Spiral-based data dissemination in sensor networks

Yang Sun
Center for Wireless Communications,
The University of Mississippi, University, MS 38677 USA
E-mail: yangsun@olemiss.edu

Guangbin Fan*
Department of Computer Science,
The University of Mississippi, University, MS 38677 USA
E-mail: gfan@cs.olemiss.edu
*Corresponding author

Shigang Chen
Department of Computer and Information Science and Engineering,
University of Florida, Gainesville, FL 32611-6120 USA
E-mail: sgchen@cise.ufl.edu

Abstract: Data dissemination is a key problem that bottlenecks the wide application of wireless sensor network (WSN). In this paper, a novel data dissemination scheme−logarithmic spiral data dissemination (LSDD)−is proposed. In LSDD, data advertisements are disseminated following a spiral-like path, which involve only a small fraction of nodes in a sensor network. By exploiting nice features of spiral, this scheme scales well for large sensor networks while prolonging the network lifetime. In numerical analysis and simulations, we show the distinct merits of LSDD as follows: lower dissemination cost ($O(\sqrt{n})$ compared to that of flooding-based schemes ($O(n)$), controllable topology by spiral parameters, excellent scalability, and good fault tolerance. We evaluate the extra delay caused by LSDD as a trade-off for its advantages, which is shown to be within a tolerable range.

Keywords: Sensor Network, Spiral; Data Dissemination; Energy Efficiency; Scalability.

Reference to this paper should be made as follows: Sun, Y., Fan, G. and Chen, S. (2005) 'Spiral-based data dissemination in sensor networks', *IJAHUC, special issue in "Pervasive Computing through Networked Sensing Devices."*

Biographical notes: Yang Sun received his M.S.E.E in Telecommunications from the University of Mississippi in 2004. He is now a Ph.D. candidate in the center for wireless communications at the University of Mississippi. His research interests include network traffic analysis and modeling, collaborative signal processing, self-localization, self-clustering, data dissemination, and other issues in wireless sensor networks.

Dr. Guangbin Fan is an assistant professor in the Department of Computer and Information Science at the University of Mississippi. He received his Ph.D. degree in Computer Science from the University of Alabama in 2003. His research interests include mixed networking, mobility, security, and sensor networks. He currently serves on the editorial board of IEEE Communications Surveys and Tutorials, and several technical program committees of International conferences.

Shigang Chen received his Ph.D. degree in computer science from University of Illinois at Urbana-Champaign in 1999, respectively. After graduation, he had worked with Cisco Systems for three years before joining University of Florida as an assistant professor in 2002. His research interests include network security, quality of service, and sensor networks.

1 INTRODUCTION

With the advances of VLSI, MEMS and wireless technology, it is practical to integrate sensor component, microprocessor, flash ROM, RF transceiver unit and battery into a box or a button as small as one cubic inch or even less (Hill et al. 2000). A sensor network (Kahn et al. 1999) is constructed with a large number of such tiny and smart sensor nodes in an ad-hoc manner. The sensor network has great potential to be applied in battle field surveillance, forest fire monitoring, natural habitat recording to intrusion detection, etc. One of the key problems faced by sensor network is data dissemination: an *observer* needs to spread his interests over the whole network, while a sensor node needs to notify the observer when interested phenomena occur. Usually, some *sink* nodes are deployed in a sensor network to act as agents between observers and sensor nodes. Thus, the problem is to find an efficient and robust data dissemination scheme between sensor nodes and sink nodes.

In the past decade, quite a few data dissemination schemes have been proposed (Tilak et al. 2002). However, most solutions for data dissemination depend on the global or local flooding to construct and maintain dissemination paths. As is well known, flooding lead to low energy efficiency and high channel congestion, which is not preferred in a sensor network. In addition, most proposed schemes use the local information of each sensor node to construct the dissemination paths. That is, a sensor node is assumed to know the location or routing information of the neighbor nodes which lie within a limited range. However, due to the ad-hoc nature of the sensor network, the resulting path under such assumption is suboptimal, unstable, and the energy consumed on path construction is unpredictable.

In this paper, we propose an energy efficient and scalable data dissemination scheme, logarithmic spiral data dissemination (LSDD), based on the local geographic information at each node as well as the logarithmic spiral geometrics. In LSDD, data advertisements are disseminated following a spiral-like path, which simulates the natural progression of knowledge propagation while involving only a small fraction of nodes in a sensor network. We also propose a complementary reversed spiral searching algorithm for queries. Compared with previous schemes, our solution provides a more predictable and stable network topology, and consequently reduces considerable traffic overhead on dissemination path construction and query/response process. Without full flooding, the channel congestion is alleviated a lot. This feature benefits the majority of WSNs, in which one channel is shared by all sensor nodes. Both numerical analysis and simulations have been conducted in this paper. In the numerical analysis, we show that LSDD is able to cover larger area with fewer nodes than the flooding-based schemes, and consequently incurs much less overhead. In addition, it is shown that the larger coverage, the more efficient LSDD can achieve than other schemes. We also derive the relationship between the spiral parameters and the dissemination ratio.

Extensive simulations have been done to evaluate the performance of LSDD from various aspects. It is compared with the other two data dissemination schemes: Sink Initiated Data Dissemination (SIDD) and Gossiping Based Data Dissemination (GBDD). Note that SIDD is actually a flooding based scheme. Besides verifying the results of numerical analysis, our simulations also show that LSDD outperforms SIDD and GBDD on scalability. The search cost of LSDD increases linearly with respect to the source-sink distance. On the contrary, the costs of SIDD and GBDD increase exponentially in both tests. When there are multiple sink nodes, the search cost of LSDD is also linear with the number of sink nodes, and at least one order of magnitude lower than SIDD and GBDD. On the other hand, LSDD is very robust against unreliable sensor nodes, which are quite common in practice. In our simulation, even when 10 percent of nodes fail, LSDD achieves a success rate higher than 90 percent on average.

In the rest of this paper, the *sensor node* is referred to as any member node in a sensor network, the observer is referred to as any application, system, or human being that queries the interested phenomena and receives the aggregated/processed results from the sensor network. The *sink node* is referred to as a communication unit relaying information between observers and the sensor network, which may be a sensor node, a base station, or a laptop carried by an observer. The functions of a sink node include initiating queries and aggregating and reporting the gathered data. Our studies presented in this paper are based on the following assumptions.

1. Flat plane: the sensor network is deployed in a flat plane, where a point in that plane can be expressed in Cartesian coordinates as (x, y) or polar coordinates as (r, θ), where
$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$$
2. The sensor nodes are densely deployed on a convex area following a uniform distribution.
3. Local geographic information: Each node is assumed to know the locations of itself and its one-hop neighbor nodes. A sensor node can acquire the location information from an embedded GPS-like unit or from other reference nodes by using some location algorithm.
4. Static network topology: every sensor node is static in the initial position, or moves within such a small area that it can be viewed as static regarding to its effective radio range, and its neighborhood list do not change. As to the case considering sensor nodes mobility, we leave it to further study.
5. Normalized query/response packet size: we assume that all messages for query and response are aggregated to the same size.

6. Symmetric links: if a sensor node can receive messages from another node, then it is also able to send message to that node.

The rest of this paper is organized as follows. In Section II, we briefly review the main proposed data dissemination schemes and protocols and summarize their advantages and disadvantages. In Section III, we introduce the logarithmic spiral data dissemination (LSDD), including a short review of the geometrics of the logarithmic spiral, the construction and maintenance of a spiral path and the query/response procedures. In Section IV, we provide a numerical analysis on the traffic overhead and scalability of LSDD. In Section V, we introduce our simulation scenarios and settings, present the simulation results, and compare LSDD with other alternative schemes. In Section VI, our conclusion is presented.

2 OVERVIEW OF DATA DISSEMINATION

As wireless sensor networks receive more and more attention from the communication society in recent years, many data dissemination schemes and protocols are proposed. Classical flooding is the simplest and most common method. Sink nodes look for the path to interested data by flooding, and the sensor nodes passively wait for the query. However, tremendous traffic overhead, due to the ad-hoc nature of sensor networks and replies from multiple sensor nodes, motivates people to search for more efficient solutions. In gossiping (Haas et al. 2002), flooding is limited by randomly forwarding messages to one of the neighbor nodes at each hop instead of full flooding. That feature helps save the energy, however, could result in unreliable and unmanageable paths. Rumor routing (Braginsky and Estrin 2002) avoids using flooding at all: a sensor node which detects interests propagates its result along a random path, while a sink node keeps sending its query along random paths too. Both keep searching until their paths meet. This scheme eliminates most disadvantages with flooding. However, since every time only one path is searched for one node, the total cost to establish a dissemination path may still be high, especially in a network with high node density. SPIN-1 and SPIN-2 are early content-based data dissemination schemes proposed in Heinzelman et al. (1999). SPIN-1 employs the meta-data to aggregate data and reduces the redundant traffic among different sensor nodes. SPIN-2 adds the energy constraint to SPIN-1. Directed Diffusion was proposed as a data-centric approach by Intanagonwiwat et. al. (2000), and a modified version optimized for higher network density is presented in (Intanagonwiwat 2002). Directed Diffusion is similar to SPIN in that data are named and cached in the sensor nodes like meta-data. Multiple versions about the same data are sent back to the sink node at low data rate, then data are aggregated in the nodes along the multi-path, and the best version is chosen to transfer at high data rate while other version are discontinued. However, flooding is still the underlying method for communications in both SPIN and directed diffusion. In SPIN, the aggregated data will rebroadcast as new data to the network, which also offsets the traffic reduction by data aggregation. Both of them have no effective ways to manage the topology of the dissemination paths. Therefore, those paths may be far from optimal in global view although it may be optimal at every hop.

Energy efficiency is another important issue for WSN, which should be considered when designing the routing protocols or data dissemination schemes. Li et al. (2002) proposed three heuristic algorithms to approximate a broadcasting tree in a static sensor network by which the broadcast energy is minimized. However, all three algorithms are centralized, therefore it is not very efficient when the resource required by information gathering is considered. Yu et al. (2002) proposed GEAR (Geographical and Energy Aware Routing), which chooses routes depending on the local knowledge of node energy and coordinates. It can lower the traffic overhead by limiting the flooding in a specific destination region, since each node is assumed to know its own location. However, another important assumption is that the destination must be known *priori*. This means GEAR cannot be used when the sink nodes have no knowledge where the interested sensor nodes are located, which limits the usefulness of GEAR.

Generally, any scheme depending on full flooding, local flooding, or any other kinds of flooding will suffer the fast-increased overhead from either the larger network size or the higher network density. A high traffic overhead means more energy per node spent on RF transmission, which is not preferred in an energy-constrained system like sensor network. It also leads to higher probability for packet collision in a wireless channel, longer delay, and more unstable response. Without global knowledge of the whole network, it is hard for sensor nodes and sink nodes to establish stable and optimal dissemination paths by only neighborhood knowledge and localized algorithms. However, it is unpractical to require every node in the sensor network to learn the whole network topology, which may consume a large amount of memory and much energy on computation and communications.

A related work is Trajectory Based Forwarding (TBF) (Niculescu and Nath 2003). TBF is proposed as a hybrid of source routing and Cartesian geographic forwarding. The authors described the general procedure of TBF, enumerated the potential applications of TBF, pointed out some possible adverse conditions for TBF, and suggested some open issues for TBF. However, the author did not provide any specific scheme based on a particular trajectory. Besides, more numerical analysis and simulation results are necessary to show the potential advantages of TBF.

In the next section, we introduce a novel data dissemination model–the logarithmic spiral data dissemination model (LSDD), which pursues the solution from a new perspective. LSDD promises to achieve a better performance on the energy efficiency, network scalability, and data redundancy.

3 LSDD SCHEME

Unlike all other data dissemination schemes, under LSDD, sensor nodes forward their advertisement messages along the path approximating a logarithmic spiral curve, while a sink node forwards its query message along the path approximating a logarithmic spiral curve reverse to the advertisement path. As a beautiful curve, spiral exists in natural for thousands of centuries, for example, nautilus and the pattern of sunflower seeds. There are many variants of spiral curve, such as golden mean spiral and equiangular spiral and so on. For LSDD, we choose the logarithmic spiral because of its versatility as well as its succinct parameter set and definition, which is very suitable for the resource-limited sensor nodes.

To focus on the LSDD itself, we assume that the parameters of the spiral are broadcasted or preprogrammed to every sensor node in advance. When a sensor node detects something interesting that needs to be disseminated, it initiates the spiral dissemination. It first runs the spiral path search algorithm (SPSA) to choose the next-hop neighbor node which fits the intended spiral path best, then sends an advertisement message to the chosen sensor node. Besides the signature of the interested phenomenon and the location of the source node (here "source node" refers to the sensor node which initiates a data dissemination), the advertisement packet also includes the location and id of the previous hop node, the spiral angle of the previous hop node, and other parameters like TTL (Time-To-Live) of the advertisement, the maximum hop number of the dissemination path, etc.

When a sensor node receives an advertisement packet, it first makes a local copy of this advertisement, then uses the same SPSA to choose one neighbor as the next hop, and forwards the advertisement message. In this way, the advertisement is forwarded hop by hop in the sensor network following a spiral-like track. Such dissemination will end when some preset condition is met, for example, there is no neighbor node along the spiral path, the spiral reaches the boundary, or the hop limit is reached.

The query procedure is similar to the dissemination procedure but in a reverse direction. A sink node initiates a query, and the query follows the reverse spiral path until it meets the dissemination spiral, or the termination condition is satisfied, for example, the maximum hop number is reached, boundary is reached or there is no node to choose. Before forwarding the query to the next-hop, the sensor node will also broadcast a message to its neighborhood to see if it meets the dissemination path of the desired knowledge.

The source node may periodically update the information along the spiral path, or work in a spontaneous mode that launch the spiral dissemination whenever the interested phenomenon is detected. To avoid redundant traffic, all sensor nodes will drop an advertisement or query packet unless it is newer than their local copy.

For a better understanding, we review some fundamental knowledge on the spiral geometrics in the next section.

3.1 Logarithmic Spiral

The logarithmic spiral (Archibald 1918) (also called as Bernoulli spiral, Fibonacci spiral) was first studied by Rene Descartes (1638). Torricelli worked on the curve independently, and found the curve's length. An illustration of the logarithmic spiral is shown in Figure 1.

In a two-dimension plane under polar coordinates, a logarithmic spiral is defined as a curve such that:

$$r = ae^{b(\theta-\theta_0)}, \tag{1}$$

where the radius r is the distance from origin to the point with angle θ, a and b are arbitrary positive constants, and θ_0 is the initial spiral angle. From equation (1), it's simple to show that for a point (r, θ) on the spiral specified by a and b, given θ_0 and r, the angle of the point is

$$\theta = \theta_0 + \frac{\log(r/a)}{b}. \tag{2}$$

Under the Cartesian coordinates, equation (1) converts to:

$$\begin{cases} x = a\cos[b(\theta)]e^{b(\theta-\theta_0)} \\ y = a\sin[b(\theta)]e^{b(\theta-\theta_0)} \end{cases} \tag{3}$$

The reverse spiral is defined by

$$r = ae^{-b(\theta-\theta_0)}. \tag{4}$$

A reverse spiral is shown in Figure 2.

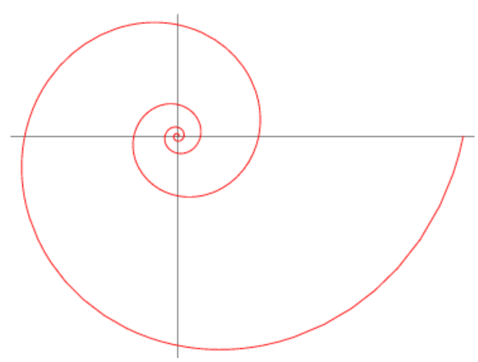

Figure 1: Logarithmic Spiral

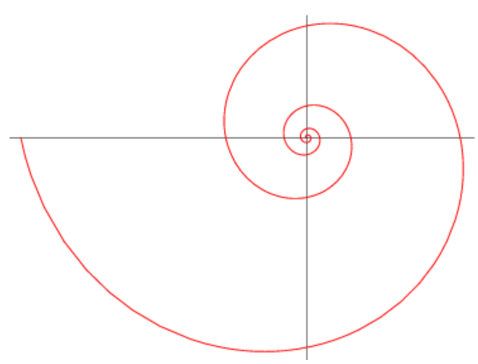

Figure 2: Reversed Logarithmic Spiral

According the above definition, the reverse logarithmic spiral is identical with the normal logarithmic spiral except that it turns in the opposite direction. Therefore, all properties of the reverse logarithmic spiral are similar to the normal one, and all parameters are calculated in a similar way as the above equations.

3.2 Spiral Path Search Algorithm

One crucial problem in the spiral dissemination is how every sensor node in the dissemination path finds the next hop so that the whole path approaches a spiral curve given the parameters of the spiral, the location information of itself and its first-hop neighbor nodes. Here we provide a localized heuristic employing the linear programming (Winston 1993) to solve this problem.

This problem can be formally stated as: at any sensor node known as the current hop of a spiral dissemination path, how to find the sensor node for the next hop along the same path within the one-hop neighborhood? Note that "current node" is referred to as the node for the current hop, "previous node" is referred to as the node for the previous hop, and "next node" is referred to as the node for the next hop. Assume that the current node knows the locations of all its one-hop neighbor nodes including the previous node, the parameters of the spiral and the coordinates of the original node, which are forwarded by the previous node. An illustration of this problem is shown in Figure 3. Let P_i denote the sensor node for the ith hop in a spiral dissemination path, and P_i^j denote the jth one-hop neighbor node. Suppose that there are totally N one-hop neighbor nodes, the neighborhood set can be written as $Neighbor_i = \{P_i^j | j = 1, 2 \ldots, N\}$. Let $R(\cdot)$ be a function to return the distance between a sensor node and the original node, $SPA(\cdot)$ be a function to return the spiral angle of a node given the original node, $SPR(\cdot)$ be a function to return the spiral radius of an angle, and $D(\cdot)$ be a function to return the distance from a node to the point on a ideal spiral which corresponds the node's spiral angle, where the spiral starts from original node. Note that the $SPR(\cdot)$ is the same as equation (1), and $SPA(\cdot)$ is based on equation (2), and

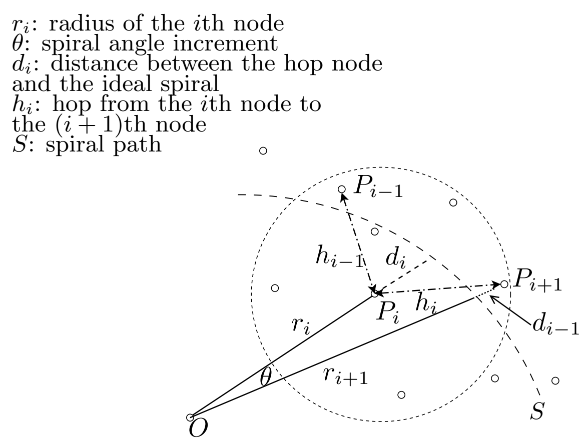

Figure 3: Spiral Dissemination Path Estimation

$$D(x) = \|SPR(SPA(x)) - R(x)\|,$$

where x denotes a node. The difference of the spiral angles of two successive nodes indicates how much the spiral path advances, and the distance from a node to the closest point on the spiral shows how much the path approximates the ideal spiral. Let K_a denote the weight on the spiral angle and K_d denote the weight on the distance, where both K_a and K_d are arbitrary constants.

This path search problem at ith hop can be formulated to an linear programming (LP1) as follows:

$$\text{Max} \sum_{j=1}^{N} x_j (k_a SPA(P_i^j) + \frac{k_d}{D(P_i^j)})$$

subject to

$\sum_{j=1}^{N} x_j = 1.$ (a)

$$x_j = \begin{cases} 1 \text{ if } P_i^j \text{ is chosen} \\ 0 \text{ else} \end{cases} \quad j = 1, 2, \ldots, N. \text{ (b)}$$

$x_j \leq I(SPA(P_i^j) > SPA(P_i))$,
where $I(\cdot)$ is the indicator function (c)

In the above formulating, the weighted sum of the interval spiral angle and the distance to the ideal spiral is used as the cost function. Another linear programming (LP2) can be formulated by changing the cost function to the ratio of the above two terms. That is, the objective in LP2 can be written as

$$\text{Max} \sum_{j=1}^{N} x_j (\frac{SPA(P_i)}{D(P_i^j)})$$

subject to

$\sum_{j=1}^{N} x_j = 1.$ (a)

$$x_j = \begin{cases} 1 \text{ if } P_i^j \text{ is chosen} \\ 0 \text{ else} \end{cases} \quad j = 1, 2, \ldots, N. \text{ (b)}$$

$x_j \leq I(SPA(P_i^j) > SPA(P_i))$,
where $I(\cdot)$ is the indicator function (c)

In both of the LP1 and LP2, constraint (c) filters out all the neighbor nodes with spiral angle less than the current node, and constraint (a) and (b) limit that only one node can be chosen. In the cost function, we make a trade-off between the path advance and the spiral approximation, and this trade-off is adjustable in LP1 by changing the values of K_a and K_d. Our simulation shows that both LP1 and LP2 work very well. Since there is little difference between the paths found by LP1 and LP2, we choose LP1 for SPSA and fix the weight values as $K_a = 1$ and $K_d = 2$. Figure 4 shows an sample path found by SPSA, and Figure 5 shows an sample path with boundary in a limited area , which is found by an extension of SPSA (ESPSA).

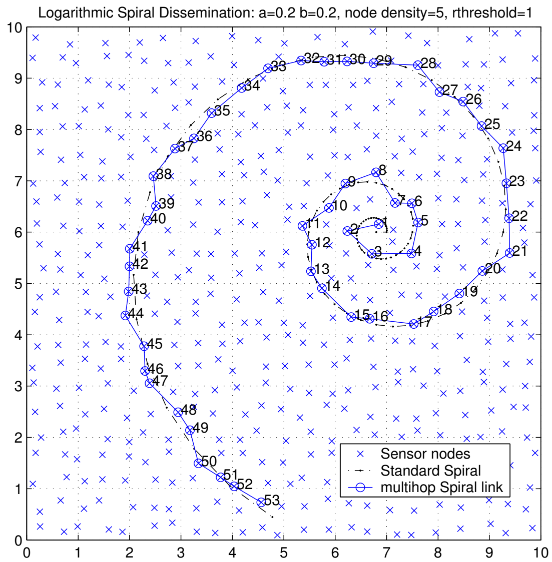

Figure 4: Spiral Dissemination Path

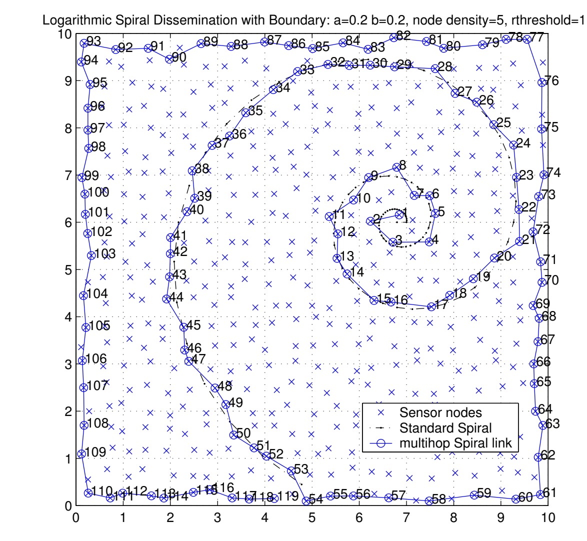

Figure 5: Spiral Dissemination Path with boundaries

3.3 Reverse Spiral Search Procedure

As stated before, the reverse spiral is used for the sink node to spread the query packet across the whole sensor network. Before a sink node initiates a query, it uses the reverse spiral path searching algorithm (RSPSA) to choose a neighbor node which fits the reverse spiral path best. Then it starts a query by sending a query packet to the chosen sensor node. A query packet includes the location of the sink node and the query message, and the parameters of the reverse spiral, and the location of the previous hop node. After receiving a query packet, a sensor node will first broadcast this query to all its neighbor nodes to see if any of them has the wanted information. If there is no reply by timeout, the sensor node makes a local copy of the query and forwards the query hop by hop by using RSPSA. Otherwise, if there is some reply before timeout, then the location of the answering node and the answer is forwarded back to the sink node by greedy geographical forwarding. The answer may include the location of the source node and the signature of the interested phenomena. If the sink node needs more data from the source node, then it may establish a data route to the source node directly by greedy geographical forwarding. The query will go on until the reverse spiral meets the dissemination spiral, or the termination condition is satisfied, for example, the maximum hop number or the boundary is reached. Figure 6 shows an illustration of such procedure. In our scheme, the dissemination path always follows the spiral curve, while the query path always follows the reverse spiral curve. Such setting guarantees that the query path and the dissemination path must meet each other at some point quickly. The dissemination path found by ESPSA makes sure that any query path initiated within the sensor network will success.

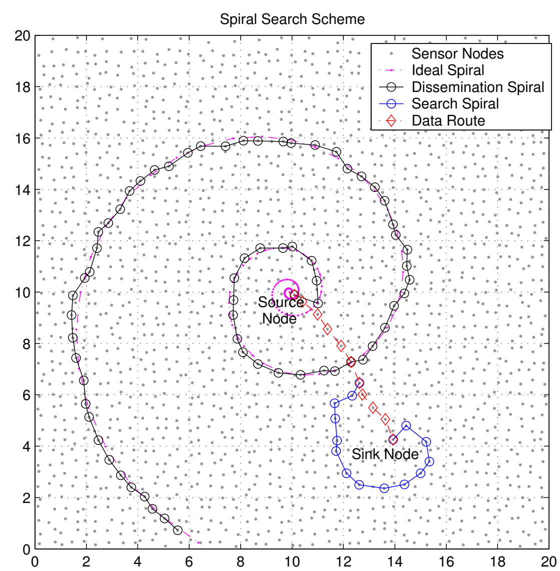

Figure 6: Reverse Spiral Search Procedure

The spiral dissemination achieves the high energy efficiency and low traffic overhead at the price of relatively long delay and blind area inside the spiral path. Therefore, LSDD is not suitable for time-urgent applications. However, LSDD is versatile in that: 1)by adjusting b, the percentage of blind area is under control. Two extreme cases are a straight line ($b \to \infty$) and full flooding ($b \to 0$); 2)by incrementing θ_0 in successive disseminations, a spiral path can sweep the whole coverage over a controllable period and leave no blind area.

4 LSDD CHARACTERIZATION

In this section, we provide a brief numerical analysis on the performance of LSDD. Here we define the dissemination cost of a scheme as the number of involved nodes, and the dissemination ratio as the number of involved nodes to the number of all nodes in a covered area. Given the spiral parameters are fixed, we showed that the dissemination cost of LSDD is in the order of $O(\sqrt{n})$ while that of flooding is in the order of $O(n)$, where n is the number of nodes in the covered area. We also derived the dissemination ratio of LSDD as a function of the spiral parameter b. Given the fact that all nodes get involved in a flooding-based dissemination, this ratio also approximates the ratio of dissemination costs of LSDD and flooding. Finally, we derive the ratio of dissemination radii of LSDD and flooding as a function of b. We found that LSDD and flooding have the same dissemination radius when $b = 0$, and LSDD reaches the longer distance than flooding as b increases given the number of covered nodes is same. Combining the above results, we notice that as b increases, both the dissemination cost and radius increase. In order to achieve an efficient dissemination, we have to make a trade-off between these two index by carefully choosing b. The issue of spiral parameter optimization is out of the scope of this paper, and we leave it for further research. In this paper, we choose $a = 0.1$ and $b = 0.2$ in our simulations, which is concluded empirically from extensive simulations.

There are two important properties owned by the logarithmic spiral. First, the changing rate of the radius is constant.

$$\frac{dr}{d\theta} = \frac{dae^{b\theta}}{d\theta} = abe^{b\theta} = b \cdot ae^{b\theta} = br. \tag{5}$$

Second, the angle between the tangent and radial line at the point (r, θ) is calculated as

$$\psi = \tan^{-1}\left(\frac{r}{dr/d\theta}\right) = \tan^{-1}\frac{1}{b} = \cot^{-1} b. \tag{6}$$

As $b \to 0$, $\psi \to \frac{\pi}{2}$, and the spiral approaches a circle. Note that a constant ψ indicates that if a point is moving along a logarithmic spiral, then its direction will keep unchanged while its radius will keep increasing exponentially. Further, although the spiral itself is a one-dimension curve, the area covered by this curve expands rapidly as the curve lengthens. In addition, by just manipulating the parameters a and b, we are able to adjust the how fast a spiral path covers. For example, when the spiral radius is larger than a threshold, we can set b equal to zero so that the dissemination is confined within a circle.

The arc length (as measured from origin) is given by

$$l = \frac{ae^{b\theta}\sqrt{1+b^2}}{b}. \tag{7}$$

And the curvature is given by

$$\kappa = \frac{e^{-b\theta}}{a\sqrt{1+b^2}}. \tag{8}$$

The coverage area is defined as the area surrounded by sensor nodes on a spiral path, and given by

$$S = \int_{\theta \ominus 2\pi}^{\theta} \int r dr d\theta = \frac{a^2}{4b} e^{2b\theta}(1 - e^{-4b\pi}), \tag{9}$$

where

$$(\alpha \ominus \beta) = \begin{cases} \alpha, \text{ if } \alpha < \beta \\ \alpha - \beta, \text{ if } \alpha \geq \beta \end{cases}$$

If the node density is K nodes per unit area, then the number of sensor nodes in a area covered by a spiral can be estimated as

$$N_c = KS = \frac{Ka^2}{4b} e^{2b\theta}(1 - e^{-4b\pi}). \tag{10}$$

The number of sensor nodes used to construct a spiral can be estimated as

$$N_s = l\sqrt{K} = \frac{\sqrt{K}ae^{b\theta}\sqrt{1+b^2}}{b}. \tag{11}$$

Let DC_f and DC_{sp} denote the dissemination cost of flooding and LSDD, respectively. Suppose that flooding and LSDD cover the same number of nodes, say, n, then it is obviously that DC_f is equal to n, therefore $DC_f \sim O(n)$. Let $N_c = n$, substitute (10) into (11), we have

$$\begin{aligned} N_s &= l\sqrt{\frac{N_c}{S}} \\ &= \sqrt{\frac{1+b^2}{b(1-e^{-4\pi b})}} \cdot \sqrt{n} \\ &= c\sqrt{n}. \end{aligned} \tag{12}$$

When b is fixed, c is a constant, and hence $N_s \sim O(\sqrt{n})$. According to the definitions, $DC_{sp} = N_s$, therefore $DC_{sp} \sim O(\sqrt{n})$.

The dissemination ratio of the LSDD can be estimated by the ratio of N_s to the sume of N_c and N_s. Here the dissemination ratio is defined as the ratio of the number of hops in the spiral path to the number of sensor nodes within the spiral path. We provide another definition of the dissemination ratio in Section V, which expresses the same idea. Let λ denote the estimated dissemination ratio, from equation (10) and (11),

$$\lambda = \frac{N_s}{N_s + N_c} = \frac{4\sqrt{1+b^2}}{4\sqrt{1+b^2} + \sqrt{K}a(1-e^{-4b\pi})e^{b\theta}}. \tag{13}$$

By substituting equation (1), we have

$$\lambda = \frac{\Gamma}{r+\Gamma} = \frac{1}{1+\frac{r}{\Gamma}}, \tag{14}$$

where

$$\Gamma = \frac{4\sqrt{1+b^2}}{\sqrt{K}(1-e^{-4b\pi})}. \tag{15}$$

Note that Γ is proportional to the reciprocal of the square root of K. If K is fixed, then Γ is a function of b. $\Gamma \to \infty$ when either $b \to 0$ or $b \to \infty$. Figure 7 shows that how Γ changes when b increases from 0.02 to 5, where $K = 5$. To get a better view, the logarithmic coordinates is used on the x axis. We can see that the lowest Γ value is located at 0.2 and 0.5.

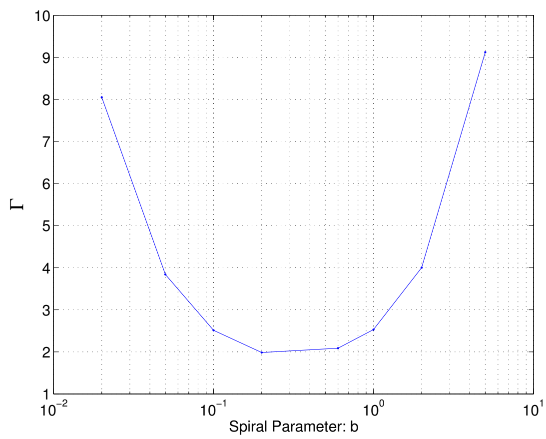

Figure 7: Γ vs. b

In Figure 8, we show that how λ changes as r increases under different b values, where $K = 5$. To get a better view, the logarithmic coordinates is used on both x axis and y axis.

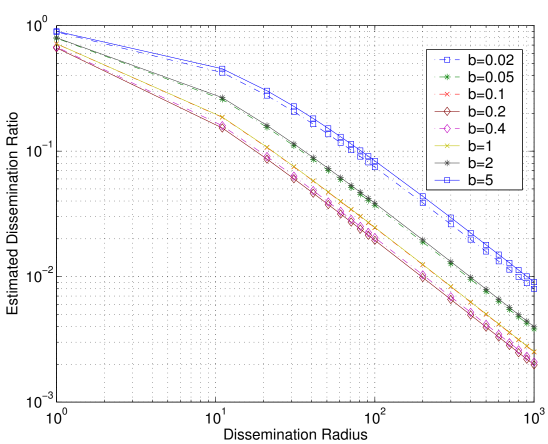

Figure 8: Estimated Dissemination Ratio under different b values

It is clear that the dissemination ratio λ decreases exponentially as the spiral radius increases exponentially. This estimation agrees with our simulation results very well, which will be shown in Section V. For different b values, the ranges of λ are different. Such change pattern complies with the relation of b and Γ as shown in Figure 7. For example, when $b = 0.2$, $\Gamma = 2$ reaches the lowest value, the corresponding curve in Figure 8 is at the bottom of all lines. By carefully choosing the b value, we can control the range of the dissemination ratio for the given dissemination radius. In our simulation, the b value is fixed at 0.2.

Given the number of covered nodes H_m, due to the dense and uniform node distribution, the coverage of flooding can be estimated as

$$S_f = H_m/K. \tag{16}$$

In addition, the maximum radius of the flooding can be approximated by

$$R_f = \sqrt{\frac{S_f}{\pi}} = \sqrt{\frac{H_m}{K\pi}}. \tag{17}$$

Substitute H_m by N_c, the ratio of the flooding dissemination radius to the spiral dissemination radius can be derived as

$$\gamma = \frac{R_f}{R_{sp}} = \frac{\sqrt{\frac{N_c}{K\pi}}}{ae^{b\theta}}$$

.

By substituting equation (10),

$$\gamma = \sqrt{\frac{1-e^{-4\pi b}}{4\pi b}}. \tag{18}$$

According equation (18), $\gamma \to 0$ as $b \to \infty$, and $\gamma \to 1$ as $b \to 0$. Such trend is illustrated in Figure 9. In a dense and uniformly distributed sensor network, the area covered by a flooding dissemination approximates a circle. For the spiral dissemination, the spiral path and coverage is close to a circle when b is close to zero, so the difference between the spiral radius and flooding radius is not much given the same number of covered nodes. However, as b becomes larger, the spiral radius increases at a faster speed so that it will reach much farther than the flooding. Moreover, when the number of covered nodes is larger than some value, the spiral path will definitely cover larger area than flooding.

5 PERFORMANCE EVALUATION

We evaluate the performance of the LSDD on dissemination cost, efficiency, search costs with single sink node and multiple sink nodes, scalability, and fault tolerance by extensive simulations. We compare the performance of LSDD in the above aspects with those of SIDD and GBDD. All simulations are carried out in MATLAB. In SIDD, sensor nodes passively wait for queries. A sink node initiates a query by flooding the query message in the network, and then waits for the interested data until the query times out. In GBDD, both data advertisement and query are routed in a gossiping manner until they meet each other.

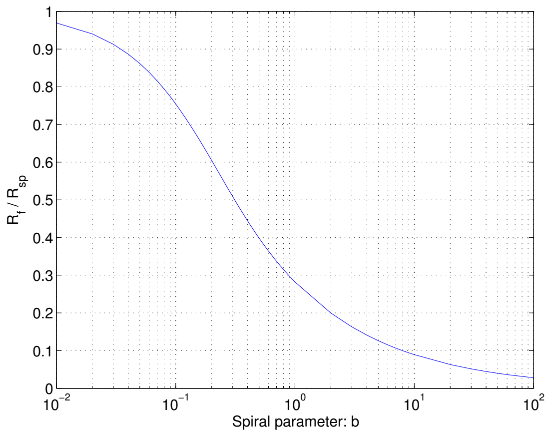

Figure 9: Ratio of the estimated dissemination radius between spiral and flooding

Here gossiping manner means every hop along a path is randomly chosen from the neighborhood list of a node. Directed diffusion (Intanagonwiwat et al. 2000), GEAR (Yu et al. 2002), SPIN-1 and SPIN-2 (Heinzelman et al. 1999) are all based on SIDD, and Gossip routing and Rumor routing Braginsky and Estrin (2002) are based on GBDD. All these schemes are enhanced by different optimization techniques like the query/data aggregation, data-centric routing, and/or geographic routing, to reduce the broadcast range and the number of messages transferred. In our simulations, we focus on the underlying dissemination mechanism and omit the upper layer optimizations, since those techniques can also be applied to LSDD. Geographic routing is not considered in our simulations since it assumes that the source (or sink) nodes know the locations of sink (or source) nodes, while in our scheme such information is unknown. Therefore, it is not fair to compare LSDD with the geographic based routing.

5.1 Simulation Settings

In our simulation, all sensor nodes are deployed in a rectangular area composed by $L \times L$ unit square cells. The side length of each cell is l, where $l = 1$ m in our simulations. The locations of the sensor nodes in each cell follow the uniform distribution. The node density is denoted as K nodes per unit cell. In our simulations, we choose $L = \{10, 20\}$ and $K = \{5, 10\}$. The spiral parameters a and b play important roles in the spiral dissemination as well as reverse spiral query. A large b leads the spiral grow faster and a smaller b makes the spiral cover more nodes in a limited area. In this paper, we assume the parameters are fixed such that $a = 0.1$ and $b = 0.2$ for all cases, and leave the effect of b as future work. Another key parameter, the effective radio range, is denoted as R_t. In our simulations, we assume that all sensor nodes has the same effective radio range as the side length of the unit cell l. A Berkeley mote (Hill et al. 2000) is used as the physical layer model of a sensor node in order to estimate the energy consumption. For details about power of RF transciever, please reference to (Chipcon 2005). We assume that CSMA/CA is the MAC layer protocol and UDP is the transport layer protocol.

With the above settings, it is easy to scale our simulation to practical scenarios by choosing appropriate L, l, K and R_t. For example, when $L = 20$, $R_t = l = 300$ m, $k = 5$, our simulation is equivalent to a sensor network 2000 sensor nodes uniformly deployed within a 6×6 km^2 square area, and each node with an effective radio range equal to 300 m. We only count the messages transmitted for data dissemination and query, and omit the messages sent for other purposes like neighborhood detection or sensed data transmission. All simulation results are averaged over three different network topologies.

To compare the coverage efficiency of different data dissemination schemes, we define the following performance metrics:

1. **Dissemination Radius**: the distance between the source node and the farthest node which receives disseminated data.

2. **Coverage**: the area covered by all the sensor nodes in a dissemination path.

3. **Information density**: the ratio of the number of informed nodes to the coverage of a dissemination path.

5.2 Comparison on energy efficiency

In this set of simulations, we compare the performance metrics of LSDD, SIDD, and GBDD on their energy consumption for message transmission. Since sending and receiving messages usually consumes much more energy than processing in sensor nodes, such limits can be regarded as the energy constraint. The simulation settings are $L = 10$, $K = 5$, $l = R_t = 1$ m. As expected, LSDD achieves the highest energy efficiency: it reaches farthest with given amount of energy, and consumes the least energy for given dissemination distance.

As shown in Figure 10, the maximum dissemination radius of LSDD increases much faster than that of SIDD or GBDD, as energy increases. This figure indicates that the LSDD covers the largest area with the same resources. Given the sensor nodes are uniformly distributed with a fixed density, it is obvious that LSDD covers the most number of sensor nodes. Such features make LSDD a better scheme for data dissemination in the large-scale dense WSNs.

5.3 Comparison on the Search Costs

Now we compare the search cost among LSDD, SIDD, and GBDD. Here we define the search cost as the number of messages sent in both data dissemination and data query processes during the setup of the connection between a source node and a sink node. After the connection is established, the traffic incurred by data transmission and mobility is not counted. The simulation settings are $L = 10$,

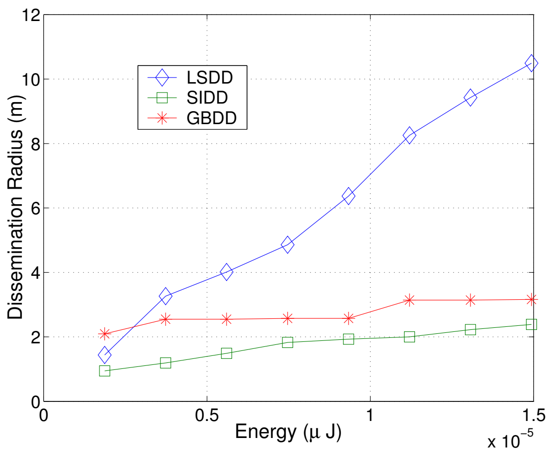

Figure 10: Maximum dissemination radius vs. energy for message transmission

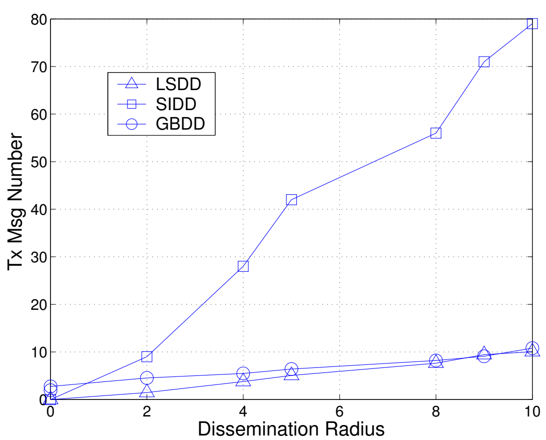

Figure 11: Energy for message transmission vs. dissemination radius

$K = 5$, $l = R_t = 1$ m. The search cost of LSDD is shown to be much lower than that of SIDD or GBDD.

Figure 12 shows the relationship between average search costs and the source-sink distance for all schemes. The energy to send one message or to receive one message is calculated according to the physical layer model. By counting the number of sent and received messages, we can estimate the amount of energy consumed solely for data dissemination process. SIDD and GBDD have similar performance, and GBDD costs a little less at large distance. When the distance is smaller than 4 units, the SIDD and GBDD work better than LSDD. However, as the distance increases, the search cost of SIDD or GBDD grows much faster than LSDD, and is about six times higher than LSDD when the distance is larger than 15.

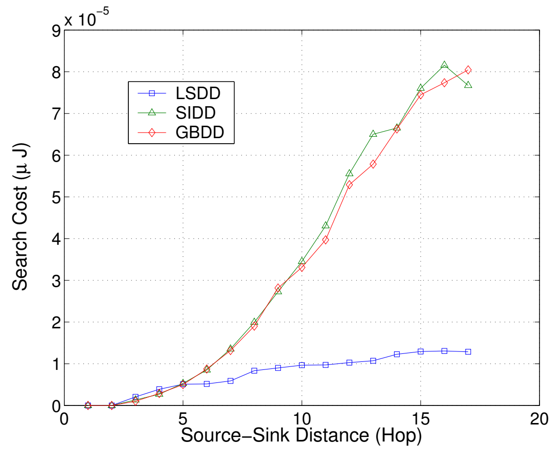

Figure 12: Average search cost comparison

Figure 13 shows the cumulative probability function (CDF) of the search costs of all three schemes at the distance 4, 6 and 9. When the distance is 4, all the CDFs locate in the same range, and the CDF of LSDD is on the right of those of SIDD and GBDD. This shows that the search cost of LSDD is usually higher than that of SIDD or GBDD. As the distance increases, LSDD CDF shifts to the left side of SIDD and GBDD, while SIDD and GBDD still lap over each other. We can also see that the SIDD CDF concentrates in a more narrow range than the GBDD CDF because of the randomness of GBDD. All three CDFs look similar to the normal distribution.

5.4 Comparison on Scalability

In this set of simulations, we test the scalability of LSDD, SIDD, and GBDD by applying them to networks with different sizes and node densities. We keep the unit cell size and the effective radio range as constant like the previous simulations, that is, $l = R_t = 1$ m. The range of L is $\{10, 20\}$, and the range of node density K is $\{5, 10\}$. There are totally 4 settings:

1. Area: 10×10, node density: 5
2. Area: 20×20, node density: 10

3. Area: 10×10, node density: 5

4. Area: 20×20, node density: 10

The results are shown in Figure 14. It is clear that

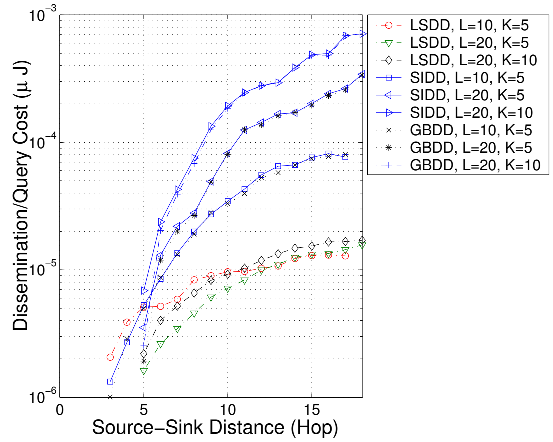

Figure 14: Scalability comparison for LSDD, SIDD, and GBDD

the search cost of LSDD increases linearly as the distance increases, and the differences between the search costs of LSDD in all 4 settings are also small which indicates the good stableness and scalability of the search algorithm. On the contrary, the search costs of SIDD and GBDD increase exponentially and the differences between search costs in 4 settings are much higher than that of LSDD. These figures indicate that LSDD can be applied to a large scale sensor network without incurring heavy traffic overhead for data dissemination and query. The superior scalability of LSDD over SIDD and GBDD is confirmed.

Figure 15 compares the total search costs of LSDD, SIDD, and GBDD with multiple sink nodes under different network settings. It is clear that as the number of sensor nodes increases (by increasing either the network size or the node density), LSDD becomes more efficient than SIDD and GBDD. SIDD and GBDD have the similar performance with multiple sink nodes. It shows that LSDD is more suitable for large scale sensor networks.

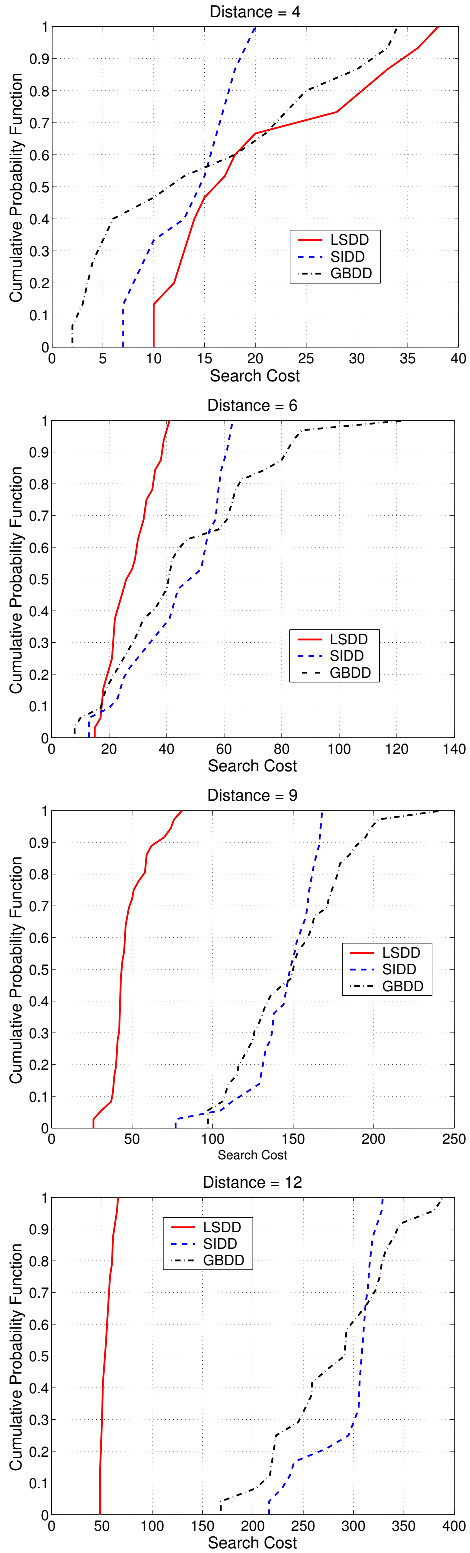

Figure 13: Comparison of the search cost CDFs given Distance

5.5 Fault tolerance

In a sensor network, each individual node is fragile to failure due to limited energy and other accidents. Therefore, the topology of a sensor network may vary when some nodes fails. In the simulation, we examine the performance of LSDD in such faulty network. We set different failure rates, and test if the advertisement packet can be disseminated over the area by LSDD. The failed nodes are uniformly chosen according to the failure rate, and the results are averaged over a large number of simulations. As shown in Figure 16, the successful rate decreases as the failure rate increases, and the successful rate under high node density is higher than that under low node density.

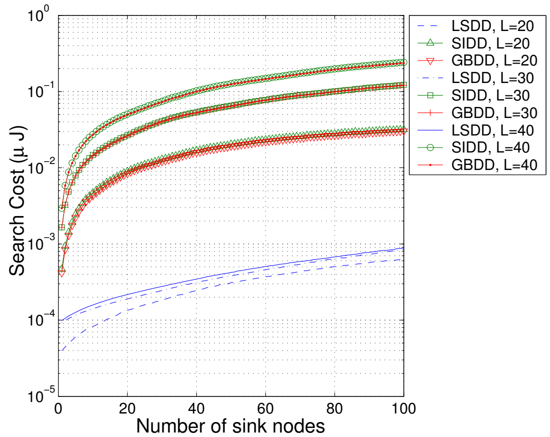

Figure 15: Search cost comparison with multiple sink nodes

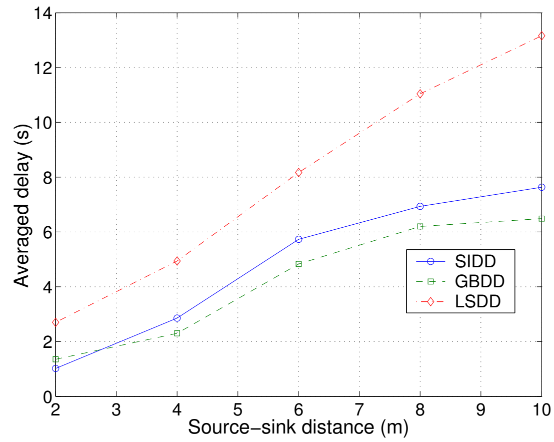

Figure 17: Comparison on delay performance

We can see that when the failure rate is lower than 0.1, the successful rate under all cases is higher than 85 percents. It indicates that LSDD is very robust to the node failures. As the network gets larger and denser, the negative effect of failure nodes gets smaller under LSDD.

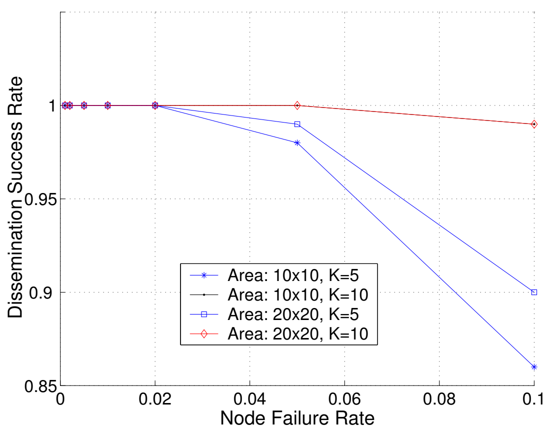

Figure 16: Fault tolerance test

5.6 Forwarding delay

In LSDD, messages are propagated along a spiral curve instead of spreading over the plane as in SIDD, which leads to a longer delay than flooding-based schemes. Such loss on delay performance can be viewed as a price for the gain on energy efficiency and robustness. To estimate the impact of this trades-off, we measured the dissemination delay of between one pair of source and sink nodes with different distances. The simulation settings are $L = 20$, $K = 5$, $l = R_t = 1$ m. The results shown in Figure 17 are averaged over 100 simulations.

As expected, LSDD spends the longest time, and GBDD takes the shortest time in most distances. The difference between LSDD and SIDD increases slowly as the distance increases. As mentioned before, LSDD is not designed for time-urgent applications which regard the short delay as the fist privilege. Instead, LSDD is oriented with energy efficiency and scalability. Besides, when source nodes and sink nodes search each other by LSDD, their paths usually meet quickly after the sum of their spiral radii become larger than the distance between these two node. Thus, the traffic overhead is moderate.

6 CONCLUSION

In this paper, we proposed a novel data dissemination scheme: logarithmic spiral data dissemination (LSDD). LSDD imitates a natural evolution of the spiral to facilitate the data dissemination and data query in sensor networks. LSDD improves the performance-resource ratio in stable structure, high energy efficiency, low traffic overhead, and flexible scalability. LSDD can also fit into large-scale networks more efficiently than other schemes. In addition, according to our simulations, LSDD is quite robust to the heavy query load caused by multiple sink nodes, as well as the faulty network caused by unreliable nodes. As is shown in this paper, LSDD is a strong candidate for the data dissemination in a sensor network.

REFERENCES

[1] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. E. Culler, and K. S. J. Pister (2000) "System Architecture Directions for Networked Sensors," in *Architectural Support for Programming Languages and Operating Systems*, pp. 93–104.

[2] J. M. Kahn, R. H. Katz, and K. S. J. Pister (1999) "Next Century Challenges: Mobile Networking for "Smart Dust"," in *Proc. of MobiCom 1999*, pp. 271–278.

[3] D. Braginsky and D. Estrin (2002) "Rumor routing algorthim for sensor networks," in *Proc. of WSNA 2002*, Atlanta, GA, Sept. pp. 22–31.

[4] W. Heinzelman, J. Kulik, and H. Balakrishnan (1999) "Adaptive protocols for information dissemination in wireless sensor networks," in *Proc. of MobiCom 1999*, pp. 174–185.

[5] C. Intanagonwiwat, R. Govindan, and D. Estrin (2000) "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in *Proc. of MobiCom 2000*, pp. 56–67.

[6] S. Tilak, N. Abu-Ghazaleh, and W. Heinzelman (2002) "A Taxonomy of Wireless Microsensor Network Models," *ACM Mobile Computing and Communications Review*, 2002.

[7] Z. J. Haas, J. Y. Halpern, and L. Li (2002) "Gossip-based Ad Hoc Routing," in *Proc. of INFOCOM 2002*, Ithaca, NY, June, pp. 1707–1716.

[8] C. Intanagonwiwat (2002) "Impact of network density on data aggregation in wireless sensor networks," in *Proc. of ICDCS-22*, Vienna, Austria, July.

[9] D. Li, X. Jia, and H. Liu (2004) "Energy efficient broadcast routing in static adhoc wireless networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 2, pp. 144–151, April-June.

[10] Y. Yu, R. Govindan, and D. Estrin (2001) "Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks," UCLA Computer Science Department, Tech. Rep. UCLA/CSD-TR-01-0023, May.

[11] D. Niculescu and B. .Nath (2003)"Trajectory based forwarding and its applications," in *Proc. of the MOBICOM 2003*, San Diego, CA, Sept., pp. 260–272.

[12] R. C. Archibald (1918) "The Logarithmic Spiral," *Amer. Math. Monthly*, no. 25, pp. 189–193.

[13] W. L. Winston (1993) *Operation Research: Applications and Algorithms*, 3rd ed. Duxbury Press.

[14] Chipcon Inc. (2005) "CC1000 Datasheet v2.3".