A Simple Model for Reliable Query Reporting in Sensor Networks *

Rajgopal Kannan
Dept. of Computer Science
Louisiana State University
Baton Rouge LA 70803
rkannan@csc.lsu.edu

Sudipta Sarangi
Dept of Economics
Louisiana State University
Baton Rouge LA 70803
sarangi@lsu.edu

S. S. Iyengar
Dept of Computer Science
Louisiana State University
Baton Rouge LA 70803
iyengar@csc.lsu.edu

Abstract – *We consider the problem of inducing the formation of reliable energy-constrained communication paths in distributed sensor networks. We label this problem of reliable information routing from reporting sensors to query nodes as Reliable Query Reporting (RQR). and analyze it using game-theoretic techniques to model the behavior of individual sensor nodes. Sensors behave strategically wherein they take into account individual costs and benefits as well as the actions of other sensors in the network. The RQR problem is shown to be NP-Hard. The optimal RQR path as defined forms an upper bound on reliable paths obtained using existing ad-hoc routing rules which are easy to implement but do not provide analytical performance bounds. We then present analytical results to identify conditions under which the optimally reliable path is congruent to well known paths such as the most reliable or cheapest path.*

Keywords: Network Reliability, Game Theory, Sensor Networks

1 Introduction

This paper introduces a new game-theoretic model that brings together elements of two emerging areas of research: strategic reliability and energy-constrained distributed sensor networks. A distributed sensor network is a web of sensors used collectively to perform a wide array of tasks ranging from military applications such as target detection, location and tracking to environmental monitoring and surveillance [1]. Each node in the network may consist of one or more sensors, ranging from motion detectors to chemical detectors, along with communication devices with portable power sources and related localization equipment, such as GPS (Global Positioning System) units. The key feature of such networks is that the nodes are unattended and untethered (independent). Hence the network must be self-configuring i.e., the nodes must make information routing/connectivity decisions in a decentralized manner. Moreover, communication must be energy efficient since battery power cannot easily be replenished.

In general, such a sensor network modeled as graph, consists of distributed query nodes which disseminate requests for information through the network. Sensor nodes that detect specific local events corresponding to these queries report information back to the appropriate query nodes, through a set of intermediate sensors. This information must be routed carefully since sensors operate unreliably and can fail. Moreover conveying information is a costly process. Individual sensors are modeled as players who make decisions about choosing intermediate nodes by weighing their own costs and benefits. We characterize the induced topology for optimal paths in such networks. Current models for communication in these networks use protocols like diffusion routing [5], which uses local 'gradients' to identify paths for sending information. However, these protocols do not optimize network wide reliability in conjunction with minimizing communication costs. Our contribution in this paper is to propose a model that explicitly optimizes over both dimensions. The techniques introduced here can also be adapted to model the trade-offs to intelligent network nodes under other optimization criteria such as throughput or delay.

Current game theoretic formulations of reliability consider situations where players derive benefits from having more information. Each player can access more information through a process of costly link formation with other players. This problem was first studied by Bala and Goyal [2] Haller and Sarangi [3] generalize the model by allowing for different link failure probabilities and introduce several variations of the initial

* This work was supported in part by DARPA-SenseIT and AFRL under grant number F30602-01-1-0551.

model to address potential shortcomings.

There is a considerable body of literature on data fusion, collaborative signal processing and communication in distributed sensor networks, for example, see [4]. However, one issue that has been largely ignored is the role of reliable energy-constrained communication between sensors. Existing communication protocols for such networks are based on localized rules for selecting next-neighbors. This is not always conducive to the formation of reliable source to destination paths. Furthermore, the lack of an existing theoretical framework in which to analyze distributed sensor networks often forces researchers to resort to simulations. Theoretical results when they exist are very specific to the model in question. This makes it quite hard to compare models and derive general conclusions.

This paper poses the following fundamental question: Can game theory be used to model the formation of reliable energy constrained *paths*[1] in sensor networks?

We believe that game theory can provide the appropriate theoretical framework to analyze distributed sensor networks. By the very nature of their deployment these networks cannot be controlled at every step by the network designer. This scenario of distributed decision making fits very well with the spirit of game theory. By designing the payoff function suitably the network designer can achieve different degrees of collaborative tasking among the sensors. More importantly, while the informational requirements in a game theoretic model will be higher than neighborhood informational requirements of the current models, we believe that this framework will be useful to obtain general conclusions about the operation of the sensor network.

This paper introduces game-theoretic modeling techniques for a distributed sensor network. We present a model of strategic reliability in a distributed sensor network. Unlike the earlier game-theoretic papers, we model node failure. All sensors on the optimal path have identical benefits while facing individual costs. Complexity of the problem is analyzed. Results identifying conditions under which the optimal path will coincide with well known computable paths like the most reliable path are derived.

This paper is organized as follows. Section 2 sets up the basic model. Results are presented in Section 3. The final section has concluding remarks about future research directions.

[1] As opposed to equilibria that form *spanning subgraphs* in other game theoretic models of reliability.

2 The Model

Let $S = \{s_1, \ldots, s_n\}$ denote the set of sensors, with generic members i and j. For ordered pairs $(i, j) \in S \times S$, the shorthand notation ij is used. We assume throughout that $s_n \geq 3$. Without loss of generality the source node $s_r = s_1$ has information of value V_r which it wishes to send to the destination node $s_q = s_n$. V_r represents an abstract quantification of the of the value of event information at sensor s_r. Information is routed through an optimally chosen set $S' \subseteq S$ of intermediate nodes by forming links. Link formation occurs by a process of simultaneous reasoning at each node leading to a path from s_r to s_q. This link formation is costly with each node incurring a cost for the link it establishes. We denote the cost of link ij by $c_{ij} > 0$. This link cost can be an abstraction of packet transmission costs in terms of required transmission power or available on-field sensor battery life, depending on the type of sensor network being modeled. For simplicity, in this paper we assume that receiving information is 'free' and nodes do not need consent before transmitting information over a link. Furthermore, we assume that sensor s_i can independently fail[2] with a probability $(1 - p_i) \in (0, 1)$. Thus $G = (S, E, P, C)$ represents an instance of a distributed sensor network in which information of value V_r is to be optimally routed from sensor s_r to sensor s_q, where S is the set of sensors interconnected by edge set E, $P(s_i) = p_i$ are the success probabilities and $C(s_i, s_j) = c_{ij}$, the cost of links in E.

In this context, we define the following problem called ***Reliable Query Reporting*** (RQR): Given that information transmission in the network is costly and not fully reliable, how can we induce the formation of maximally reliable paths in G from reporting to querying nodes where every node is also maximizing its own payoffs. The solution to this problem lies in designing payoff functions such that the Nash equilibrium of this game corresponds to the optimally reliable path. Note that such game-theoretic techniques can be used to achieve other desired network objectives as well. We now describe the different components of this strategic game.

Strategies. Each node's strategy is a vector $l_i = (l_{i1}, \ldots, l_{ii-1}, l_{ii+1}, \ldots, l_{in})$ and $l_{ij} \in \{0, 1\}$ for each $j \in S \backslash \{i\}$. 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 sensor 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 $n - 1$

[2] We assume that the destination node s_q never fails.

nodes, the number of strategies available to node i is $|\mathcal{L}_i| = 2^{n-1}$. The strategy space of all agents is given by $\mathcal{L} = \mathcal{L}_1 \times \cdots \times \mathcal{L}_n$. 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. 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 $l = (l_1, \ldots, l_n)$ becomes a **simple directed path** from s_r to s_q denoted by $\mathcal{P}$. 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 since we wish to design a decentralized sensor network that can behave in a collaborative manner to achieve a joint goal while taking individual sensor operation costs into account. Since the communal goal in this instance is reliable information transmission, the benefits to a sensor 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 path $\mathcal{P}$ from s_r to s_q where l_{-i} denotes the strategies of all the other sensors. Since every sensor has an incentive to ensure information is routed to s_q, the benefit to any sensor s_i on $\mathcal{P}$ must be a function of the path reliability from s_i onwards. Since the network is unreliable, the benefit to 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:

$$\Pi_i(l) = \begin{cases} g_i(V_r) f_i(R) - c_{ij} & \text{if } s_i \in \mathcal{P} \\ 0 & \text{otherwise} \end{cases}$$

where R denotes path reliability.

Fig. 2 illustrates this idea by looking at two adjacent nodes on a path. The expected value of information at sensor j is $p_i p_j V_i$, i.e., sensor j gets the information only when sensors i and j survive with probability p_i and p_j respectively. The expected benefit to sensor i is given by $p_j V_i$, i.e., sensor i's benefits depend on the survival probability of sensor j. Hence the payoff to sensor i is $\Pi_i = p_j V_i - c_{ij}$.

The payoff function that corresponds to this idea of communal reliability and individual costs can now be written as follows:

$$\Pi_i(l_i, l_{-i}) = V_r \prod_{t=a}^{i} p_t \prod_{t=i+1}^{q} p_t - c_{ij}$$

Figure 1: Information transfer on a path.

where $g_i(V_r) = V_r \prod_{t=a}^{i} p_t$ and $f_i(R) = \prod_{t=i+1}^{q} p_t$. We also use $\Pi_i^{\mathcal{P}}$ as the payoff to node s_i in the strategy profile represented by path $\mathcal{P}$.

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

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

Let $BR_i(l_{-i})$ denote the set of sensor i's best response to l_{-i}. A strategy profile $l = (l_1, \ldots, l_n)$ is said to be an **optimal RQR path** $\mathcal{P}$ if $l_i \in BR_i(l_{-i})$ for each i, i.e., sensors are playing a Nash equilibrium. Note that although each sensor can form only one link, multiple equilibrium paths can exist. For a given node we assume that if multiple optimal paths with identical payoffs exist, the most reliable among them is chosen.

3 Results

This section contains results on two aspects of the RQR problem. We first analyze the complexity of computing the optimally reliable paths in a given sensor network along with some approximation bounds. This is followed by some analytical results that establish congruence between optimal RQR paths and other well known path metrics.

3.1 Complexity Results

Let $G = (S, E, P, C)$ represent an instance of a distributed sensor network in which information of value V_r is to be routed from sensor s_r to sensor s_q. Only those strategy profiles that define a path from s_r to s_q are of interest and must be evaluated to compute the optimally reliable path. To compute this path each node calculates a path through a sequence of descendants whose reliability (given similar decisions by descendant nodes) relative to the immediate successor's link cost, is maximum at that node.

Theorem 1 *Given an arbitrary sensor network $G = (S, E, P, C)$ with information V, computing the optimal RQR path is NP-Hard.*

Proof: Given a solution to the RQR problem, for each node on the path verifying optimality of the successor requires exhaustively checking all possible paths to s_q. Thus RQR does not belong to the class NP.

We show that the problem is NP-Hard by considering a reduction from the Hamiltonian Path problem (*Hamiltonian path* [6]). Let $G' = (V', E')$ be any graph in which a Hamiltonian Path is to be found, where $|V'| = n$. We convert G' into another graph $G = (S, E, P, C)$ on which an instance of RQR with value $V_r = 1$, must be computed as shown in Fig. 3.1.

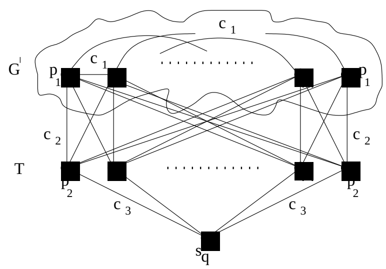

Figure 2: Reduction from Hamiltonian path.

Introduce $n+1$ new vertices to form $S = V' \bigcup T \bigcup s_q$, where $|T| = n$. E consists of the original edge set E' along with n^2 new edges from $E_2 = \{T \times V'\}$ and n new edges from $E_3 = \{T \times s_q\}$. Edges from E', E_2 and E_3 are assigned costs c_1, c_2 and c_3 respectively. All vertices $u \in V'$ and $w \in T$ are assigned success probabilities p_1 and p_2 respectively. The relationships between the probabilities and costs are as follows:

$$p_1p_2 > (\frac{3}{4})^{\frac{1}{n-1}} \quad (1)$$

$$c_1 = \frac{(p_1p_2)^n}{3} \quad (2)$$

$$c_2 = \frac{2(p_1p_2)^n}{3} \quad (3)$$

$$c_3 = (p_1p_2)^n \quad (4)$$

Let s_r and s_t be any two nodes in V'. We claim that there exists a Hamiltonian path from s_r to s_t in G' if and only if there exists an optimal RQR path of reliability $p_1^n p_2$ from s_r to s_q in G. For the only if part of the claim, assume there is a Hamiltonian Path $\mathcal{Q} = (s_r, \ldots, s_t)$ in G'. Consider the path $\mathcal{Q}$ followed by the edges (s_t, x) and (x, s_q) in G', where x is any node in T. This path has reliability $R(\mathcal{Q}) = p_1^n p_2$. The payoff for node s_t is $R(\mathcal{Q}) - c_2$ obtained by linking to node x, which is optimal since there does not exist any other unvisited node in V'. Similarly the payoff for node x is also optimal since it can only link to node s_q. Using (1)-(3) it can be seen that the payoff for the k^{th} node in $\mathcal{Q}$ is also optimal since $R(\mathcal{Q}) - c_1 > p_1^k p_2 - c_2$, which is the most reliable alternate path for this node.

For the second part of the claim, we need to show that if no Hamiltonian Path exists, there cannot be an optimal RQR path of reliability $p_1^n p_2$. For any node $s_i \in T$, it is always preferable to link to any available node in V' with cost c_2. The worst case payoff to s_i via a link of cost c_2 is $p_1^n p_2^n - c_1$, which outweighs the best possible payoff via a link of cost c_3 which is $p_1 p_2 - c_3$. Thus the optimal path must visit all nodes in V'. Similarly, any node in V' will always prefer to link to another node in V', if available. To maximize payoffs, the optimal path must have the shortest length possible. Thus it will consist of sequences of long paths in V' (the longest possible), interspersed with visits to any available node in T, such that all nodes in V' are visited with as few visits to nodes in T as possible. Since G' does not contain a Hamiltonian path there will be at least two visits to nodes in T and hence the reliability of such a path will be at least $p_1^n p_2^2$ which is less than $p_1^n p_2$. ■

The above implies that the RQR problem is still NP-Hard for uniform node success probabilities.

The optimally reliable path(s) is a Nash equilibrium, in which each node chooses its optimal neighbor in response to the links established by other nodes. Given that this problem is NP-Hard, we consider the complexity of finding 'good' sub-optimal paths. This naturally leads to consideration of the following issue: how does one evaluate the utility of *any sub-optimal* path? For example, paths in which not all nodes are playing their best response, yet are easy to compute. Given the underlying premise of decentralized decision making, any path evaluation metric must primarily account for the sub-optimality of *individual* node behavior rather than the aggregate response of nodes on the path. For instance, it is possible to find paths whose cumulative payoffs are higher than that on the optimally reliable path, with some nodes having greater payoffs at the expense of others. Thus metrics based on cumulative payoffs will fail to capture the suboptimal behavior of individual nodes. In fact, it can be shown that there exists no $(\frac{V}{3} - \epsilon)$ approximation algorithm to the RQR problem unless $P = NP$.

3.2 Analytical Results

Given the complexity of finding the equilibrium RQR path, we try to 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 Djikstra's shortest path. We also look at paths obtained when nodes select next-neighbors using a localized purely energy constrained criterion, i.e., cheapest neighbor.

Let G be an arbitrary sensor network with value V. Then the following results hold.

Observation 1 *Given $p_i \in (0,1)$ and $c_{ij} = c$ for all ij, then the most reliable path always coincides with the equilibrium path.*

Proof: Consider the most reliable path from the reporting node s_r to the querying node s_q. Clearly, the maximum payoff to s_r is obtained from this path. The payoff to any other sensor $s_i \in S$ on this path must also be maximum given uniform costs. Otherwise a more reliable path from s_r to s_q via s_i can be found. Note that when p_i is uniform, then this equilibrium path is also the cheapest path. ■

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

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

$$c^i_{\max} - c^i_{\min} < p^m(1-p)V$$

for all s_i on the most reliable path P^m_r.

Proof: Consider an arbitrary node s_i at a distance i from s_r. Let l be the length of the shortest path from s_i to s_q, where s_{i+1} is the next neighbor of s_i. For s_i, $\mathcal{P}^l_\alpha$ is optimal if

$$\begin{aligned} Vp^{i+l} - c_{ii+1} &> Vp^{i+l+\lambda} - c_{ij} \quad \lambda = 1, 2, \ldots \\ &\Rightarrow \frac{c_{ij} - c_{ii+1}}{V} < p^{i+l}(1-p^\lambda) \end{aligned}$$

where s_j is a neighbor of s_i through which there is a simple path of length $l+\lambda$. Since $m = i + l$ on $\mathcal{P}^m_r$, the reliability term above is minimized for $\lambda = 1$, whereas the cost term is maximized at $c^i_{\max} - c^i_{\min}$. ■

Note that the above result identifies sufficient constraints on costs for the most reliable path to also be optimal. It 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 now look at the situation when the probabilities of node survival are non-uniform. Let s_i and s_{i+1} be subsequent nodes on the most reliable path. Denote R_i be the reliability of the most reliable path from s_i to s_q and R'_i be the reliability along any alternative path from s_i. Let $\Delta c_i = c_{ii+1} - c_{ij}$ where s_j is any neighbor not on the optimal path and ΔR_i is defined similarly.

We define the cheapest neighbor path [CNP] from s_r to s_q as the path obtained by each node choosing its successor via its cheapest link. In a sense, this path reflects the optimal route obtained when each node merely cares about minimizing its local communication costs. The following proposition identifies when it will coincide with optimal path.

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

$$\min\{c^k \backslash c^k_{\min}\} - c^k_{\min} > Vp^l(1-p^{t-l})$$

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

Proof: Consider an arbitrary node s_k which is k hops away from s_q on the CNP. In order for this to be optimal s_k should not get higher payoff by deviating to an alternative path. We do not need to consider alternative paths that have lengths greater than k to s_q since that would decrease benefits and the CNP already has the lowest cost edges. Let m be the path length along the CNP from s_r to s_k. For alternative paths of length $i = 1, \ldots, k-1$, from s_k to s_q to be infeasible, we need

$$c_i > c_o + Vp^{m+i}(1-p^{k-i})$$

where c_o is the edge cost along the CNP and c_i, the edge cost along alternative paths. By definition for any node on the CNP $m + i \geq l$. Also at s_k we have $c_o = c^k_{\min}$, with c_i being at most $\min\{c^k \backslash c^k_{\min}\}$. Thus when $\min\{c^k \backslash c^k_{\min}\} - c^k_{\min} > Vp^l(1-p^{t-l})$ the CNP will coincide with the optimal path. ■

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 are overwhelmingly cheap compared to others, routing along CNPs may be reasonable. However, in networks where communication cost are not dissimilar, routing based on local energy (cost) gradients is likely to be less reliable.

4 Conclusion

We introduce game-theoretic techniques to model intelligent behavior in a sensor network. The formation of reliable communication paths from a source node to a querying node are analyzed. Certain extensions of the model suggest themselves immediately. The model can be easily extended to the case where the probabilities of node failure are non-uniform. The simultaneous presence of multiple reporting and querying is another obvious one. The dynamic evolution of paths in a sensor network where nodes fail over time would also be an useful extension. Other interesting versions of the problem could incorporate uncertainty and localized information. For instance, each node is perhaps aware only of the failure probabilities and costs of link formation of a neighborhood set of sensors. Decisions made under these constraints could lead to dramatically different results from the full information model analyzed here. Uncertainty in the model could be of the form where a sensor is only aware of the probability distribution from which link formation costs are drawn instead of knowing these costs precisely. We believe these extensions would be of great practical interest.

The other direction for future research would be to focus on the complexity aspect of the problem. The first task in this context would be to develop evaluation metrics for different paths that can be obtained using heuristics. In the standard version of the problem sensor nodes together maximize a joint objective function. In the game-theoretic version of the problem, every node has its own payoff. Consequently, a path metric that maximizes the payoff function of one node could very well minimize the payoff function of another node. We suggest several alternative techniques to deal with this problem. The first would be to formalize the notion of path weakness mentioned in the paper. Another approach would be to setup a group version of the problem with a common objective function. This would allow for development of easy algorithms since would be no conflict between maximizing group and individual objectives. Another heuristic-based approach would be to use genetic algorithms for identifying the optimal path. The probabilistic setup mentioned in the earlier paragraph can also be used to develop randomized algorithms.

Finally, the techniques introduced here can also be used to model certain types of behavior on the Internet. One example could be the reliable communication of information between different web servers. Alternative payoff functions based on different link-establishment criteria also need to be explored in this context. Congestion control, pricing of links and multicast issues are all possible application areas of the game-theoretic formulation introduced here.

References

[1] Brooks, R.R., C. Griffin and D.S. Friedlander, " Self-organized Distributed Sensor Network Entity Tracking," *Journal of High-Performance Computing: Special Issue on Sensor Networks*, July 2001.

[2] V. Bala and S. Goyal, "A Strategic Analysis of Network Reliability," *Review of Economic Design*, 2000.

[3] Haller, H. and S. Sarangi, "Nash Networks with Heterogeneous Agents," *Economic Theory*, under review..

[4] Journal of High-Performance Computing, Special Issue on Sensor Networking, July 2001.

[5] C. Intanagonwiwat, R. Govindan and D. Estrin, "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks," In *Proc. Sixth Annual International Conference on Mobile Computing and Networks (MobiCOM 2000)*, August 2000, Boston, Massachusetts.

[6] M. R. Garey and D. S. Johnson, "Computers and Intractability," Freeman, New York, 1979.