

# Sensor-Centric Quality of Routing in Sensor Networks

Rajgopal Kannan\*, Sudipta Sarangi†, S. S. Iyengar\* and Lydia Ray\*

\*Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803, Email: rkannan@csc.lsu.edu

† Department of Economics, Louisiana State University, Baton Rouge, LA 70803

**Abstract**—Standard embedded sensor network models emphasize energy efficiency and distributed decision-making by considering untethered and unattended sensors. To this we add two constraints - the possibility of sensor failure and the fact that each sensor must tradeoff its own resource consumption with overall network objectives. In this paper, we develop an analytical model of data-centric information routing in sensor networks under all the above constraints. Unlike existing techniques, we use game theory to model *intelligent* sensors thereby making our approach *sensor-centric*. Sensors behave as rational players in an N-player routing game, where they tradeoff individual communication and other costs with network wide benefits. The outcome of the sensor behavior is a sequence of communication link establishments, resulting in routing paths from reporting to querying sensors. We show that the optimal routing architecture is the Nash equilibrium of the N-player routing game and that computing the optimal paths (which maximizes payoffs of the individual sensors) is NP-Hard with and without data-aggregation. We develop a game-theoretic metric called *path weakness* to measure the qualitative performance of different routing mechanisms. This sensor-centric concept which is based on the contribution of individual sensors to the overall routing objective is used to define the *Quality of Routing (QoR)* paths. Simulation results are used to compare the QoR of different routing paths derived using various energy-constrained routing algorithms.

## I. INTRODUCTION

Embedded Sensor Networks are distributed systems for sensing and *in situ* processing of spatially and temporally dense data from resource-limited and harsh environments such as seismic zones, ecological contamination sites or battlefields [1]. Sensors execute tasks by routing and cooperatively processing sensed information. Information routing in sensor networks is primarily data-centric in nature. Interest queries originating from sink nodes are disseminated over the network resulting in responses from those sensors whose sensed information satisfy the query attributes. The technique of data aggregation is used to solve the problems of data implosion and overlap [7].

Sensors in embedded sensor networks operate under a set of unique and fundamental constraints which make collaborative information routing challenging.

- 1) Sensors are untethered.
- 2) Sensors are unattended.

These two constraints imply that nodes must utilize their unreplaceable and limited energy resources efficiently. For example, too many sensors being active at the same time will lead to increased energy consumption and competition

for communication resources. Additionally, nodes must make decisions independently without recourse to a central authority because of the energy needed for global communication and latency of centralized processing. Thus ensuring the effective use of collected sensor data will require the development of scalable, self-organizing, and energy-efficient solutions for data dissemination through aggregation.

Designing a sensor network that only takes into account the first two constraints will not always lead to optimal architectures. There are many applications where sensors are deployed in hazardous and hostile environments in which they can fail to operate or be destroyed with certain probabilities. Wireless sensor networks are also extremely vulnerable to data loss under denial of service (DoS) attacks [10]. In these cases the task of routing a query response from observing sensors to querying nodes should not be compromised by the inhospitability of the environment. Consider sensor networks for monitoring environmentally toxic situations, or seismic sensor networks in earthquake or rubble zones or even sensors in military battlegrounds under enemy threat. For such networks to carry out their tasks meaningfully, sensors must route strategic and time-critical information via the most reliable paths available. Hence in this paper, we introduce an additional constraint.

3. Sensor  $s_i$  can fail with probability  $q_i$ .

When a sensor node loses its energy (or is destroyed), it is unlikely to be replaced. The information utility of the sensor network (in terms of data collecting and processing ability) decreases as nodes die out. Thus, implicit in the operation of an embedded sensor network is a *fourth constraint*: To maximize network utilization and information viability, sensors must cooperate to achieve network wide objectives while maximizing their individual lifetimes<sup>1</sup>. We label this paradigm for broad sensor network operation as **sensor-centric**.

While there are many popular routing algorithms for sensor networks for minimizing energy consumption, (MECN [8] and diffusion routing [5], for example), in this paper, we analyze sensor-centric routing, i.e, routing within the bounds of all the four constraints mentioned above. The choices for untethered, unattended and unreliable sensors when seen from

<sup>1</sup>Our assumption is that the longer individual sensors survive, the better it is for the sensor network.

this perspective are a natural fit for a game-theoretic framework. Sensors are modeled as rational/intelligent agents that cooperate to find optimal network architectures that maximize their payoffs i.e., benefits to the network of this sensor's action minus individual costs (as opposed to aggregate path costs), in a network game.

The central feature of our sensor-centric paradigm is that sensors are rational and driven by self-interest. Ideally sensors should route over the most reliable paths while minimizing their own power/energy consumption rather than some aggregate energy criterion. This model of reliable energy-constrained routing has three benefits<sup>2</sup>: *First*, it is in the interests of long-term network operability that nodes survive even at the expense of somewhat longer (but not excessively so!) paths. The network will be better served when a critical sensor can survive longer by transmitting via a cheaper link rather than a much costlier one for a small gain in reliability or delay. *Second*, it takes the cost distributions of individual sensors into account while choosing good paths. The advantages of modeling rational, self-interested sensors can be seen easily from the following example. Given a path involving three sensors with absolute communication costs in the low, medium and high ranges respectively, choosing a reliable path subject to minimizing overall costs might lead to the first two nodes having to select their highest cost links as the third node is dominant in the overall cost. This would run counter to the long-term operability goal of the network. *Third*, it incorporates the extreme case when sensors only have limited and local network state information (about neighbors and link costs, for example). In this case, when information is received, a node should choose to route to the cheapest neighbor in the absence of further state information.

In data-centric routing [7], data aggregation or data fusion is used to reduce the problems of data implosion and overlap. Here, the sensor network can be perceived as a reverse multicast tree with information aggregated or fused at intersecting nodes and routed to the sink node at the root. In [9], the authors describe data-centric routing algorithms for sensor networks that take energy constraints and quality of service considerations into account. In this paper, we formalize this concept by developing a new analytical model of information routing in sensor networks. Unlike existing techniques, we use game theory to model *intelligent* sensors thereby making our approach *sensor-centric*. This sensor-centric paradigm can be applied in parallel to the data-centric information flow model. We consider a model of additive data aggregation at intersecting nodes, based on information value quantification. We show that the optimal routing tree is the Nash equilibrium [3] of the N-player routing game and that computing the

<sup>2</sup>Note that while we model reliable energy-constrained routing in this paper, our model can be extended to other network optimization criteria such as latency also. For example, we can let  $q_i$  be the probability that a given delay bound is exceeded at sensor  $s_i$  and assume a message is lost if the delay bound is exceeded at any node. This is analogous to the sensor failure probability  $q_i$  in the reliability model. More complicated models that take into account correlated and cumulative delay violation probabilities over a series of sensors can be derived, which we do not consider in this paper.

optimal paths/tree (which maximizes payoffs of the individual sensors) is *NP-Hard with and without data-aggregation*.

This leads us to consider two important questions. First, are there easily computable routing algorithms which produce approximately optimal routing paths? Secondly, in a sensor-centric network what is an approximately optimal routing path? There is as yet no formal framework for quantifying and comparing the merits of different routing algorithms in terms of the *Quality of Routing* (QoR) paths obtained. We use the term QoR path from the game-theoretic or individual sensor's perspective rather than the well known Quality of Service (QoS) based path (shortest path, for example) which is an end-to-end concept. Given the increasing prevalence of networks with 'smart' components, it is necessary to evaluate the performance gain of individual components within the overall objective. Traditional measures such as quality of service do not suffice in capturing this concept. Therefore we require new techniques for computing the QoR of routing paths, i.e. ranking them. At a more specific level, given that the optimal path is a vector of payoffs of individual nodes, how do we characterize approximately optimal paths?

In this paper, we derive a game-theoretic path performance metric labeled *path weakness*. We use this to evaluate standard routing techniques based on aggregate payoffs as well as the suboptimality of *any* routing path from the point of view of individual sensor payoffs. We address the following issues: How well do standard distributed routing algorithms perform when compared to the optimal analytical solution. Can we quantify the tradeoff of saving network state transmission overheads in a particular routing algorithm with the quality of routing paths (i.e., their weakness) obtained? Are there distributions of costs, probabilities and values under which some routes are 'less weaker' than others.

We summarize the contributions of this paper below:

- A game-theoretic model of routing in sensor networks is developed. Rational, intelligent sensors select routing paths by evaluating the trade-offs between reliability and the costs of communication.
- A sensor-centric paradigm for evaluating the quality of routing trees for data-aggregated routing in sensor networks, is proposed. This QoR concept captures the participation suboptimality of a node on the given tree, i.e., how much would a node gain by deviating from the current tree to an optimal one. A routing heuristic based on a team version of the routing game called *Team-RQR* is presented.
- Analytical results on the complexity of computing paths with bounded weakness are derived along with some sufficient conditions on costs and probabilities for well known routing algorithms such as most reliable path and least cost neighbor to be congruent to the optimal sensor-centric route.
- Simulation results comparing the QoR of paths obtained using some well known routing algorithms and identifying ranges of costs and probabilities in which they perform favorably are shown.

The paper is organized as follows: Section 2 describes our game-theoretic model set-up. Section 3 contains analytical as well as complexity results on path congruence and optimal path computability. Section 4 explains the Quality of Routing paradigm and some theoretical QoR complexity results. Simulation results comparing the QoR of different algorithms are also presented in Section 4. Finally, Section 5 concludes the paper.

## II. THE MODEL

We model data-centric routing with data-aggregation in sensor networks. In data-centric routing, interest queries are disseminated through the network to assign sensing tasks to sensor nodes. Attribute based naming is used to resolve these queries by using the attributes of the phenomenon to trigger responses from appropriate sensor nodes. Further, data aggregation at intersecting nodes can be used to reduce implosion and overlap problems in the network. With data-aggregation, the sensor network can be perceived as a reverse multicast tree with information fused at intersecting nodes and routed to the sink node at the root.

Let  $S = \{s_1, \dots, s_n\}$  denote the set of sensors, modeled as players in a routing game to be defined below, with generic members  $i$  and  $j$ . For ordered pairs  $(i, j) \in S \times S$ , the shorthand notation  $ij$  is used. Sensor  $s_i$  has information (data) of value  $v_i$  which it wishes to send to the sink node  $s_q = s_n$ , where  $v_i \in \mathbb{R}^+$  represents an abstract quantification of the value of the event sensed at node  $s_i$ ,  $1 \leq i \leq n$ . Also,  $v_i = 0$  for nodes whose sensed information does not satisfy the specified attributes of the query. Information is routed to  $s_q$  through an optimally chosen set  $S' \subseteq S$  of intermediate nodes by forming neighbor communication links. Link formation occurs by a process of simultaneous reasoning at each node leading to a path from each  $s_i$  with nonzero value  $v_i$  to  $s_q$ . For untethered sensor networks, communication energy costs are a significant constraint. We account for this by modeling link formation as costly. Each node incurs a cost  $c_{ij} > 0$  for each link  $ij$  it establishes. This link cost is an abstraction of message transmission costs in terms of required transmission power or available on-field sensor battery life.

Our routing model is rigorous enough to account for cases when some sensors can choose to participate or not participate in this routing process. By incorporating a participation cost to each sensor, we can analytically model situations where a certain proportion of sensors switch themselves off (perhaps based on neighborhood density as proposed in [2]) to conserve energy<sup>3</sup>. Further, our model selects routing path based on the ‘importance’ of the query being reported. For example, urgent messages must be treated differently and routed over more reliable paths even at higher costs. These two features of our model allow sensors to rationally decide (by computing

<sup>3</sup>In this paper we do not consider the protocol required to implement this participation mechanism, perhaps through exchange of ‘permission to transmit’ messages. Our objective is to consider routing implications of this abstraction of individual sensor self-interest.

individual payoffs) whether or not to participate in routing data of a given significance.

We assume that node  $s_i$  can fail with a probability  $(1-p_i) \in [0, 1)$ . We make no assumptions about correlations in these probabilities while formulating our abstract model, since the model primarily requires the values of path reliability, which we assume can be obtained<sup>4</sup>. For ease of calculation in our simulations (Section 4), we do assume independent failure probabilities. Also, for simplicity, we assume that the sink node  $s_q$  never fails.

Thus the graph  $G = (S, E, P, C)$  represents an instance of a data-centric sensor network in which data of value  $v_i$  is to be optimally routed from node  $s_i$  to node  $s_q$ , with  $S$  the set of sensors interconnected by edge set  $E$ ,  $P(s_i) = p_i$  the node success probabilities and  $C(s_i, s_j) = c_{ij}$ , the cost of links in  $E$ . We denote a path from any node  $s_a$  to  $s_b$  in  $G$  by the node sequence  $(s_a, s_2, \dots, s_b)$ .

In this context, we define the following problem called **Reliable Query Reporting (RQR)**: Given that data transmission in the network is costly and nodes are not fully reliable, how can we induce the formation of a maximally reliable data aggregation tree from reporting sensors (sources) to the querying (sink) node, where every sensor is ‘smart’ and motivated by self-interest, i.e., it can trade-off individual costs with network wide benefits. This optimal data aggregation tree will naturally be distinct from standard multicast trees such as the Steiner tree or shortest path trees which minimize overall network costs and therefore cannot represent the outcome of self-interested sensors. The solution to this problem lies in designing a routing game with payoff functions such that its Nash equilibrium corresponds to the optimally reliable data aggregation tree. We now describe the different components of this strategic game.

**Strategies.** Each node’s strategy is a vector  $l_i = (l_{i1}, \dots, l_{ii-1}, l_{ii+1}, \dots, l_{in})$  and  $l_{ij} \in \{0, 1\}$  for each  $j \in S \setminus \{i\}$ . The value  $l_{ij} = 1$  means that nodes  $i$  and  $j$  have a link initiated by  $i$  whereas  $l_{ij} = 0$  means that sensor  $i$  does not send information to  $j$ . The set of all pure strategies of player  $i$  is denoted by  $\mathcal{L}_i$ . We focus only on pure strategies in this paper. Given that node  $i$  has the option of forming or not forming a link with each of the remaining  $n - 1$  nodes, the number of strategies available to node  $i$  is  $|\mathcal{L}_i| = 2^{n-1}$ . The strategy space of all nodes is given by  $\mathcal{L} = \mathcal{L}_1 \times \dots \times \mathcal{L}_n$ . Notice that there is a one-to-one correspondence between the set of all directed networks with  $n$  vertices or nodes and the set of strategies  $\mathcal{L}$ . In order to keep the analysis tractable, in this model we assume that each node can only establish one link. Note that while diffusion routing based algorithms start off with nodes sending query responses to the sink over multiple paths [5], eventually a single route is established once interest gradients are determined. Our objective in this paper is to compare and evaluate these final routing paths

<sup>4</sup>While we assume static failure probabilities in developing our model, a dynamic extension would view the network in terms of failure probability snapshots in successive operational periods.

from the game-theoretic optimality point of view and hence our restriction is valid. Further, routing loops are avoided by ensuring that strategies resulting in a node linking to its ancestors yield a payoff of zero and are thus inefficient. Under these assumptions each strategy profile  $l = (l_1, \dots, l_n)$  becomes a reverse tree  $\mathcal{T}$ , rooted at the sink  $s_q$ . We now proceed to model the payoffs in this game.

A standard noncooperative game assumes that players are *selfish* and are only interested in maximizing their own benefits. This poses a modeling challenge as we wish to design a decentralized information network that can behave in a collaborative manner to achieve a joint goal while taking individual operation costs into account. Since the communal goal in this instance is reliable data transmission, the benefits to a player must be a function of path reliability but costs of communication need to be individual link costs.

**Payoffs.** Consider a strategy profile  $l = (l_i, l_{-i})$  resulting in a tree  $\mathcal{T}$  rooted at  $s_q$ , where  $l_{-i}$  denotes the strategy chosen by all the other players except player  $i$ . Since every sensor that receives data has an incentive in its reaching  $s_q$ , the benefit to any sensor  $s_i$  on  $\mathcal{T}$  must be a function of the path reliability from  $s_i$  onwards. Since the network is unreliable, the benefit to player  $s_i$  should also be a function of the expected value of information at  $s_i$ . Hence we can write the payoff at  $s_i$  as:

$$\Pi_i(l) = \begin{cases} g_i(v_1, \dots, v_{n-1})R_i - c_{ij} & \text{if } s_i \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases}$$

where  $R_i$  denotes the path reliability from  $s_i$  onwards to  $s_q$  and  $g_i$  the expectation function, is explained below.

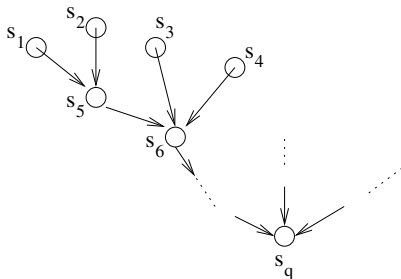


Fig. 1. Payoffs with data aggregation.

Consider the data-aggregation tree shown in Fig. 1. Let  $V_i = g_i(v_1, \dots, v_{n-1})$  denote the expected value of the data at node  $i$  and  $F(i)$  the set of its parents. Then  $V_i = v_i + \sum_{j \in F(i)} p_j V_j$ , i.e.,  $s_i$  gets information from its parents only if they survive with the given probabilities. The expected benefit to sensor  $s_i$  is given by  $V_i R_i$ , i.e.,  $i$ 's benefits depend on the survival probability of players from  $i$  onwards. Hence the payoff to  $s_i$  is  $\Pi_i = R_i V_i - c_{ij}$ . For example, the payoff to sensor  $s_5$  in the figure is  $\Pi_5 = R_5(v_5 + p_1 v_1 + p_2 v_2) - c_{56}$ .

**Definition 1:** A strategy  $l_i$  is said to be a **best response** of player  $i$  to  $l_{-i}$  if

$$0 \leq \Pi_i(l_i, l_{-i}) \geq \Pi_i(l'_i, l_{-i}) \text{ for all } l'_i \in \mathcal{L}_i.$$

Let  $BR_i(l_{-i})$  denote the set of player  $i$ 's best response to  $l_{-i}$ . A strategy profile  $l = (l_1, \dots, l_n)$  is said to be an **optimal RQR tree**  $\mathcal{T}$  if  $l_i \in BR_i(l_{-i})$  for each  $i$ , i.e., sensors are playing a Nash equilibrium. In other words, the payoff to a node on the optimal tree is the highest possible, given optimal behavior by all other nodes. A node may get higher payoffs by selecting a different neighbor on another tree, however it can only do so at the cost of suboptimal behavior by (i.e reduced payoffs to) some other node(s). Also, although each sensor can form only one link, multiple equilibrium trees can exist.

Note that the process of choosing the optimal strategy requires each node to determine the optimal tree (in the remaining graph) formed by each of its possible successors on receiving its data. The node then selects as next neighbor the node, the optimal tree through which it gets the highest payoff. Since all nodes in the graph have to perform these calculations, finding the optimal RQR tree is computationally intensive as will be shown formally in the next section. Further, given the additive nature of data aggregation, note that many of the results that hold for multiple sources are also true when considering a single source, routing to the sink. Hence we present our results mainly in terms of single source-sink paths and when necessary the result is stated in terms of trees.

### III. RESULTS

This section contains results on two aspects of the RQR problem. We first analyze the complexity of computing the optimally reliable (or equilibrium) data aggregation tree in a given sensor network. This is followed by some analytical results that establish congruence between the equilibrium RQR path and other well known path metrics such as the most reliable path, energy conserving paths etc.

#### A. Complexity Results

We begin with the following general result.

**Theorem 1:** Given an arbitrary sensor network  $G$  with sensor success probabilities  $P$ , communication costs  $C$ , and data of value  $v_i \geq 0$  to be routed from each sensor  $s_i$  to the sink  $s_q$ , computing the optimally reliable data aggregation tree  $\mathcal{T}$  (the RQR tree) is NP-Hard.

*Proof:* Given any solution  $\mathcal{T}'$  to the RQR problem, verifying the optimality of the successor for each node in  $\mathcal{T}'$  requires exhaustively checking payoffs via all possible trees to  $s_q$ . Thus RQR does not belong to NP. That the RQR problem is NP-Hard follows by reduction, using the following lemma which considers the special case of finding an optimal path, given a single source. (Note that this is equivalent to finding routing trees without data-aggregation.) ■

**Lemma 1:** Let  $\mathcal{P}$  be the optimal RQR path for routing data of value  $v_r$  from a single reporting sensor  $s_r$  to the sink node  $s_q$  in a sensor network  $G$  where  $v_i = 0 \quad \forall i \neq r$ . Computing  $\mathcal{P}$  is NP-Hard.

*Proof:* Reduction from Hamiltonian Path. See [6] for details. ■

Note that the RQR path and tree problems remain *NP-Hard* for the special case when nodes have equal success probabilities. The case when all edges have the same cost is much simpler, however, as will be shown below.

### B. Analytical Results

Given the complexity of finding the equilibrium RQR path, we next identify conditions under which this path coincides with other commonly used routing paths. In particular, we look at the most reliable path [MRP] which can be computed using well known techniques such as Dijkstra's shortest path. We also look at cheapest neighbor paths [CNP], obtained when nodes with limited network state or diffusion gradient/route quality information, select next-neighbors using only localized criteria such as communication costs.

Let  $G$  be an arbitrary sensor network with a single source node having data of value  $v_r$ . Then the following results hold. Note that the results describe only sufficient conditions for congruence with the optimal path. Also for brevity, most results are stated without proofs details of which can be found in [6].

*Observation 1: Given  $p_i \in (0, 1]$  and  $c_{ij} = c$  for all  $i, j$ , then the most reliable path always coincides with the equilibrium path. For uniform  $p_i$ , the equilibrium path is also the path with least overall cost.*

Before proceeding further, we now introduce some notation. For any node  $s_i$ , let  $c_i = \{c_{ij}\}$ ,  $c_i^{\max} = \max\{c_{ij}\}$  and  $c_i^{\min} = \min\{c_{ij}\}$ . Also  $c^{\max} = \max\{c_{ij}^{\max}\}$  and  $c^{\min} = \min\{c_{ij}^{\min}\}$ . We use  $\mathcal{P}_i^l$  to denote a path of length  $l$  from  $s_i$  to  $s_q$ .

*Proposition 1: Given  $G$  and  $P(s_i) = p \in (0, 1]$ , for all  $i$ , the most reliable path from  $s_r$  to  $s_q$  will also be the optimal path if*

$$c_i^{\max} - c_i^{\min} < v_r p^m (1 - p)$$

for all  $s_i$  on the most reliable path  $\mathcal{P}_r^m$ .

Note that the above result identifies sufficient constraints on costs for the most reliable path to also be optimal. The result shows that while the MRP can be costlier than other paths, to be optimal it cannot be 'too' much more expensive. From the above result, it also follows that when  $c^{\max} - c^{\min} < p^m (1 - p)$  the MRP coincides with the optimal, thereby providing a global bound on costs.

We define the cheapest neighbor path [CNP] from  $s_r$  to  $s_q$  as the simple path obtained by each node choosing its successor via its cheapest link (assuming such a path exists). In a sense, this path reflects the route obtained when each node has only limited network state information (about neighbor costs and probabilities) and in the absence of gradient information or route quality feedback, should merely minimize its local communication costs. The following proposition identifies when CNP will coincide with optimal path.

*Proposition 2: Given  $G$  and  $P(s_i) = p \in (0, 1]$ , for all  $i$ , the optimal path is at least as reliable as the cheapest neighbor path. Furthermore, the CNP will be optimally reliable if*

$$\min\{c_k \setminus c_k^{\min}\} - c_k^{\min} > v_r p^l (1 - p^{t-l})$$

where  $l$  is the length of the shortest path from  $s_r$  to  $s_q$  and  $t$  is the length of the CNP.

The above proposition illustrates that the CNP does not have to be the most reliable in order to be optimal, it only needs to be sufficiently close. For networks in which some paths (edges) are overwhelmingly cheap compared to others, routing along CNPs may be reasonable. However, in networks where communication costs to neighbors are similar, routing based on local cost gradients is likely to be less reliable.

## IV. QUALITY OF ROUTING

We divide this section into two subsections. In the first of these we present our route evaluation metric and some theoretical results for it. The second half provides experimental results about the quality of routes obtained different routing algorithms based on our metric. Throughout this section, we assume that there is a single source and destination pair. Thus results are presented in terms of paths instead of trees.

### A. Evaluation Metric

In an ideal sensor-centric network, optimal RQR paths are computed by individually rational sensors who maximize their own payoffs. On the other hand traditional routing algorithms optimize using a single (end-to-end) distinguishing attribute such as total cost or overall latency<sup>5</sup>. From a sensor-centric perspective these approaches are inadequate and sub-optimal since they use a single network wide criterion. How then do we compare different suboptimal paths? For example, one path may yield high payoffs for sensor  $i$  with low payoffs for sensor  $j$ , while the exact opposite situation may prevail on another path. Clearly in a framework where rational, independent sensors maximize their own payoff subject to the overall network objective, we need a new metric for evaluating the quality of different paths from an individual sensor's point of view. We introduce a metric called path weakness which captures the suboptimality of a node on the given path, i.e., how much a node would have gained by deviating from the current path to an optimal one. We believe this provides a new sensor-centric paradigm for evaluating the quality of routing in sensor networks.

We formally define our Quality of Routing metric as follows: Let  $\mathcal{P}$  be any given path from the source sensor  $s_r$  to the sink node  $s_q$ . Consider any node  $s_i$  on  $\mathcal{P}$  with ancestors  $\{s_r, \dots, s_{i-1}\}$ . Let  $\hat{\mathcal{P}}_{iq}$  be the optimal RQR path for routing information of value  $V_i = v_r \prod_{t=r}^i p_t$  (i.e., the expected value) to  $s_q$  from  $s_i$  in the subgraph  $G \setminus \{s_r, \dots, s_{i-1}\}$ , assuming such a path exists. Thus  $\hat{\mathcal{P}}_{iq}$  represents the best that node  $s_i$  can do, given the links already established by nodes  $s_r, \dots, s_{i-1}$  **and** assuming optimal behavior from nodes  $s_i$  onward, downstream. Define  $\Delta_i(\mathcal{P}) = \Pi_i(\hat{\mathcal{P}}_{iq}) - \Pi_i(\mathcal{P})$  as the payoff deviation for  $s_i$  under the given strategy profile (path)

<sup>5</sup>See [9] however, for an elegant model in which the authors develop data-centric routing algorithms for sensor networks that take both energy constraints and Quality of Service considerations into account. However the model contrasts from ours in not being sensor-centric

$\mathcal{P}$ . A negative deviation represents the fact that  $s_i$  is benefiting more from this path (perhaps at the expense of some other sensor). Conversely, a positive deviation indicates  $s_i$  could have done better. We set  $\Delta_i(\mathcal{P}) = v_r$  whenever  $\Pi_i(\mathcal{P})$  is negative. This positive deviation from the optimal payoff is intended to represent the fact that  $s_i$  is participating in a path which is giving it negative payoffs i.e, the communication cost on the edge out of  $s_i$  in  $\mathcal{P}$  outweighs the benefits to  $s_i$  of participating in this route. Also note that it is possible that no optimal path from  $s_i$  exists, even if its payoff on  $\mathcal{P}$  is positive. For example, all of  $s_i$ 's neighbors might have very high communication costs and cannot participate in any optimal path, making  $s_i$  in a sense isolated. In such cases, we set  $\Delta_i(\mathcal{P}) = -\Pi_i(\mathcal{P})$ .

$\bar{\Delta}(\mathcal{P}) = \max_i \Delta_i(\mathcal{P})$  represents the payoff deviation at the node which is 'worst-off' in  $\mathcal{P}$ . What can be said about this parameter for optimal and sub-optimal paths?

*Observation 2:*  $0 < \bar{\Delta}(\mathcal{P}') \leq v_r$  for all non-optimal paths  $\mathcal{P}'$ .

However observe that  $\Delta_i(\mathcal{P}')$ —the weakness of individual nodes on sub-optimal paths can take both positive and negative values. On the other hand,  $\bar{\Delta}(\mathcal{P}) = 0$  if and only if  $\mathcal{P}$  is the Nash equilibrium path of the game. Thus from a global point of view,  $\bar{\Delta}(\mathcal{P})$  identifies the maximum degree to which a node on the path can gain by deviating. This allows us to rank the 'vulnerability' of different paths, which embodies the idea that a path is only as good as its weakest node. We label this QoR measure **path weakness**.

Note that the weakness metric can be similarly defined for data-aggregation trees. Given a sensor on any tree  $\mathcal{T}$ , its weakness can be calculated as its payoff deviation from the optimal tree that would have been obtained, given the expected value at that sensor along with the distribution of values in the remaining nodes in the graph. As mentioned before, we focus on single-source single-destination paths in the rest of this paper.

We now compute bounds for finding paths with low path weakness. We will show that there exist networks not containing paths of bounded weakness. Our proof relies on constructing a specific example of a network whose best suboptimal paths satisfy certain weakness characteristics. This network is constructed below.

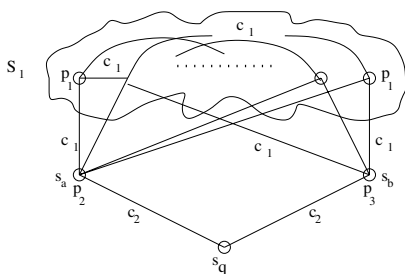


Fig. 2. Network for illustrating path weakness.

Consider an arbitrary sensor network  $G = (S, E)$  as shown in Fig. 2 with the following parameters: The vertex set  $S$  is the union of vertex set  $S_1$  with nodes  $s_a, s_b$  and  $s_q$ .  $G' = (S_1, E_1)$  is an arbitrary network, where  $|S_1| = \{s_r = s_1, \dots, s_n\} = n$ . The edge set  $E$  for  $S$  is the union of disjoint edge sets  $E_1, E_2$  and  $E_3$ , where  $E_2 = \{(s_a, s_i)\} \cup \{(s_b, s_i)\}, \forall s_i \in S_1$ , and  $E_3 = (s_a, s_q) \cup (s_b, s_q)$ . There are two types of edge costs in  $C$ -edges in  $E_3$  cost  $c_2$  with all other edges costing  $c_1$ . The node success probabilities are  $P(s_i) = p_1, \forall s_i \in S_1$ ,  $P(s_a) = p_2$  and  $P(s_b) = p_3$ .  $v_r$  is the value of information to be routed from  $s_r$  to  $s_q$ . These parameters are related to each other as follows:

$$p_3 < p_1^{n-2} \quad (1)$$

$$p_3 < p_2 \quad (2)$$

$$p_1 p_2 (1 - p_1 p_3) v_r < c_2 - c_1 \quad (3)$$

$$c_1 < c_2 < p_1^n p_2 p_3 v_r \quad (4)$$

We now look at the strategy choices for nodes in  $G$  on any path from  $s_r$  to  $s_q$ , when receiving  $v_r$ . Condition (4) ensures that all edges in the network are feasible since all payoffs are greater than zero. Also,  $s_q$  is reachable only through  $s_a$  and  $s_b$  and all edges from any node in  $S_1$  have identical costs. Thus if  $s_a$  ( $s_b$ ) is the parent of any node in  $S_1$ , this node will immediately prefer to link to  $s_b$  ( $s_a$ ) to maximize its payoff. Coupled with condition (3), this implies that if node  $s_a$  is visited before  $s_b$  in any path,  $s_a$  prefers to link to any available node in  $S_1$  instead of linking to  $s_q$ , regardless of the number of nodes visited in  $S_1$  prior to  $s_a$ . A similar situation holds true for  $s_b$  if it is visited before  $s_a$ .

Now consider paths  $\mathcal{P}_{ik} = (s_1, \dots, s_k, s_i, s_q)$ , where  $i = a, b$  is the penultimate node for  $k = 1, 2, \dots, n$  and similarly  $\mathcal{P}_k = (s_1, \dots, s_k, s_a, s_{k+1}, s_b, s_q), k = 1, \dots, n-1$ , assuming they exist. The observations above can be used to calculate the path weakness of  $\mathcal{P}_{ak}$  as follows. First,

$$\Delta_a(\mathcal{P}_{ak}) = \begin{cases} v_r p_1^k p_2 (p_1 p_3 - 1) + (c_2 - c_1), & 1 \leq k \leq n-1, \\ 0 & k = n \end{cases} \quad (5)$$

Also, for each node  $s_j, 1 \leq j \leq k$ ,

$$\Delta_j(\mathcal{P}_{ak}) = \begin{cases} v_r p_1^k p_2 (p_1^{n-k} - 1), & 1 \leq k \leq n-1, \\ & \text{if } \mathcal{P}_{an} \text{ exists} \\ v_r p_1^j p_2 (p_1 p_3 - p_1^{k-j}), & 1 \leq k \leq n-1, \\ & \text{otherwise} \\ 0, & k = n \end{cases} \quad (6)$$

To understand (5–6), first note that  $\mathcal{P}_{ak}$  cannot be the equilibrium RQR path whenever  $k < n$ . The optimal choice for  $s_a$  is always to link to any available node in  $S_1$ . Condition (1) implies that nodes in  $S_1$  would prefer to link to nodes in  $S_1$  and  $s_a$  and avoid visiting  $s_b$  en route to  $s_q$ , if possible.

Hence, the optimal payoff for  $s_j$  is via  $\mathcal{P}_{an}$  if it exists, and via the path  $(s_1, \dots, s_j, s_a, s_{j+1}, s_b, s_q)$ , otherwise.

It can be seen that  $\Delta_j(\mathcal{P}_{ak}) \leq 0$ , for all  $j$  and  $k$ . Thus  $\overline{\Delta}(\mathcal{P}_{ak})$ , the path weakness of  $\mathcal{P}_{ak}$ , is given by  $\Delta_a(\mathcal{P}_{ak})$ . Similarly,  $\overline{\Delta}(\mathcal{P}_{bk})$  can be obtained by interchanging  $p_2$  and  $p_3$  in (5).

Now consider paths of type  $\mathcal{P}_k$ ,  $1 \leq k \leq n-1$ .  $\Delta_a(\mathcal{P}_k) = \Delta_{k+1}(\mathcal{P}_k) = \Delta_b(\mathcal{P}_k) = 0$ , since these three nodes are choosing their neighbors optimally. Therefore the path weakness of  $\mathcal{P}_k$  is given by

$$\overline{\Delta}(\mathcal{P}_k) = \Delta_1(\mathcal{P}_k) = \begin{cases} v_r p_1^{k+1} p_2 (p_1^{n-k-1} - p_3), & \text{if } \mathcal{P}_{an} \text{ exists} \\ v_r p_1^2 p_2 p_3 (1 - p_1^{k-2}), & \text{otherwise} \end{cases} \quad (7)$$

Similarly, it can be shown that all paths in which  $s_b$  is visited before  $s_a$  or in which multiple nodes in  $S_1$  are visited in between  $s_a$  and  $s_b$ , are weaker than the above paths.

The following lemma can be used to compute a lower bound on the path weakness of suboptimal paths.

*Lemma 2: For any  $\epsilon \in (0, \frac{v_r}{3}]$  in the network  $G$ , there exists a path  $\mathcal{Q}$  and probabilities  $p_1, p_2, p_3$ , such that either  $\mathcal{Q}$  is the optimal RQR path or  $0 < \frac{v_r}{3} - \overline{\Delta}(\mathcal{Q}) < \epsilon$  and there is no other suboptimal path weaker than  $\mathcal{Q}$ .*

*Proof:* Consider all paths  $\mathcal{P} \setminus \mathcal{P}_{an}$  in  $G$ .

$$\min_{\mathcal{P} \setminus \mathcal{P}_{an}} \{\overline{\Delta}(\mathcal{P})\} = \min_k \{\mathcal{P}_{ak}, \mathcal{P}_{bk}, \mathcal{P}_k\}$$

where  $\mathcal{P}_{ak}$ ,  $\mathcal{P}_{bk}$  and  $\mathcal{P}_k$  are as defined before.

Using (2),  $\overline{\Delta}(\mathcal{P}_{bk}) > \overline{\Delta}(\mathcal{P}_{ak})$ , and hence  $\mathcal{P}_{bk}$  is always weaker than  $\mathcal{P}_{ak}$ . Additionally from (5), and (7), it can be seen that

$$\min_{\mathcal{P} \setminus \mathcal{P}_{an}} \{\overline{\Delta}(\mathcal{P})\} = \min \{\mathcal{P}_{a1}, \mathcal{P}_1\} \quad (8)$$

To obtain the result in the lemma, we set  $\mathcal{Q}$  to be the path  $\mathcal{P}_1$  and solve to obtain the corresponding  $p$  values as below.

$$p_1^{n-2} > p_3 > \max \left( \frac{1}{p_1(1+p_1^{n-2})}, \frac{1+p_1^{n-1}}{p_1(2+p_1^{n-2})} \right) \quad (9)$$

The first term in the maximum is obtained using conditions (3) and (4) simultaneously for the network  $G$ . Solving for situations when  $\mathcal{P}_{a1}$  exceeds  $\mathcal{P}_1$  and then using (3) gives the second term. Thus a network  $G$  with the above probability values will have the property that

$$\min_{\mathcal{P} \setminus \mathcal{P}_{an}} \{\overline{\Delta}(\mathcal{P})\} = \overline{\Delta}(\mathcal{P}_1) = \begin{cases} v_r p_1^2 p_2 (p_1^{n-2} - p_3), & \text{if } \mathcal{P}_{an} \text{ exists} \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Identifying the *upper* bound of  $\overline{\Delta}(\mathcal{P}_1)$ , subject to the constraints in (9) yields the desired bounded weakness result. ■

Since  $G'$  is an arbitrary subgraph of  $G$ , the above lemma implies the existence of infinitely many graphs without any suboptimal paths of weakness bounded by  $(\frac{v_r}{3} - \epsilon)$ . Stated another way, path weakness better than  $\frac{v_r}{3}$  is difficult to achieve, as shown in the next result.

*Theorem 2: There exists no polynomial time algorithm to compute approximately optimal RQR paths of weakness less than  $(\frac{v_r}{3} - \epsilon)$  unless  $P = NP$ .*

The proof follows from Lemma 2. Details can be found in [6].

## B. Experimental Results

In this section, we simulate the performance of different routing algorithms to answer the following question: What are the quality of paths compared to that of the optimal RQR path? This allows us to identify the different ranges of node reliabilities and edge costs in which a particular algorithm performs better than the others.

The setup for our experiments is as follows: In every iteration a random graph with 20 nodes and edge density of 30% is generated. The source and destination pair are randomly chosen and the value of data at the source node is normalized to one. For each run, we choose a node survival probability, which is identical for all nodes. Communication costs over each edge are drawn randomly from a given parameter range in every iteration. For each set of node success probabilities and edge costs, we have presented results for 15 different source and destination pairs (we have verified that this is a representative sample). In each simulation run, for a particular source and destination pair, routing paths are generated by several algorithms and the corresponding path weakness (QoR) is calculated. The data have been used to construct graphs which are presented at the end of the paper. We have used the following algorithms:

1. *Most Reliable Path (MRP)*: This produces the most reliable path from source to the sink. Since, in our setup, each node has the same success probability the MRP is always the shortest path as evaluated by Dijkstra's standard shortest path algorithm.

2. *Overall Cheapest Path (MCP)*: This algorithm is also Dijkstra's shortest path algorithm, with the weight of each edge being the communication cost.

3. *Cheapest Next Node Path (CNP)*: This provides a path where each node chooses its cheapest available edge leading to the sink node.

4. *Team RQR Path (TRQR)*: This path is obtained by considering a 'team' version of the RQR game in which all nodes on the path share the payoff of the worst-off node on it.

Rather than selecting a neighbor to maximize their individual payoffs as in the original game, nodes in the team-RQR model compromise by maximizing their least possible payoff. Formally, let  $\overline{P}_c$  represent the most reliable path from  $s_r$  to  $s_q$  that does not traverse any link exceeding cost  $c$ . Then  $\overline{P}$ , the equilibrium path of the team-RQR game is given by

$$\overline{P} = \arg \max_{c_i \in C} \left\{ v_r R(\overline{P}_{c_i}) - c_i \right\} \quad (11)$$

for each *distinct edge cost*  $c_i$  in  $C$ . An intuitive technique for computing the optimal team-RQR path is to repeatedly determine the most reliable path in the graph that is obtained by successively removing edges of decreasing distinct cost. In the worst case  $m$  most reliable path calculations are made, where  $m$  is the number of distinct edge costs in the network.

**5. Genetic Algorithm Path (GA):** Here, we use a genetic algorithm for solving the optimal RQR problem based on the GA for the bicriteria shortest path problem provided in [4]. A path has been encoded according to the priority-based method. In this procedure, a set of  $n$  random numbers ( $n$  being the total number of sensor nodes) is generated so that the  $i$ -th random number is the priority of the  $i$ -th node. A path is sequentially constructed led by the highest priority feasible nodes i.e., nodes which do not lead to a dead end or a cycle. The genetic operators used here are position-based crossover and swap mutation. A next generation is chosen by tournament method. We stop if the difference between the fitness values of the best paths of two adjacent generations is equal to zero.

The first three algorithms are standard routing algorithms. The fourth algorithm is our heuristic derived from a game theoretic point of view. Genetic algorithm is a standard technique applied to problems which are NP-complete or NP-hard. We have used it here to check if there is any range of node success probabilities and costs where it does well.

### Interpretation of Results:

Our simulation results are illustrated in Fig. 3 and Fig. 4. In the first five graphs, nodes are assigned very high success probabilities. Edge costs are low and chosen from a distribution such that every path is feasible (all node payoffs are positive). In case I and II, we keep the node success probability fixed at 0.99 and vary the maximum edge cost from 0-0.05 and 0-0.01 respectively.

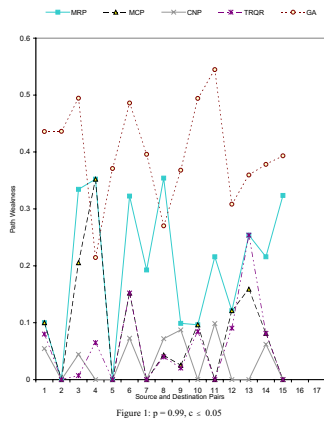
In case I, the path weakness ranges from 0 to 0.6. MCP and TRQR have average weaknesses 0.08 and 0.05 respectively in spite of their of their considerable deviations. Since the cost range and hence the cost differences among various edges are not significantly large, all the above three algorithms that try to reduce the overall cost in different ways behave reasonably well. However, the range of path weakness of MRP (0-0.4) suggests that the cost range is so high that a path which relies solely on maximizing reliability (MRP) cannot perform well. The maximum edge cost is then reduced to 0.01 in case II. Consequently, the overall range of path weakness reduces to

0-0.14. Significant improvement takes place in the behaviour of MRP and TRQR as they coincide with the optimal path for more than 90% of the source and destination pairs. The fact that MRP always coincides with the optimal path indicates that the very high node success probability and very small cost range together have reduced the length of the optimal path. The diminished variation within different edge costs allows MCP and CNP to perform well. Since the behaviour pattern and the range of path weakness of CNP do not vary significantly from case I to case II, we can conclude that performance of CNP is invariant over a large cost range when reliability is kept very high.

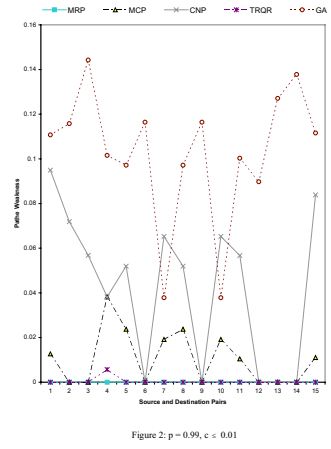
For Cases III, IV and V, we make the maximum edge cost a decreasing function of the node success probability. Then, we slowly increase node success probability to observe the impact. In case III, where the node success probability is 0.992 and the cost range is 0-0.227, the range of path weakness is quite high (0-0.35). When we raise the value of the success probability, the optimal paths can have longer lengths without sacrificing too much reliability. Therefore CNP, which tends to have a longer length, has lower path weakness now (average weakness being 0.035 approximately). The TRQR heuristic, which tradesoff both the overall path reliability and the overall cost performs as well as CNP producing an average path weakness of 0.32. As expected MCP's weakness does not differ too much from that of CNP or TRQR. The above mentioned feature of the optimal path can also explain MRP's unstable pattern and the high range of path weakness in spite of very high node success probabilities. In case IV, the success probability is increased to 0.998 and the cost range is reduced to 0-0.058. This accounts not only for the relatively small range of path weakness (0-0.1) but also for the good performance of MCP, CNP and TRQR. The congruence of TRQR and MCP is well explained by the significantly large difference between the success probability and the maximum edge cost. In case VI, we explore the consequences of restricting the likely optimal path length using one low node success probability (0.5) and maximum edge cost  $(1/2)^4$ . MRP, the shortest path, always coincides with the optimal path even though the success probability is quite low. So do TRQR and MCP. However, since the CNP usually has longer path lengths, its QoR is quite weak, in most cases.

When we compare the first 5 graphs, we observe that the increment in the node success probabilities together with the decrement in the maximum edge costs gradually leads to improvements in the behaviour of all five algorithms. In general, MRP will be a good heuristic for obtaining good QoR paths only when path reliabilities are low. The behaviours of TRQR and MCP are quite stable (with a little variation in the weakness ranges) in all the ranges of our experiment and on average, provide better QoR. CNP provides good QoR when the success probability increases and the maximum edge cost decreases accordingly.

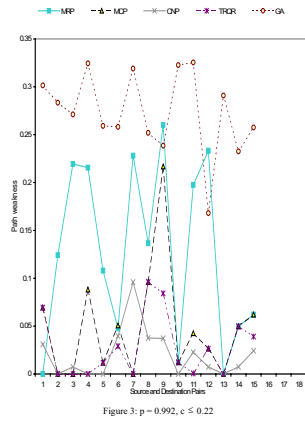




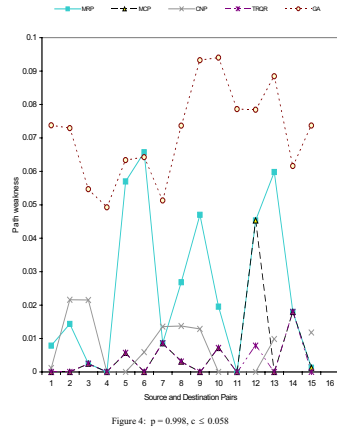
(a) Case I



(b) Case II

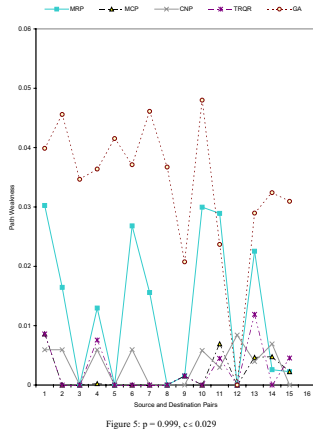


(c) Case III

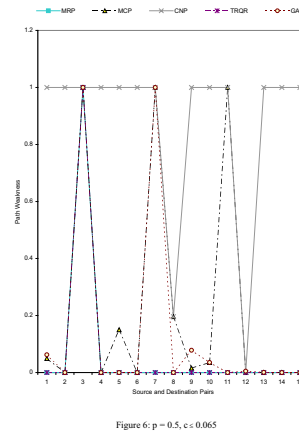


(d) Case IV

Fig. 3. Simulation results



(a) Case V



(b) Case VI

Fig. 4. Simulation results

## V. CONCLUSION

In this paper we formulate a sensor-centric model of intelligent sensors using game theory. The problem of routing data in such a network is studied under the assumption that sensors are rational and act to maximize their own payoffs in the routing game. Further, nodes in our model are susceptible to failure and each node has to incur costs in routing data. To evaluate the contribution of individual nodes in the routing tree, we develop a metric called path weakness. This individual sensor-oriented evaluation criteria provides a new paradigm for examining paths which we call Quality of Routing. While the optimal routing problem turns out to be computationally hard, our experimental results show that standard path routing mechanisms like MRP and MCP find reasonably good paths. Our game-theoretically oriented algorithm - Team RQR also performs well. For future work we plan to consider extensions using distributed games and dynamic data routing.

## ACKNOWLEDGMENT

This work was supported by DARPA SensIT administered under AFRL grant # F30602-02-1-0198.

## REFERENCES

- [1] I.F. Akyldiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Wireless Sensor Networks: A Survey," *Computer Networks*, Vol. 38, No. 4, pp. 393-422, March 2002.
- [2] A. Cerpa and D. Estrin, "Ascent: Adaptive Self-Configuring sensor Network Topologies," in *Proc. INFOCOM 2002*, New York, June 2002.
- [3] D. Fudenberg and J. Tirole, *Game Theory*, MIT Press, 1991.
- [4] M. Gen and R. Cheng, "Genetic Algorithms and Engineering Optimization", Wiley-Interscience Publication, New York, December 1999.
- [5] C. Intanagonwivat, R. Govindan and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks" (2000), in *Proc. Sixth Annual International Conference on Mobile Computing and Networks (MobiCom 2000)*, August 2000, Boston, Massachusetts.
- [6] R. Kannan, S. Sarangi, "Reliable Query Reporting in Sensor Networks," LSU Computer Science Tech Report CS-TR-010, March 2002.
- [7] B. Krishnamachari, D. Estrin, S. Wicker, "Modeling Data-Centric Routing in Wireless Sensor Networks," in *Proc. INFOCOM 2002*, New York, June 2002.
- [8] V. Rodoplu and T.H. Meng, "Minimum Energy Mobile Wireless Networks," *IEEE J. Sel. Areas in Comm. (JSAC)*, Vol. 17, No.8, pp. 1333-1344, August 1999.
- [9] K. Sohrabi, J. Gao, V. Ailawadhi and G. Pottie, "Protocols for Self-Organization of a Wireless Sensor Network," *IEEE Personal Communications*, pp. 16-27, October 2000.
- [10] A. Wood and J. Stankovic, "Denial of Service in Sensor Networks," *IEEE Computer*, pp. 54-62, October 2002.