bound on the number of nodes in an $\frac{r}{2}$-neighborhood, $f = 1 - s$.

For a fixed confidence level and a large network, as the r-neighborhood size grows, the over-sampling rate of the parallel algorithm to that of the sequential algorithm grows linearly with the local neighborhood size (which is quadratic in r). The performance gap between the sequential algorithm and the randomized parallel algorithm is rather significant. This result motivated us to design a protocol that finds the middle ground between the two algorithms.

C. The Protocol

During the network initialization phase, we select a set of landmarks by computing an approximate r-sampling of the network nodes, and the shortest path trees rooted at these landmarks. We propose to use a randomized, distributed scheme that combines both processes. First, each node in the network generates a number v uniformly at random in $[0, 1)$. Each node then waits for Kv time, where K is a network parameter, and the unit of time is the time it takes to send a packet from a node to its neighbor. If during this waiting period, a node is not suppressed by any other node, it declares itself a landmark. The new landmark broadcasts its status to its r-neighborhood, suppressing all non-landmark nodes in it.

When K is large, the above scheme effectively gives us a sequential algorithm for computing an r-sampling. When K is small, some supressing messages arrive too late due to the delay incurred in communication. Consequently, additional landmarks are selected. When the nodes are distributed approximately regularly (as defined below), we can upper bound the expected number of landmarks.

Definition 4: We say that the node distribution is (α, β)-regular ($\beta \geq \alpha \geq 1$) if the $\lfloor r \rfloor$-neighborhood contains between αr^2 and βr^2 nodes, for any $r \in \mathbb{R}$.

Theorem 2: If the node distribution is (α, β)-regular, then any r-neighborhood contains, in expectation, at most $(9\beta/Kr^2 + 16\beta - 1)$ landmarks.

Proof: See Appendix VI-A. ■

Theorem 2 shows that by adjusting K, we can tradeoff between the over-sampling rate and the time requires to complete the sampling process. For $K = \Theta(r)$, the over-sampling rate grows linearly with r. To have a constant over-sampling rate, we need $K = \Theta(r^2)$.

III. Landmark-descent — the Greedy Rule

To route a packet to a destination, the source node selects, out of the landmarks that address the destination, the one that maximizes the ratio of distances to source and destination. It then moves towards this landmark along the shortest path, until the landmark becomes equally distant from the source and the destination, at which point, the process repeats.

Effectively, the source is trying to equalize its vector of landmark distances to that of the destination, by only *reducing* the distances, i.e. going *toward* landmarks, not away from them. If the source and the destination coincide, these vectors are indeed equal. However, the converse is not true because multiple nodes may have the same hop count distances to the same set of landmarks. Nevertheless, in the continuous domain we will show that not only does the converse hold, but the distance vector can be made equal using a simple greedy strategy, i.e. decreasing the coordinate that maximizes the discrepancy (the ratio to the corresponding destination coordinate).

In the discrete domain, our experimental results show that such a greedy strategy succeeds with high probability. However, it is possible that the packet gets stuck into a loop before reaching its destination. Based on our simulations, by remembering the IDs of the last 8 *extreme nodes* (nodes at which the packet stopped proceeding to its current landmark), we can detect such a situation and declare the packet gets stuck.

Once a packet gets stuck, it is greedily forwarded to the destination using L_1 and L_∞ norms, and if the destination is still not reached, a scoped flooding is used to deliver the packet.

In the following sections, we first describe the continuous version of the landmark descent routing and provide theoretical results. We then describe the routing algorithm in a discrete network. We assume continuous, simply connected (without "holes") sensor domain with an outer boundary in the form of a closed curve.

A. Greedy Success in the Continuous Domain

Suppose there is a set of *landmark nodes* in the domain. The algorithm will use only *distances* to landmark nodes for routing to non-landmark points. Each destination is addressed by the vector of distances from a *subset* of landmarks. We will see later, by theory and simulation, that this number can be made small (independent of the network size).

In landmark-descent, the source is trying to equalize its vector of landmark distances to that of the destination. If the source and the destination coincide, these vectors are indeed equal. In this section we show that under certain conditions on the landmark placement the converse is also true, i.e. the destination is *the only* point with this property. Furthermore, we show that the vector can be made equal using a simple greedy strategy.

For simplicity of notation, throughout we use s, d and l to denote the source, destination and landmark points that are used in one routing step. Furthermore, we denote by s' the point to which the source moves after the step has been taken.

The following lemma says that unless it has reached the destination, the source can always find a discrepancy in the coordinates, i.e. the process can never terminate prematurely. This is under the condition that the destination be an interior point.

Lemma 1: If d is an interior point, then as long as $s \neq d$, there is a landmark l that satisfies $|sl| > |dl|$.

Proof: Assume $s \neq d$, and consider the intersection of $B(d, 4r)$ and the open halfplane defined by the line through d perpendicular to sd, and not containing s (Figure 2). This area is contained in the domain, since d is an interior point, and it contains a ball of radius r. Therefore it contains a landmark l, by the r-sampling condition. Clearly $\angle sdl > 90° > \angle dsl$, so $|sl| > |dl|$. ■

The next result says that by equalizing the coordinates, the source indeed gets closer to the destination in terms of the straight-line distance.

Lemma 2: If $|sl| > |dl|$, then $|s'd| < |sd|$.

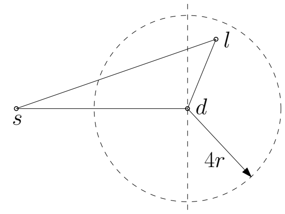

Fig. 2. Landmark descent can always proceed if d is an interior point.

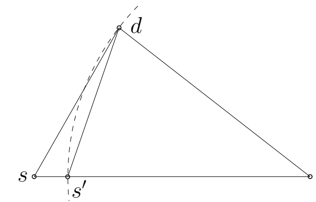

Fig. 3. Landmark descent always makes progress if $|sl| > |dl|$.

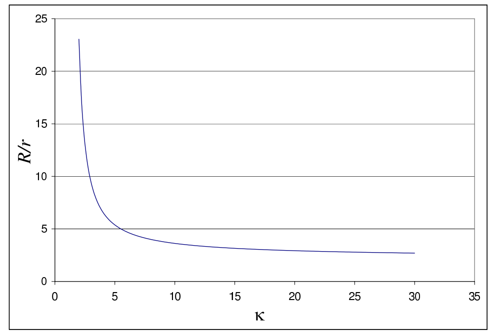

Fig. 5. Upper bound on R/r for various κ.

Proof: Refer to Figure 3. Since $\triangle ds'l$ is isosceles, $\angle ds'l < 90°$, so $\angle ss'd > 90°$. This implies $\angle dss' < 90°$, hence $|s'd| < |sd|$. ■

From the previous two results it is obvious that the routing process eventually terminates by reaching the destination, if the latter is an interior point. We summarize this part in the following theorem.

Theorem 3: If d is an interior point, the landmark descent routing to d is always successful.

B. Bounded Stretch in the Continuous Domain

In this section, we analyze the theoretical performance of our routing approach. Let s be the current point, d the destination, l a landmark, $\rho = |sl|/|dl|$, and $\phi = \pi - \angle sdl$ (see Figure 4). Assume that $\rho > 1$ and let s' be the point on the line sl such that $|s'l| = |dl|$. The ratio $\kappa = |ss'|/(|sd| - |s'd|)$ we call the *one-step stretch factor* of routing from s to d using l.

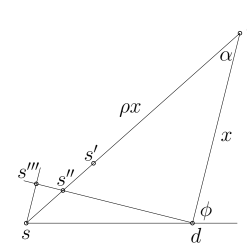

Fig. 4. Bounding the one-step stretch κ by a function of ρ or ϕ.

We first relate κ to the minimum distance between d and its addressing landmarks (including l).

Lemma 3: If the landmarks form an r-sampling and the distance between d and its addressing landmarks is at least R, then $f(\kappa) \geq R/r$, where

$$f = \frac{4}{1 + u - \sqrt{w}}\,, \qquad u = 1 - \frac{1}{\kappa^2}\,, \qquad v = 1 - \frac{1}{\kappa}\,,$$
$$w = (1 - u^2) + 2u^2(1 - v^2) + 2u\sqrt{(1 - v^2)(1 - u^2v^2)}\,.$$

Proof: See Appendix VI-B ■

It is straightforward to show that for $\kappa > 1$, $f(\kappa)$ is monotonically decreasing with range $(2, \infty)$. It follows that its inverse f^{-1} exists, and $\kappa \leq f^{-1}(R/r)$ for $R > 2r$. The plot of f is shown in Figure 5. Apparently, the lower the stretch upper-bound, the farther away the addressing landmarks need to be. To have an upper bound of 5 on the per-hop stretch, we need each node to be addressed by its distances to landmarks that are at least $5r$ away. We will show in the next corollary that the number of addressing landmarks for each node is an scale-free constant. We have thus proved:

Theorem 4: If $R > 2r$, then κ is bounded.

In the rest of the paper we assume $R = 2r$ for simplicity.

Now we can bound the number of addressing landmarks, a quantity essential for scalability of the system, under the regularity assumption of node distribution in the previous section.

Corollary 2: If the node distribution is (α, β)-regular, then the expected number of addressing landmarks for any node is at most $(144\beta/Kr^2 + 16\beta - 1)p$.

Proof: Not difficult, but omitted for brevity. ∎

C. Greedy Landmark Descend Routing (GLDR) in Networks

Each GLDR packet carries in its header, (i) the address of the destination, (ii) the ID of the landmark toward which the packet is forwarded, (iii) a list of recently encountered extreme nodes.

The address of a destination is a vector of distances to its addressing landmarks, together with its ID (multiple nodes may have the same set of nearest landmarks). A packet's header is updated only at extreme nodes, nodes with the same distance to the current landmark as the destination. At an extreme node, if there is no landmark that is farther from the node than from the destination or if the extreme node is found in the packet header, the packet is stuck. If not, the current extreme node is added to the header. The packet is then forwarded toward the landmark that maximizes the ratio of distances to the current node and to the destination.

If a packet gets stuck, we resort to flooding. As this is expensive, we actually forward the packet greedily using L_1 norm, then using L_∞ norm. If the destination is still not reached, we perform scoped flooding. The pseudo code to implement the algorithm is shown in III-C.

To keep the packet header small, we do not store all extreme nodes that we encountered and only keep the last 8 recently encountered extreme nodes in the header. While there is no theoretical proof that such a short list suffices to detect looping, in our extensive simulation we never encountered a loop. This is because looping is usually caused by too small a number of landmarks being available for routing.

We would like to emphasize that GLDR is simple and easy to implement. A local greedy rule guides the greedy routing at each step. The protocol is stateless unless loop occurs in the route, which rarely happens.

```
if p.mode = GLDR then
   if dist(p.bestBase) > dist(p.dest) then
      FORWARD p toward p.bestBase
   else
      if current node is in p.extremeNodes then
         SET p.mode = L1
      else
         RECORD current node into p.extremeNodes
         COMPUTE new base p.bestBase
         if p.bestBase = null then
            SET p.mode = L1
         else
            FORWARD p toward p.bestBase
         end if
      end if
   end if
end if
if p.mode = L1 then
   if some neighbor nb is closer to p.dest using L1 then
      FORWARD p to nb
   end if
   SET p.mode = L∞
end if
if p.mode = L∞ then
   if some neighbor nb is closer to p.dest using L∞ then
      FORWARD p to nb
   end if
   FLOOD to find p.dest
end if
```

Algorithm 1. Pseudocode for routing a packet p.

D. Handling the Boundary Effect

Recall that the results we discussed up to now apply to *interior nodes* only. For the non-interior nodes, the routing is more challenging. Intuitively speaking, for virtual coordinate based methods, routing to destinations near the network boundaries is more difficult because there are less and/or only one-sided reference points to navigate the packets. Specifically in our case, the assumptions of our continuous domain results (Lemma 2, Theorems 3, and 4) are not even approximately satisfied near the boundary. This setback is common in all virtual coordinate based methods.

To alleviate this "boundary effect", we propose to *virtually expand* the sensor field. Suppose that we duplicate the *non-interior* part of the network, and place it immediately below the original so that they meet along the boundary. The expanded network is obtained by "unfolding" the replica. This effectively provides "pulling forces" on packets towards the boundary.

We first extend the definitions of shortest paths and distances. Paths between the original (*real*) and replicated (*virtual*) nodes are required to visit a boundary node, and are thus called *reflected paths*. The distance between two such nodes (*reflected distance*) is the length of shortest reflected path connecting them. The reflected and the standard straight-line distance between x and y are denoted by $\langle xy \rangle$ and $|xy|$, respectively.

Now we can define an augmented landmark selection procedure and the matching routing algorithm to be essentially the same as in the base case. This is possible because the algorithms operate only on shortest paths.

If the boundary is not very curved compared to r, it locally looks like a straight line (Figure 6 left). Thus using a virtual landmark is almost the same as using its "mirror image" with respect to the boundary, viewed as a real landmark.

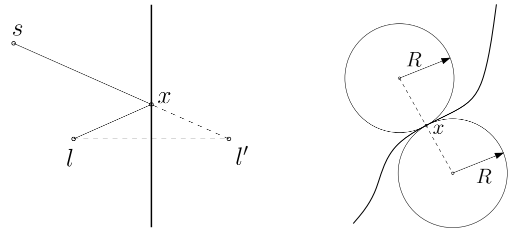

Fig. 6. Left: When the boundary is "locally straight", replacing a virtual landmark l by a "reflected" real landmark l' preserves the distances involved (e.g. $|sl'| = \langle sl \rangle$), and does not change the routing step. For a general low-curvature boundary, the change will be small. Right: Assumption on the curvature of the boundary.

From the above discussion it should be clear that the results will depend on some measure of *curvature* of the boundary. As the boundary is not always known in a realistic setting, and even more complex issues arise when there are large 'holes' [12], [13] in the network, we at this stage study this method more or less as a theoretic interesting possibility rather than a practical solution to ad-hoc networks. In particular, we assume that the boundary is a smooth curve such that the two tangential balls of radius R at any point x on on the curve, do not intersect the curve anywhere except in x (Figure 6 right). In practice the boundary can often be approximated by such a curve to within an accuracy comparable to discretization error.

Theorem 5: If R is large enough compared to r, the augmented-GLDR guarantees packet delivery along a path of bounded stretch factor. In particular, if $R \geq 320r$, and the stretch factor of the base case routing is $\kappa < 2.5$ (cf. Theorem 4), then the stretch factor of the augmented-GLDR is at most $7\kappa/(5-2\kappa)$.

Proof: The key idea is to replace a virtual landmark used in a given routing step by a carefully designed real landmark which results in a similar routing step. The base case results can then be applied to the latter. For technical details, see Appendix VI-C. ∎

The reason for using virtual landmarks is to compensate for the lack of addressing landmarks for non-interior nodes along the field boundary. They are used, and only incur extra storage cost, in a local neighborhood along the boundary. Specifically, virtual landmarks need only to flood $12r$ hops inward from the boundary, to let nodes in this region learn their distances to virtual landmarks, and non-interior nodes learn their local addresses. It is easy to see that only nodes within $12r$ hops away from the boundary need to learn their distances to virtual landmarks: all the other nodes are interior, so they must have at least one *real* addressing landmark; after only one step of landmark descent (always to a real landmark), the new source position s' will be within $8r$ hops from d (for it suffices to pick *any* of d's real addressing landmarks, which are all at most $4r$ hops away from d), i.e. at most $12r$ hops away from the boundary.

IV. Simulations

We implemented our landmark selection and routing algorithms in java, for heuristic study. We simulate the network at the network level, not taking into account networking issues such as packet loss or delay.

A. Default simulation setup

In our simulations, we generated 3200 nodes randomly in a unit square, see Figure 1. The communication radius was set to make the average node degree roughly 10. The field was fairly dense, but still had many small holes, which made greedy routing challenging.

We generated the landmarks to form an 10-sampling. The default K was 1200. For a given sensor field, we ran our landmark selection schemes 100 times, and for each set of landmarks obtained, we generated 320 pairs of random source and destination points and computed the routes between them using that set of landmarks. When routing, we used the distance to the nearest 10 landmarks of a node as its address.

In each simulation, we varied a single parameter and kept the other parameter as in the above default setting. We collected statistics on all the 32000 paths computed for different parameters we were interested in. For comparison, we ran BVR and computed statistics on the same 32000 paths on the same sensor field.

In this default setup, there were on average 28.2 landmarks. Our routing algorithm achieved a success rate of 90.9% without flooding. When the routing failed, the hop distance from the node where the algorithm terminated to the destination was 4.47 on average, with a standard deviation of 2.34.

B. Successful delivery rate

We considered the performance of the routing algorithm by varying the number of routing landmarks, the r-sampling parameter (landmark separation distance, roughly), and the average network degree. The success rate is shown in Figure 7.

First we gradually increased the number of *routing* landmarks and found that, as expected, the success rate also increased, from 85% with only 6 landmarks, up to 92% with 18 landmarks. We observed that the nodes where GLDR failed to delivered were often near the boundary so that they were not "well surrounded" by their routing landmarks. GLDR must often rely on greedy routing using L_1 norm and L_∞ norm to reach boundary nodes, and increasing the number of routing landmarks does little to help for this phase.

When the number of routing landmarks was small, GLDR was more successful than BVR; as the number increased, the performance gap narrowed. BVR's random landmark selection often produces highly irregular landmarks and may not work well when only a few routing landmarks are used.

We also examined the influence of the *total* number of landmarks. We varied the separation distance r in the landmark selection algorithm between 6 and 18, and the number of landmarks from around 60 to around 20. The performance improved with more landmarks, mostly because more nodes near the boundary were "well covered" by the extra landmarks.

Finally, we varied the communication radius so that the number of communication neighbors, averaged over all nodes, was between 6 and 18. As the average degree increased, fewer landmarks were used, so that the landmark density remained relatively fixed. GLDR performance improved when network degree was high, as the network became more regular.

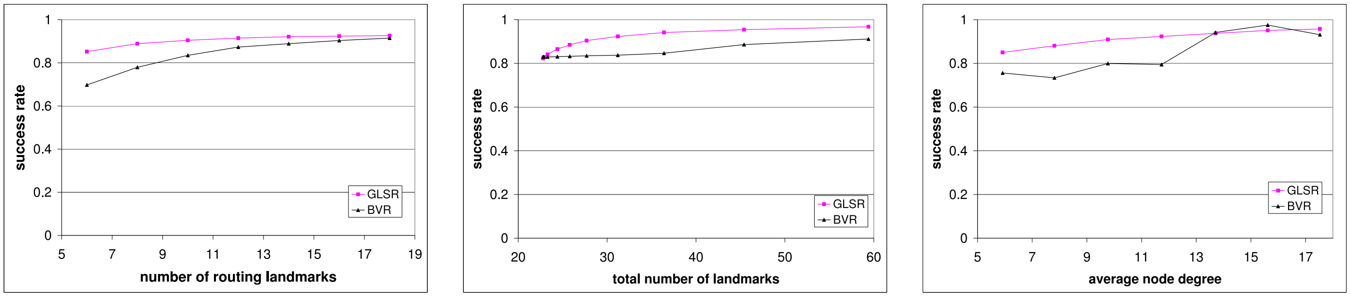

Fig. 7. The success rate of GLDR is high for various settings. GLDR compares favorably to BVR.

In all cases, the success rate of GLDR was consistently over 80%, and frequently over 90%, which is good, considering the relatively low degree of the network.

C. Remaining paths and stretch factors

We varied the density of the landmarks by changing the landmark separation distance parameter between 6 and 18 in the landmark selection algorithm. For pairs of nodes on which the routing got stuck, we computed the hop count from the terminating nodes to the destinations. This measure gave us some idea on how much we need to flood in order to reach the destination. When we could route successfully, we also computed the shortest path between the source and the destination, and computed the stretch factor. The average remaining path (Figure 8) was comparable with the separation distance between landmarks. The stretch factor of the path (Figure 9) was quite small, between 1.04 and 1.09. A typical route was about 30 hops, which means that our routes were only $1-3$ hops longer than the shortest paths.

GLDR compared favorably to BVR in these measures.

D. Landmark selection

We measured the effectiveness of the landmark selection algorithm in GLDR. For this purpose, we ran the selection algorithms for various values of K, see Table I.

As K increased, the number of landmarks decreased, as landmarks had more time to suppress nearby nodes, preventing them from becoming landmarks. From Table I, we noted that when K increased from 600 to $3,000$, the number of landmarks increased, but the routing performance stayed relatively constant. This suggests that tightly clustered landmarks might not contribute much to the routing quality, and that the K value should not be too small. On the other hand, when K increased above $1,200$, the number of landmarks started to stabilize. As the network initialization time was dominated by K, K should be small, at the expense of having more landmarks in the network.

We compared the performance of GLDR with landmarks chosen uniformly at random, and according to II-C. The success rate degraded significantly with random landmarks, as there were more nodes that are not "surrounded" well by landmarks. Thus, our landmark selection is crucial for effectiveness of GLDR.

E. Improving BVR with r-sampling landmark selection

Our landmark selection scheme can be used to improve other landmark-based algorithms. For example, it helped improve the success rate of BVR (Figure 10). The performance of BVR with our landmark selection scheme is very similar to GLDR.

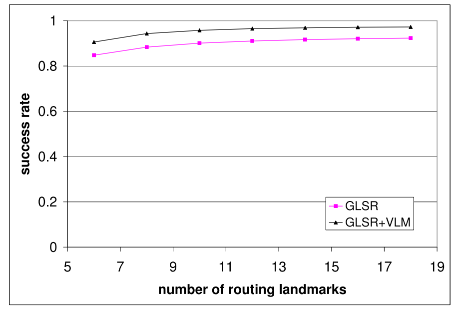

Fig. 11. Virtual landmarks improve routing performance.

F. Improving routing performance using virtual landmarks

We tested the idea of virtual nodes and landmarks for routing to nodes near the sensor field boundary (Section III-D). To simplify the implementation, we selected virtual landmarks to coincide with real landmarks which are near the boundary (instead of choosing them in the same way as in the base case, only using a different notion of distance for some landmark pairs). We found that the virtual landmarks improve the routing quality of GLDR, see Figure 11.

V. Conclusion

We proposed a practical landmark selection protocol that uniformly samples a network of sensors. Combined with a simple greedy routing based on distances to landmarks, we can achieve good routing performance which is provable in the continuous domain and manifested in our extensive network simulations. GLDR, while comparable to BVR in performance categories such as packet delivery rate and routing path length, has lower routing overhead. Further, the landmark selection algorithm is also beneficial to some other landmark-based routing schemes. Our experiments showed that BVR's performance improves significantly with a set of landmarks/beacons selected using this algorithm.

Landmark selection is seldom a stand-alone problem, but rather strongly tied to the overlying application, be it point-to-point network routing or robot navigation. This paper provides a practical solution in the context of routing in wireless ad-hoc networks. The general problem remains important and challenging. We hope this paper could encourage more research interest in this direction in the networking community.

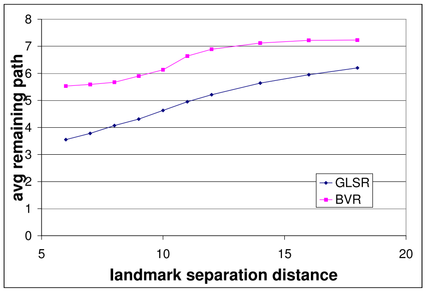

Fig. 8. Unsuccessful routes ended up near the destination.

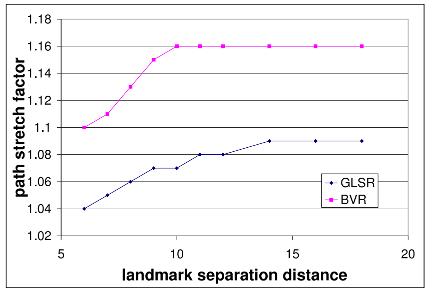

Fig. 9. Successful routes had low stretch factor.

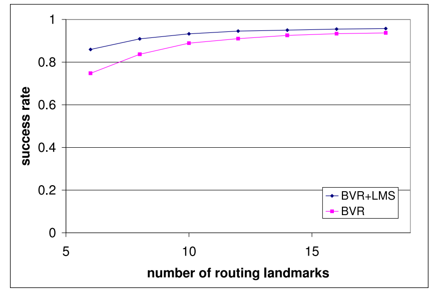

Fig. 10. Good landmark selection scheme can be used to improve the performance of BVR.

TABLE I
THERE ARE FEWER LANDMARKS WHEN K INCREASES.

K	300	600	900	1200	1500	1800	2100	2400	2700	3000
#LM	57.1	37.5	30.7	27.6	25.8	24.5	23.7	23.2	22.8	22.4
GLDR	0.921	0.904	0.901	0.904	0.906	0.904	0.906	0.906	0.905	0.905
No LMS	0.861	0.841	0.831	0.832	0.828	0.822	0.814	0.808	0.805	0.800

REFERENCES

[1] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," *Wireless Networks*, vol. 7, no. 6, pp. 609–616, 2001. [Online]. Available: citeseer.ist.psu.edu/bose01routing.html

[2] B. Karp and H. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, 2000, pp. 243–254.

[3] F. Kuhn, R. Wattenhofer, Y. Zhang, and A. Zollinger, "Geometric ad-hoc routing: Of theory and practice," in *Proc. 22^{nd} ACM Int. Symposium on the Principles of Distributed Computing (PODC)*, 2003, pp. 63–72.

[4] Y.-J. Kim, R. Govindan, B. Karp, and S. Shenker, "Geographic routing made practical," in *Proceedings of the 2nd USENIX Symposium on Networked Systems Design and Implementation (NSDI '05)*, Boston, MA, USA, May 2005.

[5] B. K. S. S. Young-Jin Kim, Ramesh Govindan, "On the pitfalls of geographic face routing," in *Proceedings of the Third ACM/SIGMOBILE International Workshop on Foundations of Mobile Computing*, September 2005.

[6] A. Rao, S. Ratnasamy, C. Papadimitriou, S. Shenker, and I. Stoica, "Geographic routing without location information," in *Proc. 9th Annual International Conference on Mobile Computing and Networking (MobiCom 2003)*. San Diego: ACM Press, September 2003, pp. 96–108.

[7] Q. Fang, J. Gao, L. Guibas, V. de Silva, and L. Zhang, "GLIDER: Gradient landmark-based distributed routing for sensor networks," in *Proc. of the 24th Conference of the IEEE Communication Society (INFOCOM)*, March 2005.

[8] R. Fonseca, S. Ratnasamy, J. Zhao, C. T. Ee, D. Culler, S. Shenker, and I. Stoica, "Beacon vector routing: Scalable point-to-point routing in wireless sensornets," 2005.

[9] A. Caruso, A. Urpi, S. Chessa, and S. De, "GPS free coordinate assignment and routing in wireless sensor networks," in *Proc. of the 24th Conference of the IEEE Communication Society (INFOCOM)*, March 2005.

[10] R. Williams, *The Geometrical Foundation of Natural Structure: A Source Book of Design*. New York: Dover Publications, 1979.

[11] R. Motwani and P. Raghavan, *Randomized Algorithms.* Cambridge University Press, 1995.

[12] A. Kroller, S. Fekete, D. Pfisterer, and S. Fischer, "Deterministic boundary recognition and topology extraction for large sensor networks," in *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*. New York, NY, USA: ACM Press, 2006, pp. 1000–1009.

[13] Q. Fang, J. Gao, and L. J. Guibas, "Locating and bypassing holes in sensor networks," *Mobile Networks and Applications*, vol. 11, no. 2, pp. 187–200, 2006.

[14] N. Amenta and M. W. Bern, "Surface reconstruction by Voronoi filtering," in *Symposium on Computational Geometry*, 1998, pp. 39–48. [Online]. Available: citeseer.ist.psu.edu/article/amenta98surface.html

VI. APPENDIX

A. Proof of Lemma 2

Consider any node c. Let T be the number of landmarks in $B(r,c)$. Suppose that $T \geq t$, where $t \geq t^* := \frac{\beta}{\alpha}(3+\sqrt{\alpha})^2$. Let $2\rho+1$ be the smallest odd upper bound on the minimum inter-landmark distance. ρ-neighborhoods of the landmarks are disjoint, and contained in $B(c, r+\rho)$. It follows $\rho \leq \frac{r}{\sqrt{\alpha t/\beta - 1}}$, by a standard packing argument. Using $t \leq \beta r^2$ we find $1 \leq \sqrt{\alpha} r / \left(\sqrt{\alpha t/\beta} - 1\right)$, hence

$$2\rho + 1 \leq \frac{(2+\sqrt{\alpha})r}{\sqrt{\alpha t/\beta} - 1}\,. \tag{1}$$

Since $t \geq t^*$, the right hand side of (1) is at most r, so the two landmarks p and q which realize the minimum inter-landmark distance are at most r hops from each other. Hence

$$|v_p - v_q| \leq \frac{(2+\sqrt{\alpha})r}{K(\sqrt{\alpha t/\beta} - 1)}\,. \tag{2}$$

The probability of this event is at most the value of the right hand side (which is at most 1, assuming $K \geq r$). By definition, $\mathrm{E}[T] \leq (t^* - 1) + \sum_{t=t^*}^{\beta r^2} \Pr[T \geq t]$. We bound the sum using (2).

$$\begin{aligned} \sum_{t=t^*}^{\beta r^2} \frac{(2+\sqrt{\alpha})r}{K(\sqrt{\alpha t/\beta} - 1)} &\leq \frac{(2+\sqrt{\alpha})r}{K} \int_{t=t^*}^{\beta r^2} \frac{dt}{\sqrt{\alpha t/\beta} - 1} \\ &\leq \frac{2\beta(\frac{2}{\sqrt{\alpha}}+1)r^2}{K} \left(1 + \frac{\ln(r\sqrt{\alpha})}{r\sqrt{\alpha}}\right)\,. \end{aligned}$$

We can simplify the expression using $r \geq 1$, $\alpha \geq 1$, and $\frac{\ln x}{x} \leq \frac{1}{2}$ (for $x > 0$), to obtain the desired result.

B. Proof of Lemma 3

Lemma 4: $\kappa \leq \sqrt{\rho/(\rho-1)}$.

Proof: Without loss of generality assume $|sd| = 1$. Let $|dl| = x$ and $\angle sld = \alpha$. It follows $|sl| = \rho x$ and $|s'd| = 2x\sin(\alpha/2)$ (see Figure 4). We have

$$\begin{aligned} \kappa &= \frac{(\rho-1)x}{1 - 2x\sin\frac{\alpha}{2}} && \text{(definition of } \kappa)\,, \\ 1 &= \rho^2 x^2 + x^2 - 2\rho x^2\left[1 - 2\sin^2\frac{\alpha}{2}\right] && \text{(cosine law } \triangle sdl)\,. \end{aligned}$$

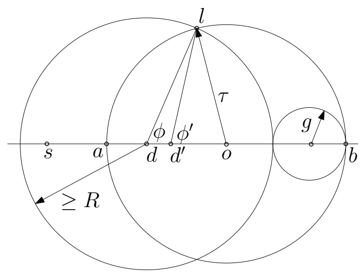

Fig. 12. Bounding the one-step stretch factor.

Eliminating $\sin(\alpha/2)$,

$$\kappa = \frac{(\rho-1)x}{1-\sqrt{(1-(\rho-1)^2x^2)/\rho}}\,. \tag{3}$$

Triangle inequality $|\rho x - 1| \le x$ implies $x \in [\frac{1}{\rho+1}, \frac{1}{\rho-1}]$. Tedious but straightforward calculations show that the right hand side of (3) is minimized over this interval for $x = 1/\sqrt{\rho(\rho-1)}$, at which point $\kappa = \sqrt{\rho/(\rho-1)}$. ■

Lemma 5: If $\phi < \pi/2$, then $\kappa \le 1/(1-\sin\phi)$.

Proof: Let s''' be such that ss''' is parallel to dL and $s'''d$ is perpendicular to both ss''' and dL. Let s'' be the intersection of $s'''d$ and ss', see Figure 4. Note that $\angle ds'L = \angle s'dL < \pi/2$, and when when $\phi < \pi/2$, it is easy to see that s'' is between s''' and d, and s'' is also between s and s'.

From triangle inequality, $|ss'| = |sL| - |dL| \le |sd|$. We also have that $|s'd| \le |s''d| \le |s'''d| = |sd|\sin\phi$. Thus $\kappa = |ss'|/(|sd| - |s'd|) \le 1/(1-\sin\phi)$. ■

Note that as s approaches d, ρ approaches 1 and ϕ may or may not be acute. Thus both Lemma 4 and 5 do not yield any bound on κ. We now restrict the routing landmarks of d to only landmarks that are further than some threshold distance R from d. We also assume that we select the landmark L such that the ratio ρ for L is the largest among those.

Let o be the center and τ be the radius of the Apolonius circle of sd with ratio ρ, the circle $\mathcal{C}$ of all points p satisfying $|ps|/|pd| = |ls|/|ld| = \rho$. It is well-known that o is on the line sd. Let a and b be the intersection of this circle with sd, with a is between sd and b is outside (Figure 12). By the choice of l, there is no routing landmark of d inside $\mathcal{C}$.

From $\rho = |ss|/|sd| = |bs|/|bd|$, it is easy to see that $|da|/|db| = (\rho-1)(\rho+1) < 1$, and thus d is between a and o. By Lemma 4, $|da|/|db| \le 1/(2\kappa^2 - 1) = (1-u)/(1+u)$, and thus $|da| \le \tau(1-u)$, $|do| \ge \tau u$, $|db| \ge \tau(1+u)$.

Let d' be between d and o such that $|d'b| = \tau(1+u)$. Let $g = (|db| - |dl|)/2$. Since there is a disk of radius g empty of landmarks (Figure 12), we have $r \ge g \ge (|d'b| - |d'l|)/2 = \tau(1+u-x)/2$, where $x = |d'l|/\tau$ can be bounded as follows.

Lemma 6: For w as in the statement of Lemma 3, $x \le \sqrt{w}$.

Proof: Suppose $x > \sqrt{w}$. Then

$$\cos\phi \ge \cos\phi' = \frac{x^2 - (1-u^2)}{2x} > 0\,,$$

so $\phi < \pi/2$. By Lemma 5, $\sin\phi \ge 1 - 1/\kappa = v$, and thus

$$\sin^2\phi + \cos^2\phi \ge \frac{(x^2-(1-u^2))^2}{4x^2} + v^2 > 1\,,$$

a contradiction. ■

Since $R \le |dl| \le 2\tau$, $r \ge R(1+u-x)/4$. From Lemma 6, $R/r \le 4/(1+u-\sqrt{w}) =: f(\kappa)$.

C. Proof of Theorem 5

We give only the proof for the case when s, d, s' are real and l is virtual (Figure 13), the other cases are similar.

Let x and y be the boundary points where the shortest paths sl and dl cross the boundary. Let l' be such that $|sl'| = \langle sl\rangle$ and $|dl'| = \langle dl\rangle$

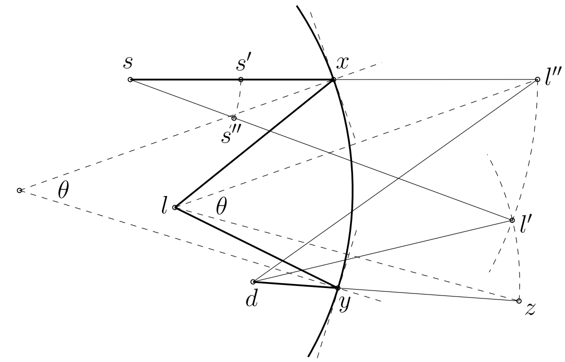

Fig. 13. Analysis of the augmented-GLDR when s, d, s' are real, and l is virtual.

(in general, there are two such points symmetric with respect to the line sd, but we define l' the be the one on the same side of the line as the local boundary).

Notice that l was picked as a routing landmark because the distances (possibly reflected) among s, d and l satisfy certain relations. By construction, s, d and l' preserve this set of distances. It follows that if the base case protocol is applied to s, d and l', then Theorem 4 provides a stretch guarantee, but for a "new source position" s'', which is different than s' in general. The "true" one-step stretch is

$$\frac{|ss'|}{|ds| - |ds'|} \le \frac{|ss''| + |s's''|}{|ds| - |ds''| - |s's''|} \le \frac{|ss''| + |s's''|}{\frac{|ss''|}{\kappa} - |s's''|}\,, \tag{4}$$

which tends to κ as $|s's''|/|ss''|$ becomes smaller. Thus if we prove that $|s's''|/|ss'| \ll 1$, we are done.

We bound the ratio as follows. Let θ be the (small) angle between the boundary normals at x and y. Let l'' and z be the mirror images of l about the tangents to the boundary at x and y, respectively. We find that $|dl''| \approx \langle dl\rangle$, specifically

$$\begin{aligned} & ||dl''| - |dz|| \le |zl''| \text{ (triangle inequality)} \\ \Rightarrow\quad & ||dl''| - \langle dl\rangle| \;\le\; \max\{|ll''|, |lz|\}\theta \;\le\; 2\langle sl\rangle\theta\,. \end{aligned}$$

Now, $\triangle sdl''$ and $\triangle sdl'$ have two pairs of sides of the same lengths ($|sd|$ and $\langle sl\rangle$) and one which differs only slightly (at most $2\langle sl\rangle\theta$). The small change in the length implies a small change in the internal angles, more precisely

$$\angle l'sl'' \le 2\cdot\frac{||dl''| - |dl'||}{|sl''|} \le 2\cdot\frac{2\langle sl\rangle\theta}{\langle sl\rangle} \le 4\theta\,.$$

(multiplying the fraction by 2 is far too sloppy, but illustrates the point). Hence $|s's''| \le 4\theta\cdot|ss'|$.

So we have bounded the ratio by a multiple of the angle θ between the normals, which we expect to be small if the curvature is low. Formally, we know that x and y are close – at most $8r$ apart (both are within $4r$ from l) – and we invoke the following well known geometric lemma.

Lemma 7 (Amenta and Bern [14]): The angle between the normals whose footpoints are at most $\rho \le R/4$ apart is at most $4\rho/R$. It follows that if $R \ge 320r$, then $\theta \le 4\cdot 8r/(320r)$, hence $|s's''|/|ss'| \le 0.4$. Substituting into (4) we find that the one-step stretch is bounded.

Notice that the proof so far did not assume (or require) guaranteed delivery, as it was concerned only with a single routing step. However, we can use the single-step analysis to prove that even in the augmented case the destination is always reached. As in the one-sided case, it suffices to prove, in the same setting as above, that $|sd| > |s''d|$. But this simply follows by the following chain of inequalities

$$|sd| - |s''d| \ge |sd| - |s'd| - |s's''| \ge (\kappa - 4\theta)|ss'|\,,$$

which is positive, under our low curvature requirement.

Once we have the delivery guarantee, we can extend the one-step stretch bound to the entire path. This completes the proof.