Analysis of Gradient-Based Routing Protocols in Sensor Networks*

Jabed Faruque, Konstantinos Psounis, and Ahmed Helmy

Department of Electrical Engineering,
University of Southern California, Los Angeles, CA 90089
`{faruque, kpsounis, helmy}@usc.edu`

Abstract. Every physical event results in a natural information gradient in the proximity of the phenomenon. Moreover, many physical phenomena follow the diffusion laws. This natural information gradient can be used to design efficient information-driven routing protocols for sensor networks. Information-driven routing protocols based on the natural information gradient, may be categorized into two major approaches: (*i*) the single-path approach and (*ii*) the multiple-path approach. In this paper, using a regular grid topology, we develop analytical models for the query success rate and the overhead of both approaches for ideal and lossy wireless link conditions. We validate our analytical models using simulations. Also, both the analytical and the simulation models are used to characterize each approach in terms of overhead, query success rate and increase in path length.

1 Introduction

Sensor networks are envisioned to be widely used for habitat and environmental monitoring where the attached tiny sensors sample various physical phenomena. More specifically, advances in the MEMS technology make it possible to develop sensors to detect and/or measure a large variety of physical phenomena like temperature, light, sound, radiation, humidity, chemical contamination, nitrate level in the water, etc. Every physical event leaves some fingerprints in the environment in terms of the event's effect; e.g., fire increases the temperature, chemical spilling increases the contamination, nuclear leakage increases the radiation so on. Moreover, most of the physical phenomena follow the diffusion law[13][14] with distance, i.e., $f(d) \propto \frac{1}{d^\alpha}$, where d is the distance from the point having the maximum effect of the event, $f(d)$ is the magnitude of the event's effect, and α is the exponent of the diffusion function that depends on the type of effect and the medium; e.g., for light $\alpha = 2$, and for heat $\alpha = 1$. Hence, routing protocols used in sensor networks for habitat and environmental monitoring applications can exploit this natural information gradient to efficiently forward the queries toward the source of the event. Throughout this document the term "*source*" is used to refer to the source of the event; e.g., the contaminant, the epicenter of an earthquake, etc.

---

* This research has been partially supported by NSF award 0435505.

V. Prasanna et al. (Eds.): DCOSS 2005, LNCS 3560, pp. 258–275, 2005.
© Springer-Verlag Berlin Heidelberg 2005

Traditional data-centric routing protocols for sensor networks are based mostly on flooding [1] or random-walks [3], [4]. However, these approaches do not utilize the domain-specific knowledge, i.e., the information gradient about the monitored phenomenon. Here, we keep our focus on the routing protocols that exploit the information gradient to route a query efficiently in sensor networks from the sink to the source.

In real life, sensors are not always perfect and are subject to malfunction due to obstacles or failures. Also, the characteristics of the sensor nodes, e.g., limited battery life, the energy expensive wireless communication, and the unstructured nature of the sensor network, make the data-centric routing protocols based on the information gradient a challenging problem. Several routing protocols have been proposed to exploit the information gradients. These query routing protocols use greedy forwarding and can be broadly classified in two categories: (1) *Single-path approach*[5][6][8], where the query reaches the source from the sink through a single path, and (2) *Multiple-path approach*[7], where the query uses multiple paths to reach the source.

In this paper, we do not aim to design new routing protocols per se. Rather, *the objective of the research is the evaluation and the analysis of the general approaches to route a query using the natural information gradient in the sensor networks.* In particular, we use probabilistic modeling methods to derive analytical expressions for the energy overhead and the query success rate of each approach. Also, we compare the performance of the query routing approaches using carefully selected performance metrics. Our analysis is validated through extensive simulations. For the analysis and the simulations, we only consider sensor networks with static nodes, which is usually the case for environmental monitoring, and we assume that the queries are triggered from a sink to identify the origin (i.e., source) of the event, after the event's occurrence. To keep the analysis simple, we ignore potential packet collisions, which can be (and usually is) effectively reduced by inserting a random delay time before forwarding the query packet. However, wireless link loss is considered in both the analysis and the simulations. The main contributions and findings of this paper include:

- Analytical models for the query success rate and the overhead of the single-path and the multiple-path approaches to route a query using the information gradients in sensor networks. Validation of the analytical models using extensive simulations.
- Performance analysis of both routing approaches using the analytical models and simulations in ideal and lossy link conditions.
- In the ideal wireless link case, it is found that the multiple-path approaches are more energy efficient than the single-path approaches when the source is relatively close to the sink. Also, the multiple-path approaches yield shorter paths than the single-path approaches. Further, as the number of malfunctioning nodes in the network increases, the query success rate of the multiple-path approaches degrades a lot slower than that of the single-path approaches.
- In the lossy wireless link case, the query success rate of the single-path approaches drops drastically while the multiple-path approaches are quite resilient.

2 Related Work

Several approaches have been proposed for routing in sensor networks. The major difference between the information-gradient based approach[5][6] and the flooding[1][2] and the random-walks[3][4][9] based approaches is that the former uses the sensors measurements about the event's effect for routing decisions. In [9], Servetto and Barrenechea have shown that multiple random-walks improve load balancing and minimize latency with increased communication cost. Also, they analyzed the random-walk approach in regular/irregular and static/dynamic grid topology, but they did not consider the existence of information gradients. We now briefly summarize previously proposed gradient based routing protocols.

Chu, Hausseker and Zhao propose CADR (Constrained Anisotropic Diffusion Routing) mechanism[5], especially designed for localization and target tracking. CADR uses a proactive sensor selection strategy for correlated information based on a criterion that combines information gain and communication cost. CADR is a single path greedy algorithm that routes a query to its optimal destination using the local gradients to maximize the information gain through the sensor network.

Later work by Liu, Zhao and Petrovic [6] proposed the min-hop routing algorithm to overcome the limitation of CADR to handle local maxima and minima. The algorithm uses a multiple step look-ahead approach with single path query forwarding. Here, the initial network discovery phase determines the minimum look-ahead horizon (in hops) so that the path planning phase can avoid network irregularities. The algorithm improves the success rate of routing message with additional search cost. Also, the increase in the neighborhood size causes more communications between the cluster leaders and their neighbors.

Some recent studies on information-driven routing protocols also use the single-path approach [8]. It is important to note that all the above information-driven protocols based on the single-path approach, use a proactive phase to prepare the gradient information repository. In our study we analyze the performance of the query routing mechanisms without considering the cost of the proactive phase.

In [7], a multiple path exploration mechanism is proposed to discover a route or an event. It is a reactive and distributed routing mechanism to effectively exploit the natural information gradient repository. The protocol controls the instantiation of multiple paths using a probabilistic function based on the simulated annealing concept. It uses flooding to forward the query when no information gradient is available. The efficiency of the protocol depends on properly selecting the parameters of the probabilistic function.

According to [7], the diffusion information in the environment consists of a flat (i.e. zero) and a gradient information region. However, the single-path protocols are unable to forward a query in the flat information region. In this work, we only consider the performance of various routing approaches in the gradient information region, while malfunctioning sensors are uniformly distributed.

In this work, our interest is primarily focused on the systematically analyzing the performance e.g., the query success rate and the overhead, of the single-path and the multiple-path approaches to design data centric routing protocols in the presence of a natural information gradient. In the analysis, we use a probabilistic framework to develop simple analytical models for the success rate and the overhead for both approaches in ideal and lossy wireless link conditions. We also simulate these protocols with more realistic scenarios.

3 Query Routing Approaches

According to the discussion of Section 2, two major approaches, (*i*) the single-path approach and (*ii*) the multiple-path approach, are used for query routing protocols to exploit the information gradients. To properly describe these routing approaches, we need to define two terms: (1) *Active node*, a node which is currently holding the query, and (2) *Candidate node*, a node which has never received the query. Now, a brief description of both routing approaches is given below:

Single-Path Approach: The query follows a single path to reach the source from the sink. At each step of the query forwarding, the active node uses a look-ahead parameter r, $r \geq 1$, to collect information from all candidate nodes within r-hops. For $r > 1$, all nodes within $r - 1$ hops need to transmit the request of the active node to gather information about the event. Note that for $r = 0$, the single-path approach becomes a random-walk and is unable to utilize the gradient information repository.

Single-path approach based protocols can be designed in several ways using different selection policies for the next active node. In our study, we consider the following two policies:

a) *Basic single-path approach*: In this policy, the protocol always selects the node with the maximum information among all candidate nodes within r-hops of the active node, when the node's information is higher than that of the active node. This selection policy is sensitive to local maxima and arbitrarily high readings of the malfunctioning nodes that cause these local maxima. The resilience of the protocols based on this approach can be improved by using filters to avoid such arbitrarily high readings.
b) *Improved single-path approach*: In this policy, the active node forwards the query to a node having the maximum information among all candidate nodes within r-hops of the active node. So, the information content of all candidate nodes can be less than that of the active node. Here, the query forwarding ends either at the source node or at an active node having no candidate nodes within r-hops.

Multiple-path Approach: This approach forwards the query through multiple paths towards the source without any look-ahead phase. These paths may not be disjoint paths. Usually the active nodes forward the query greedily when information level improves. In the presence of malfunctioning nodes having wrong information, the protocols based on this approach can use probabilistic forwarding. For example, the protocol proposed in [7] uses a diffusion function for probabilistic forwarding. It creates some extra paths but the protocol can adaptively change the forwarding probability to control the instantiation of these extra

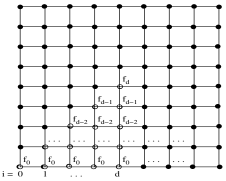
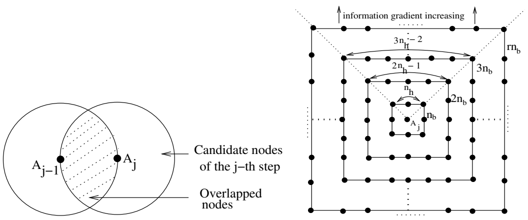


Fig. 1. A regular grid topology. Here, f_j indicates the magnitude of information and $f_0 < f_1 < \cdots < f_d$. The triangular pattern represented by white dots is present *eight times* in the grid. Information magnitude near the source is f_d and gradually reduces towards the edge



(a) A_{j-1} and A_j are the active nodes of steps $(j-1)$ and j.

(b) Active node A_j with r-hop neighbors.

Fig. 2. Single path approach with look-ahead r

paths. To capture this in the analysis, the forwarding probability is considered different at each step of the query forwarding.

All query routing protocols considered in this paper use unique query IDs to suppress duplicates and to avoid loops.

4 Analytical Model

In this section, we derive models to describe the characteristics of the approaches used to design information-driven routing protocols for sensor networks.

4.1 Assumptions and Metrics

Let a sensor network consist of N nodes and the nodes be deployed as a regular grid as shown in Fig.1. Assume that only one event occurs and the effect of the event follows the diffusion law as previously described. Assume also that the information gradient is available in the whole network, i.e., there is no flat information region. Further, consider that the malfunctioning nodes have arbitrary information and that these nodes are uniformly distributed in the network which may cause failure during the route discovery. Let p_f be the probability that a node is malfunctioning. The stored information in the malfunctioning node can be arbitrarily high or low and this is equally likely. Finally, assume each node is able to communicate via broadcast with its eight neighbors on the grid.

Suppose that the querier, i.e. the sink, is located d hops away from the source node. The query is forwarded step by step, where the term "*step*" is defined as follows:

1) Single-path approach: The active node collects information from all candidate nodes within r-hops. Then it forwards the query to the next active node which is r-hops away.

2) Multiple-path approach: The active nodes forward the query either greedily or probabilistically via broadcasts. Then the query reaches the candidate nodes which are 1-hop away.

Due to greedy forwarding based on the information gradient, after each step of the query forwarding, the query reaches one step closer to the source with some probability.

Here, we are interested to develop analytical models for the following two metrics:

1) *Query Success rate*, i.e. success probability, is the probability that the query initiated from the sink reaches the source.
2) *Overhead* in terms of energy dissipation, calculated as the number of transmissions required to forward the query to the source and to get the reply back from the source using the reverse path.

4.2 Single-Path Approach

Let n_b be the number of nodes that are one hop away from the active node. Overlap of the sensor nodes radio coverage causes some nodes to receive the same query multiple times. The query ID is used to suppress duplicate queries. If we consider the radio coverage of a node to be circular and the radius to be the same for all nodes, then using simple geometry, it can be shown that the overlap is one-third. For all except the first step of the query forwarding, let $n_c = \frac{2}{3}n_b$ denote the number of candidate nodes within one-hop of the active node. Now, it is easy to show that the total number of neighbors and candidate nodes within r-hops of the active node equal $n_B = \frac{r(r+1)}{2}$ and $n_C = \frac{2}{3}n_B$ respectively. However, for the first step of the query forwarding, $n_C = n_B$.

Within one-hop of the active node, let n_h and n_l be the number of candidate nodes having high and low information respectively according to the diffusion pattern of the event's effect in the grid, where $n_c = n_l + n_h$. Thus, except for the first step of the query forwarding, it can be shown from Fig.2 that the total number of high information candidate nodes within r-hops of the active node equals $n_H = \frac{r(r+1)}{2}n_h - \frac{r(r-1)}{2}$. Finally, $n_L = n_C - n_H$ denotes the number of low information candidate nodes within r-hops, since $n_C = n_L + n_H$.

Basic Single-path Approach: At each step of the query forwarding, the protocol selects the node that has the highest information among all nodes (including the active node) within r-hops and forwards the query to that node. When all candidate nodes function properly, the query forwarding ends at the source. However, the query forwarding also halts at a local maxima. A malfunctioning node contains arbitrarily high information with probability $\frac{p_f}{2}$. Let $f(i)$ be the diffusion function value when the source is i-hops away and R_{max} be the maximum reading of the sensor attached to the source node. The arbitrarily high reading of a malfunctioning node can be between 0 and R_{max}. But at the j-th step, the query forwarding halts for any candidate node having value between $f\left(\left(\lceil \frac{d}{r} \rceil - j\right) r\right)$ and R_{max}. Therefore, the probability of success equals

$$P_{single_B} = \prod_{j=1}^{\left\lceil \frac{d}{r} \right\rceil} \left[1 - \frac{p_f}{2} \cdot \frac{R_{max} - f\left(\left(\left\lceil \frac{d}{r} \right\rceil - j\right) r\right)}{R_{max}} \right]^{n_C} . \qquad (1)$$

Here, the product term is the probability that at each step of the query forwarding, there is no malfunctioning candidate node having arbitrarily high information when the source is i-hops away and a total of $\left\lceil \frac{d}{r} \right\rceil$ such steps are required to reach the source. In the expression, we assume that after each forward of the query, the next active node is r-hops away. This may not be possible if all high information candidate nodes that are r-hops away are malfunctioning. However, a protocol can still select a node that is r-hops away as the next active node and notify the maximum information collected from all of the nodes within r-hops.

At each step of the query forwarding, to gather information for the active node, each node within $(r-1)$-hops from the active node sends the request to its n_b neighbors and finally all nodes within r-hops send the reply to the active node through their one-hop away neighbors. Thus, the number of transmissions required to forward the query to the next active node is $\left(1 + n_b \sum_{i=1}^{r-1} i\right) + n_b \sum_{i=0}^{r-1} (r-i) + r$, that equals $1 + r^2 n_b + r$. Now, the total number of transmissions require to forward the query from the sink to the source and get the reply equals

$$T_{single_B} = \left\lceil \frac{d}{r} \right\rceil (1 + r^2 n_b + r) + d. \qquad (2)$$

Here, $1 + r^2 n_b + r$ is the required number of transmissions at each step of the query forwarding and total $\left\lceil \frac{d}{r} \right\rceil$ such steps are required. Also, it requires d transmissions to reply to the sink through the reverse path. If we consider that nodes in the overlapping regions respond only one time, then the total number of transmissions equals

$$T_{single_{B_O}} = (1 + r^2 n_b + r) + \left(\left\lceil \frac{d}{r} \right\rceil - 1\right) \left[1 + \frac{2r^2 n_b}{3} + r\right] + d, \qquad (3)$$

since except for the first step of the query forward, each remaining step requires $1 + \frac{2r^2 n_b}{3} + r$ transmissions for non-overlapping nodes.

Improved Single-Path Approach: In this policy, at each step of the query forwarding, the protocol selects the node with the maximum information among all candidate nodes within r-hops of the active node and forwards the query to that node. For protocols based on this approach, the query success rate and the overhead depend on the length of the path followed by the protocol. Though the sink node is d-hops away from the source, in the presence of arbitrary information in the malfunctioning nodes, the query may follow a path other than the shortest path. Let l_j denote the length of the path after the j-th forward of the query. If all sensor nodes in the network were perfect, the query should follow the shortest path and $l_j - l_{j-1} = r$. However, due to malfunctioning nodes, some low information candidate nodes may contain arbitrarily high information with probability $\frac{p_f}{2}$. The probability of selecting such a node as the next active node is $\left(\frac{p_f}{2}\right) \frac{n_L}{n_C}$ and the path length difference per step equals $l_j - l_{j-1} = r + L_{err}$, where L_{err} is the average path length increase per step. For simplicity of the analysis, if we consider that each step of the query forwarding is independent, then the length of the path after the j-th step can be expressed as

$$l_j = \begin{cases} \frac{\frac{p_f}{2}.n_L}{n_C}\left(l_{j-1} + r + L_{err}\right) + \left(1 - \frac{\frac{p_f}{2}.n_L}{n_C}\right)\left(l_{j-1} + r\right), & \text{for } j > 1 \\ r, & \text{for } j = 1. \end{cases}$$

Thus, l_d denotes the length of the path followed by the protocol while the actual distance of the source is d.

Here, the query forwarding halts either at the source node or at an active node with no candidate nodes within r-hops. In the gradient information repository, there are always some candidate nodes as the query forwarding proceeds from low to high information nodes. However, due to malfunction, with probability $\frac{p_f}{2}$, high information candidate node(s) may contain arbitrarily low information which may be lower than that of some low information candidate nodes. When all high information candidate nodes are malfunctioning and containing arbitrarily low information, the query forwarding proceeds through a low information candidate node. Such low information candidate nodes are unable to find any candidate node as its all neighbors may already have received the query and the query fails to reach the source. Therefore, the probability of query success equals

$$P_{single_I} = \left[1 - \left(\frac{p_f}{2}\right)^{n_H}\right]^{\left\lceil \frac{l_d}{r} \right\rceil}. \tag{4}$$

Here, $1 - (\frac{p_f}{2})^{n_H}$ is the probability that not all high information candidate nodes are malfunctioning at each step and a total of $\left\lceil \frac{l_d}{r} \right\rceil$ such steps are required.

Similar to the Equation(2), here the total number of transmissions required to forward the query from the sink to the source for path length l_d and get the reply equals

$$T_{single_I} = \left\lceil \frac{l_d}{r} \right\rceil \left(1 + r^2 n_b + r\right) + l_d. \tag{5}$$

Here, l_d transmissions are required to reply to the sink using the reverse path.

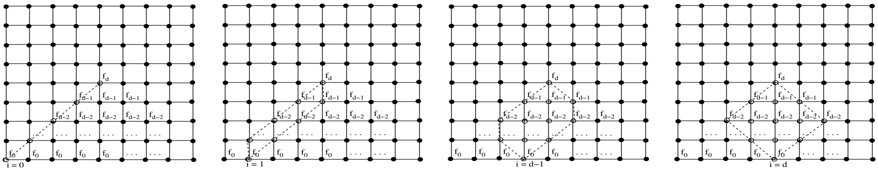
Further, considering that nodes in the overlapping regions respond only one time, the total number of transmissions equals

$$T_{single_{I_O}} = \left(1 + r^2 n_b + r\right) + \left(\left\lceil \frac{l_d}{r} \right\rceil - 1\right)\left[1 + \frac{2r^2 n_b}{3} + r\right] + l_d, \tag{6}$$

since except for the first step of the query forward, each remaining step requires $1 + \frac{2r^2 n_b}{3} + r$ transmissions for the non-overlapping nodes.

4.3 Multiple-path Approach

In this routing approach, except for the first step of the query forwarding, multiple active nodes may forward the query to the candidate nodes without any look-ahead phase. The active nodes with lower information forward the query probabilistically. Let p_j denote the probability of forwarding the query probabilistically at the j-th step of the query forwarding. So, at the j-th step of



(a) Sink at $i = 0$. (b) Sink at $i = 1$. (c) Sink at $i = d - 1$. (d) Sink at $i = d$.

Fig. 3. Query forwarding pattern using the multiple path approach. Depending on the position of the sink patterns are different. Here, the white dots indicate the participating nodes to forward the query towards the source and d is the distance between the source and the sink

the query forwarding, a high information candidate node fails to forward the query with probability $q_j = \left(\frac{p_f}{2}\right)(1 - p_j)$, where $\left(\frac{p_f}{2}\right)$ is the probability that the high information candidate node is malfunctioning and containing low information. For simplicity of the analysis, we assume that the query forwarding steps are independent. This simplified model still captures the characteristics of the multiple-path approach, while the analysis is kept tractable.

Let $P_{multiple}$ and $T_{multiple}$ denote the query success rate and the overhead of the multiple-path approach. Consider that i denotes the position of the querier, i.e. the sink, in the last row of the grid as shown in Fig.1 and Fig.3. The query forwarding patterns, i.e. the number of participating nodes, are different for different values of i as shown in Fig.3. Also, it is easy to show that according to the diffusion pattern in the grid, the average number of low information candidate nodes is four at each step of the query forwarding. These nodes forward the query probabilistically.

According to the query forwarding patterns as shown in Fig.3, for an even value of i, $0 \le i \le d$, the query success probability equals

$$P_{multiple_e}(i) = (1 - q_1) \cdot \left[\prod_{m=\frac{i}{2}}^{d-\frac{i}{2}} \left(1 - q_{m+1}^{i+1}(1 - p_{m+1})^4\right)\right] \cdot \prod_{m=1}^{\frac{i}{2}-1} \left(1 - q_{m+1}^{2m+1}(1 - p_{m+1})^4\right)\left(1 - q_{d-m+1}^{2m+1}(1 - p_{d-m+1})^4\right),$$

and for an odd value of i, $0 \le i \le d$, it equals $P_{multiple_o}(i)$, where the only difference with the above expression is the limits of the products (i.e., $\lceil \frac{i}{2} \rceil \le m \le d - \lceil \frac{i}{2} \rceil$ and $1 \le m \le \lfloor \frac{i}{2} \rfloor$ in the first and the second products respectively). Here, the terms having the form q_y^x, $1 \le y \le d$, in the above equation expresses the probability of not forwarding the query at the y-th step by all high information candidate nodes, x, as they are malfunctioning and containing less information. Also, the surrounding four low information candidate nodes fail to forward the query with probability $(1 - p_y)^4$. Thus, at the y-th step, $1 - q_y^x(1 - p_y)^4$ is the probability of forwarding the query towards the source. Detailed derivation of these equations and all the equations of the remaining document are presented in [15].

Similarly, to compute the energy dissipation, we also consider the different forwarding patterns of the query as shown in Fig.3. The total number of transmissions required to forward the query to the source for an even value of i, $0 \leq i \leq d$, equals

$$T_{multiple_e}(i) = \left[\frac{i^2}{2} - 1 + (i+1)(d-i+1)\right] - \left[\sum_{m=1}^{\frac{i}{2}} (2m-1)\, q_m \right.$$
$$\left. + \sum_{m=1}^{\frac{i}{2}-1} (2m+1) q_{d-m+1} + (i+1) \sum_{m=1}^{d-i+1} q_{\frac{i}{2}+m} \right] + 4 \sum_{m=1}^{d} p_{m+1}, \qquad (7)$$

and for an odd value of i, $0 \leq i \leq d$, it equals $T_{multiple_o}(i)$, where the differences with Equation(7) are the first term ($\left[\frac{(i+1)^2}{2} - 1 + (i+1)(d-i)\right]$) and the limits of the second term (i.e., $1 \leq m \leq \frac{i+1}{2}$, $1 \leq m \leq \frac{i+1}{2} - 1$ and $1 \leq m \leq d - i$ respectively). Here, the first term of the above equation computes the number of transmissions due to high information candidate nodes, if all such nodes are working properly. However, some of these nodes are malfunctioning and unable to forward the query with probability q_y, $1 \leq y \leq d$, at the y-th step of the query forwarding. This reduces the overhead. The second term of the above equation computes this reduction. Finally, the third term computes the overhead due to the probabilistic forwarding of the four low information candidate nodes.

Now, if each value of i, $0 \leq i \leq d$, is equally likely, then the average probability of success equals

$$P_{multiple} = \frac{1}{d+1}\left[\sum_{k=0}^{\lfloor \frac{d}{2} \rfloor} P_{multiple_e}(2k) + \sum_{k=1}^{\lceil \frac{d}{2} \rceil} P_{multiple_o}(2k-1)\right]. \qquad (8)$$

Similarly, the average number of transmissions required to forward the query from the sink to the source and get the reply using the reverse path equals

$$T_{multiple} = \frac{1}{d+1}\left[\sum_{k=0}^{\lfloor \frac{d}{2} \rfloor} T_{multiple_e}(2k) + \sum_{k=1}^{\lceil \frac{d}{2} \rceil} T_{multiple_o}(2k-1)\right] + d. \qquad (9)$$

4.4 Wireless Link Loss

Recent experimental studies on wireless sensor networks [11][12] have shown that in practice, the wireless links of the sensor networks can be extremely unreliable and deviate from the idealized perfect-reception-range models at a large extent. Due to the lossy links, the transmissions of a node may not reach to some of its neighbor nodes. This affects the performance of the routing protocols. In this section, we consider the wireless link loss in the derivation of analytical models of the query success rate and the overhead for both routing approaches. Since the *basic single-path* approach shows poor performance even for the ideal wireless link condition (see Fig.4), only the *improved single-path* and the *multiple-path* approaches are considered in this section.

Let, p_c be the probability of a link loss, and assume that the lossy links are uniformly distributed in the network. Assume also that no automatic repeat request (ARQ) is used to broadcast or to forward the query towards the source, which is usually the case in sensor networks for energy conservation. However, notice that the ARQ mechanism is used to send the reply (if the query is successful) to the sink using the reverse path.

Improved Single-Path Approach: Due to the lossy links, at each step of the query forwarding, the broadcast of the active node may not reach to all candidate nodes within r-hops. Thus, at each step of the query forwarding, $n_H(1-p_c)$ high information candidate nodes receive the broadcast. Similarly, the active node receives responses from $n_H(1-p_c)^2$ high information candidate nodes. Also, the probability to forward the query to the next active node, which is r-hops away, is $(1-p_c)^r$. Thus, for the improved-single path approach, the probability of success equals

$$P_{single_{Ic}} = \left[\left(1 - \left(\frac{p_f}{2}\right)^{n_H(1-p_c)^2}\right)(1-p_c)^r\right]^{\left\lceil \frac{l_d}{r} \right\rceil}. \tag{10}$$

Here, $\left(1 - \left(\frac{p_f}{2}\right)^{n_H(1-p_c)^2}\right)(1-p_c)^r$ is the probability of success at each step of the query forwarding and a total of $\left\lceil \frac{l_d}{r} \right\rceil$ such steps are required.

In this routing approach, the first step of the query forwarding requires $1 + \frac{r(r-1)}{2} n_b(1-p_c) + \frac{r(r+1)}{2} n_b(1-p_c) + r(1-p_c)$ transmissions, which equals $1 + r(1-p_c)(rn_b+1)$. With probability p_c, the nodes of the overlapped region can be candidate nodes of the current active node as they failed to receive the broadcast of the previous active node due to the lossy links. Further, consider that the overlapped region nodes respond only one time. So, each remaining step of the query forwarding requires $1 + \frac{1}{3}\left[\frac{r(r-1)}{2} n_b(1-p_c) + \frac{r(r+1)}{2} n_b(1-p_c)\right] p_c + \frac{2}{3}\left[\frac{r(r-1)}{2} n_b(1-p_c) + \frac{r(r+1)}{2} n_b(1-p_c)\right] + \frac{r}{1-p_c}$ transmissions, which equals $1 + \frac{1}{3} r^2 (1-p_c)\left[n_b(p_c+2) + \frac{3}{r}\right]$. Thus the total number of transmissions equals

$$T_{single_{IOc}} = 1 + r(1-p_c)(rn_b+1) + \left(\left\lceil \frac{l_d}{r} \right\rceil - 1\right) \cdot \left[1 + \frac{1}{3} r^2 (1-p_c)\left(n_b(p_c+2) + \frac{3}{r}\right)\right] + \frac{l_d}{1-p_c}. \tag{11}$$

Since except for the first step of the query forwarding, we need to consider the overlapping region nodes for the remaining steps. Also, note that to reply to the sink through the reverse path requires $\frac{l_d}{1-p_c}$ transmissions.

Multiple-path Approach: In this approach, multiple active nodes forward the same query to the candidate nodes and reduce the possibility of the failure due to the wireless link loss. Since, the probability of lossy links, p_c is small, so in the analytical models, we consider that $1 - p_c^n \approx 1, n \geq 2$, where n is the number of copies of the same query received by a candidate node. Results with this assumption are later compared to (and validated with) simulations in Section 5.5. Now, according to the query forwarding patterns as shown in Fig.3, for an even value of i, $0 \leq i \leq d$, the success probability equals

$$P_{multiple_{e_c}}(i) = (1-q_1)\left[\prod_{m=\frac{i}{2}}^{d-\frac{i}{2}}\left(1-q_{m+1}^{i+1-p_c}\left(1-p_{m+1}\right)^{4+3p_c}\right)\right]\cdot$$
$$\prod_{m=1}^{\frac{i}{2}-1}\left(1-q_{m+1}^{2m+1-2p_c}\left(1-p_{m+1}\right)^{4-4p_c}\right)\left(1-q_{d-m+1}^{2m+1}\left(1-p_{d-m+1}\right)^{4-2p_c}\right),$$

and for an odd value of i, $0 \leq i \leq d$, it equals $P_{multiple_{o_c}}(i)$, where the only difference with the above expression is the limits of the products (i.e., $\lceil\frac{i}{2}\rceil \leq m \leq d-\lceil\frac{i}{2}\rceil$ and $1 \leq m \leq \lceil\frac{i}{2}\rceil$ for the first and the second products respectively) Here, the terms having the form $1-q_y^x\left(1-p_y\right)^z$ in the above equation expresses the probability of forwarding the query at the y-th step of the query forwarding, where x and z are the number of nodes that perform greedy and probabilistic forwarding respectively.

Similarly, to compute the overhead, we consider the different forwarding patterns of the query as shown in Fig.3. Due to the lossy links, the query may fail to reach some nodes. So the overhead reduces due to the less number of participating nodes. Thus, using the Equations (7), the total number of transmissions required to forward the query to the source for an even value of i, $0 \leq i \leq d$, equals

$$T_{multiple_{e_c}}(i) = T_{multiple_e}(i)$$
$$-\,p_c\left[2\sum_{m=2}^{\frac{i}{2}}\left(1-q_m\right)+\sum_{m=1}^{d-i+1}\left(1-q_{\frac{i}{2}+m}\right)+4\sum_{m=2}^{\frac{i}{2}}p_m+3\sum_{m=1}^{d-i+1}p_{\frac{i}{2}+m}+2\sum_{m=1}^{\frac{i}{2}-1}p_{d-m+1}\right],$$

and for an odd value of i, $0 \leq i \leq d$, it equals $T_{multiple_{o_c}}(i)$, where the differences expression are the first term ($T_{multiple_o}(i)$) and the limits of the second term (i.e., $2 \leq m \leq \frac{i}{2}$, $1 \leq m \leq d-i+1$, $2 \leq m \leq \frac{i}{2}$, $1 \leq m \leq d-i+1$, and $1 \leq m \leq \frac{i}{2}-1$ respectively). Here, the second term of the above equation computes the overhead reduction due to the lossy links.

Now, similar to the Equation (8) and (9), the average success probability and the average number of transmissions can be computed. Here, the reply to the sink through the reverse path requires $\frac{d}{1-p_c}$ transmissions.

5 Simulations and Results

In this section, we validate our analytical models by conducting extensive simulations. The objective of these simulations is to compare the simulation results about the performance of the routing approaches with the analytical models. Following performance metrics are considered in the simulations:

1) *Success probability*, is the ratio of the total number of queries that reached the source over the total number of queries sent.
2) *Overhead* in terms of energy dissipation is the average number of transmissions required to forward a query successfully to the source and to get the reply using the reverse path.

3) *Path quality* in terms of path length increase factor is the ratio of the average length of the path discovered by a routing approach over the length of the shortest path from a set of sinks to the source. Here, the shortest path length between the source and any sink node in the set is the same. This metric is important for long-lived continuous queries.

5.1 Simulation Model

In our simulations, we use a $100m \times 100m$ grid with 10^4 sensor nodes placed at distance $1m$ from each other. Except for the border nodes, each sensor node is able to communicate with eight neighbors. For all simulations, the exponent of the phenomenon diffusion function, i.e., the parameter α, is set to 0.8. To be consistent with the analytical models, the information gradient is available in the whole network and the malfunctioning nodes are uniformly distributed with some arbitrary values.

The querier, i.e., the sink, and the source are different and can be any node. We use a flooding technique to find the set of sink nodes that are specific shortest distance away from the source. In the simulations, we use only single-value queries, that search for a specific value and have a single response.

The simulated protocol based on the single-path approach uses a look-ahead parameter $r = 1$. For $r = 1$, it can be easily shown from Fig.1 that $n_B = 8, L_{err} \approx 2$ and $n_H \approx 2.5$. So, using the expressions of Section 4.2, we get $n_C = \frac{2}{3} n_B \approx 5$ and $n_L \approx 2.5$. These parameter values are used in the analytic models of the single-path approach to compare the analytical results with the simulation results.

The simulated protocol based on the multiple-path approach uses a probabilistic diffusion function with exponent β as specified in [7] for probabilistic forwarding. Thus $p_j = f(j) = \frac{1}{j^\beta}$, where j is the hop count in the information gradient region and $\beta < \alpha$.

5.2 Query Success Rate i.e., Probability of Success

The basic single-path approach is not resilient to local maxima. The query success rate of the protocol based on this approach is shown in Fig. 4, where the analytical result closely matches with the simulation result. With the increase of malfunctioning nodes, local maxima increases and the query success rate drops close to zero. However, using a filter to avoid the nodes having arbitrarily high information, the query success rate of this approach can be improved. Detailed derivation for the filter and corresponding simulation results are presented in [15]. Also, The remaining results of the performance analysis of this approach are detailed in [15].

For the improved single-path approach, the query success rate of the routing protocols depends on the availability of high information candidate nodes. From Fig.5, it is obvious that the analytically results are more or less in line with the simulation results. The number of high information candidate nodes reduces with the exploration of more nodes, especially for large d and causes some minor differences between the analytical and the simulation results. The improved single-path approach is resilient to local maxima due to its selection policy for

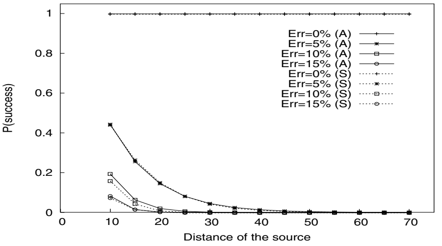


Fig. 4. Query success rate of the basic single-path approach. 'A' and 'S' indicate the analytical and the simulation results respectively

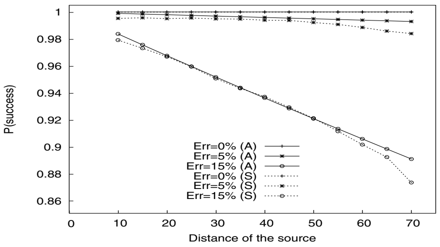


Fig. 5. Query success rate of the improved single-path approach

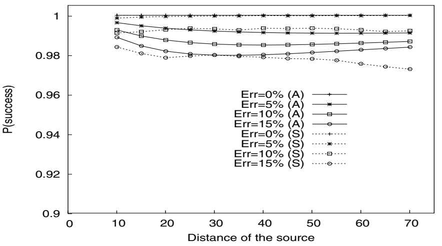


Fig. 6. Probability of success for the multiple path approach. The exponent of the probabilistic function $\beta = 0.65$.

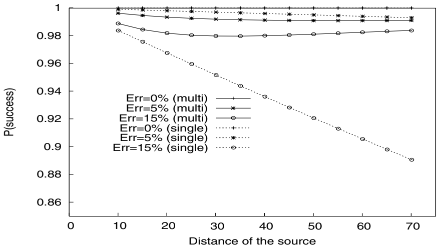


Fig. 7. Comparison of the query success rate of the improved single-path and the multiple-path routing approaches using analytical results. (Simulation results yield very similar plots)

the next active node. So, the query success rate of this approach is significantly higher than that of the basic single-path approach.

In the analytical model for the multiple-path approach, we consider that each step of the query is independent, and that low information candidate nodes forward the query probabilistically. However, due to correlation with previous steps of the query forwarding, some extra nodes may also forward the query and create few more extra paths, which actually improve the query success rate when less number of nodes are malfunctioning. Also, with the increase of malfunctioning nodes, active nodes use more probabilistic forwarding that results less number of paths and the query success rate drops. For these reasons, we notice some minor difference between the analytical and the simulation results in Fig.6.

The use of multiple paths and the probabilistic forwarding in the presence of malfunctioning nodes improves the query success rate of the multiple-path approach with compare to that of the single-path approach as shown in Fig.7. For the single-path approach, it is important to notice that the query success rate drops fast as the number of the malfunctioning nodes in the network increase.

5.3 Overhead i.e., Energy Dissipation

Fig.8 shows the overhead in terms of the average energy dissipation of the improved single-path approach. In both models, the overlapped region nodes respond only one time. The analytical and the simulation results are very similar. The minor discrepancy is due to following reason. With the exploration of more nodes, the number of high information candidate nodes reduces and the path length increases due to choosing the malfunctioning nodes to forward the query.

Both analytical and simulation results for the overhead of the multiple-path approach are given in Fig.9. The analytical results are quite close to the simulation results with a small discrepancy. In the analytical model, we assume that the sink is located at the last row of the grid i.e., an edge node as shown in Fig.1 and 3. However, in our simulation scenarios, the sink is not always at an edge node and the initial high value of the probabilistic function causes the exploration of some nodes in the low information gradient region. Also, as we have already explained in Section 5.2, due to correlations between the query forwarding steps, this routing approach creates some extra paths and increases the overhead which is not considered in the analytical model. Thus, simulation results show sightly more overhead than the analytical results.

In Fig.10, the overhead of both approaches is compared using the analytical results. It is obvious that in our model, if the source is less than 22 hops away from the sink then the multiple-path approach is more energy efficient; otherwise, the single path approach is preferable when energy dissipation is only considered. The overhead of the multiple-path approach increases more due to the extra paths created by probabilistic forwarding.

Using analytical models, the percentage of energy savings of the multiple-path approach over the single-path approach is shown in Fig.11. As the number of malfunctioning nodes increase, the overhead of the single-path approach increases. Since, the length of the path followed by this approach increases. On the other hand, with the increase of malfunctioning nodes, the multiple-path approach uses more probabilistic forwarding. This creates less number of paths, and the overhead reduces.

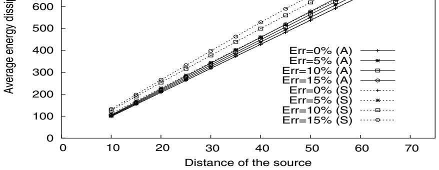


Fig. 8. Overhead for the single path approach. 'A' and 'S' indicate the analytical and the simulation results respectively

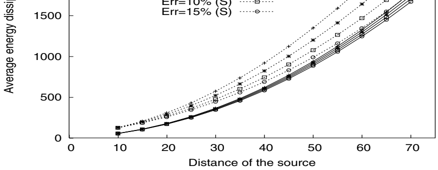


Fig. 9. Overhead for the multiple path approach. The exponent of the probabilistic function is $\beta = 0.65$

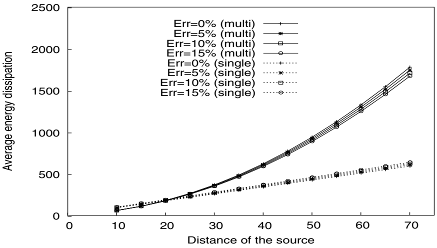


Fig. 10. Comparison of the overhead of the improved single-path and the multiple-path approaches using analytical results

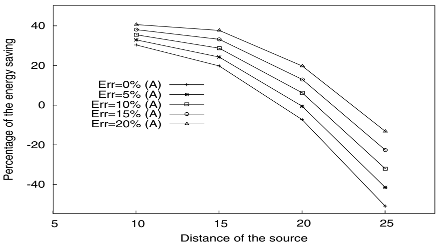


Fig. 11. Percentage of energy saving of the multiple path approach over the improved single-path approach for $d \leq 25$ using analytical models

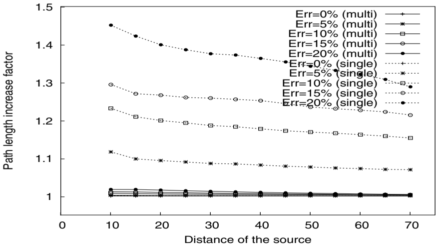


Fig. 12. Path length increase factor for the improved single-path and the multiple-path approaches. The exponent of the probabilistic function $\beta = 0.60$

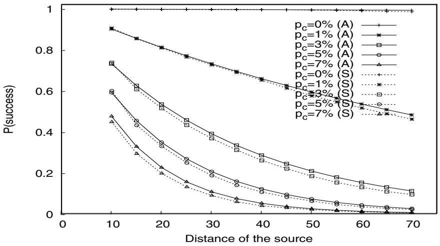


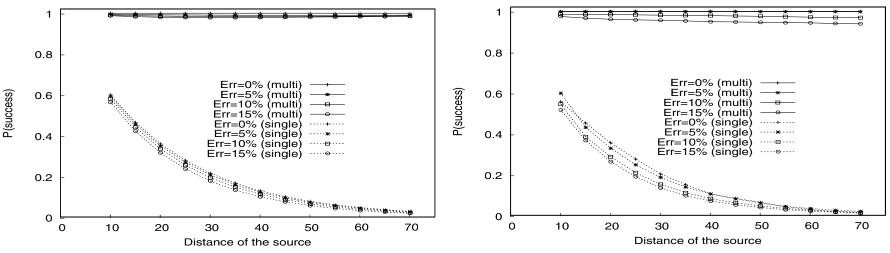
Fig. 13. Query success rate of the improved single-path approach with the varying lossy link conditions. The probability of malfunctioning nodes is $p_f = 0.05$

5.4 Path Quality

Fig.12 shows the path length increase factor of both routing approaches. The multiple-path approach results the shorter paths which are very close to the shortest path length. We notice that the path length for the single-path approach increases with the increase of the malfunctioning nodes. As expected, in the presence of malfunctioning nodes, the single path approach fails to follow the shortest path towards the source. On the other hand, the instantiation of multiple paths and probabilistic forwarding help the multiple path approach to alleviate the problem for malfunctioning nodes.

5.5 Wireless Link Loss Effect

So far, all the results presented in Section 5 consider ideal wireless links. However, the wireless links are lossy and likely affect the query success rate of the routing protocols significantly. Fig.14 shows the query success rate of both approaches in the presence of lossy links with probability $p_c = 0.05$. Here, the analytical (Fig.14(a)) and simulation (Fig.14(b)) results are identical. In both cases, the query success rate of the multiple-path approach drops quite slowly and it is



(a) Using analytic models. (b) Using simulation models.

Fig. 14. Comparison of the query success rate of the improved single-path and the multiple-path approaches in the presence of link loss, $p_c = 0.05$. The exponent of the probabilistic function $\beta = 0.65$

more than 93% even at the presence of 15% malfunctioning nodes in the sensor network. The ability to send the same query from multiple active nodes towards a candidate node improves the resilience of this approach significantly.

For the improved single-path approach, the query success rate drops drastically with the increase of the distance between the source and the sink. Also, Fig.13 shows that the query success rate drops further with the increase of loss probability of the lossy links, i.e., p_c. From Equation (10), it is obvious that the term $(1 - p_c)^r$, which corresponds to forwarding the query from the current active node to the next active node in the presence of lossy links, is responsible for the low success rate of this approach. Using ARQ, the query success rate can be improved significantly [15].

6 Summary and Conclusions

In this paper, we have presented a detailed performance analysis of information-driven routing approaches in ideal and lossy wireless link conditions using analytical models and simulations. We consider the effect of (various kinds of) noise, malfunctioning nodes and node failures in our analysis.

From our study, it is found that the query success rate of the single-path approach drops quite fast as the number of malfunctioning nodes in the network increase while the multiple-path approach retains very high query success rate. Also, it is found that the multiple-path approach is more energy efficient when the source is less than 22 hops away from the sink; otherwise, the single-path approach is more energy efficient. For example, in Fig. 11, for 5% malfunctioning nodes and a source 15 hops away from the querier, the overhead of the multiple-path approach is only 75% of that of the improved single-path approach. Further, the multiple-path approach results in shorter paths which are close to the shortest path. Finally, in the lossy link case, the query success rate of the single path approach drops drastically with the increase of the link loss probability and the distance between the source and the querier. On the other hand, the multiple-path approach achieves over 93% success rate even at the presence of 15% malfunctioning nodes in the sensor network.

The analytical models of both routing approaches can be used to determine the performance and the bottleneck of a protocol for a large sensor network without simulations or when simulations are not possible due to resource constraints. Further, the performance of a new protocol, based on either of these two approaches, can be determined using our models.

From the analytical models, it is obvious that more efficient information-driven routing protocols can be designed based on these two approaches by tuning the parameters of the models, which can be one possible future research direction.

References

1. C. Intanagonwiwat, R. Govindan and D. Estrin, “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks”, MobiCom 2000.
2. C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, “Impact of Network Density on Data Aggregation inWireless Sensor Networks”, ICDCS’02.
3. D. Braginsky and D. Estrin, “Rumor Routing Algorithm for Sensor Networks”, WSNA 2002.
4. N. Sadagopan, B. Krishnamachari, and A. Helmy, “Active Query Forwarding in Sensor Networks (ACQUIRE)”, SNPA 2003.
5. M. Chu, H. Haussecker, and F. Zhao, “Scalable Information-Driven Sensor Querying and Routing for ad hoc Heterogeneous Sensor Networks”, Int’l J. High Performance Computing Applications, 16(3):90-110, Fall 2002.
6. J. Liu, F. Zhao, and D. Petrovic, “Information-Directed Routing in Ad Hoc Sensor Networks”, WSNA 2003.
7. J. Faruque, A. Helmy, “RUGGED: RoUting on finGerprint Gradients in sEnsor Networks”, IEEE International Conference on Pervasive Services (ICPS), 2004.
8. Q. Li, M.D. Rosa and D. Rus, “Distributed Algorithms for Guiding Navigation across a Sensor Network”, MobiCom 2003.
9. S. D. Servetto and G. Barrenechea, “Constrained Random Walks on Random Graphs: Routing Algorithms for Large Scale Wireless Sensor Networks”, WSNA 2002.
10. Fan Ye, Gary Zhong, Songwu Lu and Lixia Zhang, “GRAdient Broadcast: A Robust Data Delivery Protocol for Large Scale Sensor Networks”, ACM WINET (Wireless Networks).
11. A. Woo, T. Tong, and D. Culler. “Taming the Underlying Issues for Reliable Multhop Routing in Sensor Networks”. ACM SenSys, November 2003.
12. J. Zhao and R. Govindan. “Understanding Packet Delivery Performance in Dense Wireless Sensor Networks”. ACM Sensys, November 2003.
13. D.R. Askeland, The Science and Engineering of Materials, PWS Publishing Co., 1994.
14. J.F. Shackelford, Intro to Materials Science For Engineers, 5th Ed., Prentice Hall, 2000.
15. J. Faruque, K. Psounis, A. Helmy, “Analysis of Gradient-based Routing Protocols in Sensor Networks”, CS Technical report, USC 2005.