FLOW: An Efficient Forwarding Scheme to Mobile Sink in Wireless Sensor Networks

Rahul Urgaonkar and Bhaskar Krishnamachari
Department of Electrical Engineering-Systems
University of Southern California, Los Angeles, CA 90089
Email: {urgaonka, bkrishna}@usc.edu

***Abstract*— We propose a learning-based approach to efficiently forward data to a mobile sink in a wireless sensor network. Specifically we assume a push application where the mobile sink does not initiate the query. Furthermore, the sink moves in a randomized pattern within the sensor field. In the presented scheme, *Moles* (nodes that sense the sink) learn the sink's movement pattern over time and statistically characterize it as probability distribution functions. This information collected at the moles is used in a distributed fashion to calculate the likelihood of a node being on a good path to the sink which is used to determine whether to forward data through the node.**

I. INTRODUCTION

Consider a forest patrol that makes periodic patrols along certain paths in the forest. These patrols aim at preventing poaching and monitoring endangered wildlife. However, their efficacy is limited because they cannot cover the entire area assigned to them. This problem can be solved to an extent by deploying a wireless sensor network that senses unusual events in the area. Information about these events needs to get routed to the moving patrol. Besides flooding, a naive way of doing this is to route to a stationary base-station at the periphery of the network and then directly transmit to the sink. However this is not always desirable because of energy and delay constraints. Instead, a better approach is to route data to the patrol (mobile sink) while it is *on the move*. Note that the patrol does not know when to query. Periodic query might not be feasible due to the energy and delay constraints as well as the fact that that unusual events are not likely to happen very often. In such situations, it is desirable for the mobile sink to passively listen for incoming data from the sources. Routing decisions can exploit the fact that patrols move in a certain pattern that can be characterized by some spatio-temporal relation. Over time, the network can learn this pattern and by adding a temporal dimension to the routing decision, data can be pushed towards the moving patrol with a degree of confidence about energy efficiency and delivery guarantees.

Forwarding using Likelihood-based Weights (FLOW) makes use of the underlying pattern in the sink's movement to discover *good* data delivery paths. We define *Moles* to be the nodes that lie in the vicinity of the path that the sink takes. It is assumed that a Mole can somehow detect the presence of the sink near itself. Every mole characterizes the sinks presence in its vicinity as a probability distribution. These are then used by each node to calculate its likelihood of being on a good path to the sink in a distributed fashion using only *local* information obtained from the neighboring nodes. Forwarding decisions at each node are made using these values with the intent of forwarding to a node/set of nodes that are most likely to lead to the sink.

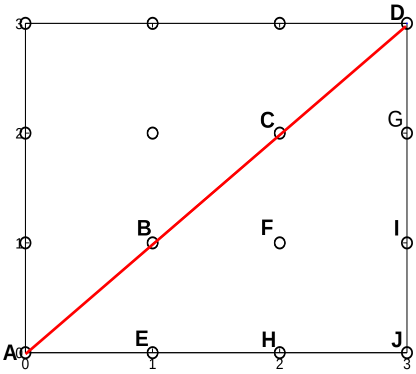

Fig. 1. A section from a Grid-based sensor field showing the Moles A, B, C, D and source nodes E, F, G, H, I, J

II. FLOW: FORWARDING USING LIKELIHOOD-BASED WEIGHTS

Consider a deployment of sensor nodes in a grid-based structure as shown in Fig. 1. Each sensor node is assumed to have a unit communication range so that it can communicate with nodes at its left, right, top and bottom. The mobile sink moves periodically along the trajectory as shown. The nodes that lie on this are assumed to be the Moles (A, B, C, D in Fig. 1). We construct a hierarchical structure of nodes by assigning the level k to nodes that are k hops away from a Mole. Thus, nodes E, F, G belong to level 1, nodes H, I belong to level 2 and node J belongs to level 3 in Fig. 1.

Define the event A_i as the event that the sink is in the vicinity of Mole A between time t_i and t_{i+1}. The probability of this event is denoted by $P_i(A)$. In a similar fashion, $P_i(B)$, $P_i(C)$, $P_i(D)$ etc. can be defined. With this set-up, we define the event E_i at node E on level 1 as the event that the sink is in the vicinity of Mole A or Mole B or both, i.e., $E_i = A_i \cup B_i$. The probability of this event can be calculated at E as:

$$P(E_i) = P(A_i \cup B_i) = P_i(A) + P_i(B) - P_i(A_i \cap B_i) \quad (1)$$

Similarly, at node H, we can define the event $E_i \cup F_i =$

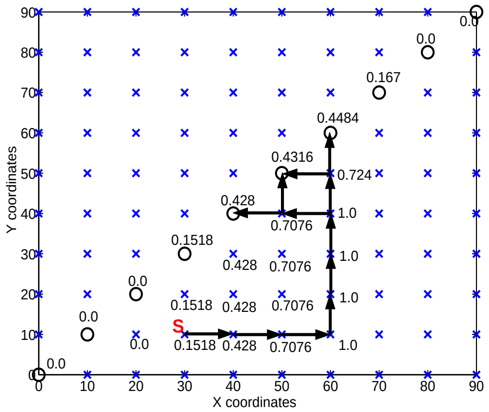

Fig. 2. The set of routes obtained by FLOW at time slot 50 for source node S. $\epsilon_1 = 0.9$, $\epsilon_2 = 0.25$

$A_i \cup B_i \cup C_i$. The probability of this event will be:

$$\begin{aligned} P(H_i) &= P_i(A) + P_i(B) + P_i(C) \\ &- P_i(A_i \cap B_i) - P_i(B_i \cap C_i) - P_i(C_i \cap A_i) \\ &+ P_i(A_i \cap B_i \cap C_i) \end{aligned} \quad (2)$$

Since the sink cannot *simultaneously* be in the vicinity of A and C, the above expression reduces to

$$\begin{aligned} P(H_i) &= P_i(A) + P_i(B) + P_i(C) \\ &- P_i(A_i \cap B_i) - P_i(B_i \cap C_i) \\ &= P(E_i) + P(F_i) - P(B_i) \end{aligned} \quad (3)$$

In a similar fashion, $P(J_i) = P(H_i) + P(I_i) - P(F_i)$

In general, for any source node S, we have

$$P(S_i) = P(S_i^{LEFT}) + P(S_i^{TOP}) - P(S_i^{TOPLEFT}) \quad (4)$$

where S_i^{LEFT} and S_i^{TOP} are the neighbors of S in the lower level while $S_i^{TOPLEFT}$ is their common neighbor. Thus, each node can calculate the probability of the event associated with it using only *local* information. We use this property to propose a forwarding scheme to the mobile sink.

Note that for a source node S with coordinates (X, Y), $P(S_i)$ denotes the probability of the sink being in the vicinity of the Moles that lie on the portion of trajectory that lies inside the region $(x < X, y > Y)$ of $\Re^2$. We call this region the *likelihood region* of S.

In this scheme, first a hierarchical structure is created where nodes that are the same number of hops away from the Moles belong to the same level. Once such a structure is created, the probability of the event associated with each source node is calculated using the values obtained from the neighboring nodes at the lower level in the hierarchy. Thus, the nodes at higher levels get progressively updated.

A source node S that wishes to send data to the mobile sink uses the following algorithm to decide which neighbor to forward its packet to.

1) $P_i(S) > \epsilon_1$: This means that the probability of the sink being in the likelihood region of S is at least ϵ_1. In this case, the node S calculates $P_i(S_{TOP}) - P_i(S_{LEFT})$ and $P_i(S_{LEFT}) - P_i(S_{TOP})$. This can lead to three cases:
    a) $P_i(S_{TOP}) - P_i(S_{LEFT}) > \epsilon_2$: This means that the probability of the sink being in the likelihood region of S_{TOP} is more than that of S_{LEFT} by ϵ_2. The source node forwards the packet only to S_{TOP}.
    b) $P_i(S_{LEFT}) - P_i(S_{TOP}) > \epsilon_2$: This means that the probability of the sink being in the likelihood region of S_{LEFT} is more than that of S_{TOP} by ϵ_2. The source node forwards the packet only to S_{LEFT}.
    c) $P_i(S_{TOP}) - P_i(S_{LEFT}) < \epsilon_2$: This means that the sink is likely to be in the likelihood regions of *both* S_{TOP} and S_{LEFT} and the source node forwards the packet to both of them.
2) $P_i(S) < \epsilon_1$: In this case, the probability of the sink being in the likelihood region of S is low. It threfore forwards it to one of the nodes *higher* up in the hierarchy. It selects between X_{RIGHT} and X_{BOTTOM}, the node with higher probability.

All other nodes that receive a packet forward it using the same algorithm, except that if they have received the packet from a node higher in the hierarchy, they strictly forward it to a node lower in the hierarchy.

Next we present preliminary simulation results for the proposed forwarding approach. We considered a grid-based deployment of 100 nodes. The mobile sink is assumed to move along the trajectory in such a way that the PDFs learnt by the Moles are Gaussian in nature. The average time of completion of a tour is taken to be 100 units. The probabilities at each source node are obtained for all time intervals using Eqn. 4. In the first set of experiments, a node is randomly chosen as the source node and it is scheduled to send data to the sink at a randomly chosen time interval within the tour duration. The FLOW algorithm is used at each node to obtain a set of paths to the moles. The outcome of one such experiments is shown in Fig. 2 (time = 50). The likelihood values for this time instant at each node are also indicated. The values of ϵ_1 and ϵ_2 were chosen to be $\epsilon_1 = 0.9$ and $\epsilon_2 = 0.25$.

It can be seen from Fig. 2 that FLOW finds a set of paths from the source node to those moles that have a high probability of being in the sink's vicinity. Furthermore, when the probabilities are similar, the algorithm selects *both*

TABLE I

Delivery Ratios and Energy Cost for different combinations of ϵ_1 and ϵ_2. Source node and time interval fixed

ϵ_1, ϵ_2	$\epsilon_1 = 0.25$	$\epsilon_1 = 0.50$	$\epsilon_1 = 0.75$	$\epsilon_1 = 1.00$
$\epsilon_2 = 0.25$	(0.2, 3)	(0.4, 4)	(0.5, 6)	(0.8, 10)
$\epsilon_2 = 0.50$	(0.2, 3)	(0.4, 6)	(0.6, 9)	(0.85, 11)
$\epsilon_2 = 0.75$	(0.2, 3)	(0.45. 6)	(0.7, 11)	(0.9, 12)
$\epsilon_2 = 1.00$	(0.2, 3)	(0.45, 6)	(0.7, 11)	(1.0. 14)

neighboring nodes. This results in *multipath* routing which offers more robustness over single path routing.

To capture the effects of the parameters ϵ_1 and ϵ_2, a source node and a time domain was fixed and for different combinations of ϵ_1 and ϵ_2, delivery ratio and energy-cost associated with the set of paths generated by the algorithm were obtained. These are shown in Table I. We define delivery ratio as the *Union* of the probabilities of all Moles that received the packet. It quantifies the probability of successfully delivering to the sink. The energy cost was obtained as the cost of the multi-path mesh structure generated by the algorithm, assuming unit cost per edge. It can be seen that there exists a trade-off between the delivery ratio and energy-cost. This trade-off is nicely captured by the tunable parameters ϵ_1 and ϵ_2.

The key advantages offered by this scheme are:

1) *Energy Efficieny* : FLOW offers energy efficiency because it takes advantage of the underlying patten of the sink's movement to select only those paths that lead to moles with high likelihood values.
2) *Distributed Nature* : In this scheme, each node can use only local information to calculate its likelihood values (Eqn. 4).
3) *Scalability* : Since each node needs to maintain information about only its local neighborhood, this scheme is scalable in terms of storage requirements at each node.
4) *Robustness* : The multi-path nature of the routes discovered makes this approach robust to individual node failures.
5) *Tunable parameters* : It can be seen from Table I that there exists a trade-off between the delivery ratio and enegy-cost. This trade-off is nicely captured by the tunable parameters ϵ_1 and ϵ_2.

We would like to generalize the current setting that considers a regular grid-based sensor field to random sensor fields as part of future work.

ACKNOWLEDGEMENTS

We would like to thank Pritam Baruah, Department of Computer Science, USC for many useful discussions.

REFERENCES

[1] F. Ye, G. Zhong, S. Lu, L. Zhang, "GRAdient Broadcast: A Robust Data Delivery Protocol for Large Scale Sensor Networks," *to appear in ACM Wireless Networks (WINET)*, Vol. 11, No.2, March 2005