

AFRL-IF-RS-TR-2003-73
Final Technical Report
April 2003



**DYNAMIC PROTOCOLS FOR RELIABLE QUERY
REPORTING IN SENSOR NETWORKS:
ANALYTICAL FRAMEWORK AND PROTOCOLS**

Louisiana State University

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK**

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2003-73 has been reviewed and is approved for publication.

APPROVED:



SIAMAK TABRIZI
Project Engineer

FOR THE DIRECTOR:



WARREN H. DEBANY, Technical Advisor
Information Grid Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE APRIL 2003	3. REPORT TYPE AND DATES COVERED Final Jun 01 – Aug 02	
4. TITLE AND SUBTITLE DYNAMIC PROTOCOLS FOR RELIABLE QUERY REPORTING IN SENSOR NETWORKS: ANALYTICAL FRAMEWORK AND PROTOCOLS			5. FUNDING NUMBERS C - F30602-01-1-0551 PE - 61101E PR - M280 TA - SE WU - NS	
6. AUTHOR(S) Rajgopal Kannan, S. Sitharama Iyengar, and Sudipta Sarangi				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Louisiana State University Department of Computer Science 298 Coates Hall Baton Rouge Louisiana 70803			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/IFGC 525 Brooks Road Rome New York 13441-4505			10. SPONSORING / MONITORING AGENCY REPORT NUMBER AFRL-IF-RS-TR-2003-73	
11. SUPPLEMENTARY NOTES AFRL Project Engineer: Siamak Tabrizi/IFGC/(315) 330-4823/ Siamak.Tabrizi@rl.af.mil				
12a. DISTRIBUTION / AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words) An analytical model was formulated for reliably routing queries in sensor networks constrained to operate with unattended sensors possessing limited energy for communication under the possibility of sensor failure due to malfunction or enemy attack. By modeling the sensors as 'intelligent', game theory was used to define optimally reliable yet energy-constrained communication paths from the individual sensors point of view. Determining the optimal routing path was shown to be computationally intensive. Bounds on link costs and sensor failure probabilities under which the optimal routing path becomes congruent to practical routing paths such as the most reliable or maximally energy efficient path, were derived. The proposed game-theoretic model sets the stage for deriving practical distributed query routing algorithms that are reliable and energy-efficient from a sensor-centric point of view.				
14. SUBJECT TERMS Sensor Networks, Querying, Reliable Routing			15. NUMBER OF PAGES 27	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Table of Contents

1. Introduction	1
2. Project Information	1
2.1 Programmatic Information	1
2.2 Project Description	2
2.2.1 Research Objectives	2
2.2.2 Technical Approach	4
3. Technical Report	9
3.1 Project Progress	9
3.1.1 Progress Against Planned Objectives	9
3.1.2 Technical Accomplishments in this Project	10
3.1.3 Publications	20
3.1.4 Meetings and Presentations	21
4. Conclusion	22

List of Figures

Figure 1: Payoffs with data aggregation.	7
Figure 2: Reduction from Hamiltonian path.	14

1. Introduction

This final status report describes the outcome of the project titled `` Dynamic Protocols for Reliable Query Reporting in Sensor Networks: Analytical Framework and Protocols'' in support of AFRL and DARPA carried out by the participants from Louisiana State University during the reporting period. The purpose of the project was to investigate the problem of *reliable routing in sensor networks* operating under specific constraints. The project utilized the following participants from Louisiana State University: Dr. Rajgopal Kannan, Dr. S. Sitharama Iyengar and Dr. Sudipta Sarangi along with one graduate student, Mr. Ysaswi Rachakonda, employed as a research assistant. The organization of this report follows the guidelines as set forth in the CDRL.

2. Project Information

2.1 Programmatic Information

Administrative data relevant to this effort is summarized below.

ARPA Order Number: M280/00

Assistance Instrument Number: CFDA 12.910

Performance Period: 1 June 2001 - 31 July 2000

Project Title: Dynamic Protocols for Reliable Query Reporting in Sensor
Networks-Analytical framework and Protocols.

Performing Organization: Louisiana State University

Performing Organization Contacts:

Principal Investigator Contact:

Rajgopal Kannan

298 Coates Hall

Department of Computer Science

Baton Rouge, LA 70803

Email: rkannan@csc.lsu.edu

Phone: (225) 578 2225

Fax: (225) 578-1465

Administrative Contact:

James L. Bates, Director
Office of Sponsored Programs
Louisiana State University
330 Boyd Hall
Baton Rouge, LA 70803
Email: osp@lsu.edu
Phone: (225) 578-3386
Fax: (225) 578-5403

DARPA Program Manager:

Dr. Sri Kumar
DARPA/ITO
3701 North Fairfax Drive
Arlington, VA 22203-1714
Phone: (703) 696-0174
Fax: (703) 696-4534
Email: skumar@darpa.mil

DARPA Program: SensIT

2.2 Project Description

2.2.1 Research Objectives

2.2.1.1 Problem Description

This project was aimed at the development of a formal theory behind the analysis of the *Reliable Query Reporting* (RQR) problem in an adaptive sensor network. We model data-centric routing in sensor networks with unattended, untethered sensors operating in hazardous failure-prone environments. 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. Different sensors may have partial

information concerning a query initiated by the control node. Given a distribution of costs, sensor survival probabilities, and the quantified values of information, what is the optimal reliable yet energy-constrained route to the destination?

2.2.1.2 Research Goals

Standard models for query routing in sensor networks emphasize the constraints of communication energy efficiency and distributed decision-making. However, in many cases, sensors are deployed in hostile environments and thus the *reliable* routing of sensitive time critical information becomes vital. Therefore query routing must be considered in the context of two additional constraints - the possibility of sensor failure and the fact that each sensor must tradeoff its own resource consumption with overall routing objectives. Thus the goal of the proposed effort is to develop an analytical model of reliable data-centric query routing in sensor networks that optimize communication energy efficiency and distributed decision-making under the possibility of sensor failure. Unlike existing techniques, we use game theory to model *intelligent* sensors thereby making our approach *sensor-centric*.

The objective is to model RQR by mapping the problem onto a graph topological reliability domain. By quantifying the parameters governing individual and communal (group) sensor node behavior, our goal was to derive game/decision theoretic techniques that control the actions of intelligent sensor nodes in terms of routing link formation, leading to the dynamic formation of RQR topologies based on local topological decisions.

2.2.1.3 Expected Impact

We have developed a new analytical framework within which reliable routing in sensor networks from reporting sensors to querying nodes can be examined in a quantitative manner. In particular, different standard routing protocols can be compared and quantitatively evaluated with respect to reliability in conjunction to communication energy efficiency. Further, the proposed game-theoretic model sets the stage for deriving practical distributed query routing algorithms that are reliable and energy-efficient from a sensor-centric point of view.

2.2.2 Technical Approach

2.2.2.1 Detailed Description of Technical Approach

We develop the formal theory behind the analysis of the RQR problem in an adaptive sensor network by solving the following tasks:

1. Formally define and map the RQR problem onto a graph topological reliability domain.
2. Quantify the parameters governing individual and communal (group) sensor node behavior.
3. Derive game/decision theoretic techniques that control the actions of intelligent sensor nodes, leading to the dynamic formation of RQR topologies based on local topological decisions. This task involves rigorous analytical treatment and derivation of relationships between the various parameters including information values, communication costs and query reporting reliability among others.

2.2.2.1.1 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 $v_i \in \mathcal{R}^+$ 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 o. (perhaps based on neighborhood density as proposed in Cerpa 2001) to conserve energy¹. 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 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 obtained². 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.

¹ In this work 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.

² 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.

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 L_i . We focus only on pure strategies in this effort. 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 $|L_i| = 2^{n-1}$. The strategy space of all nodes is given by $L = L_1 \times \dots \times L_n$. 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 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 (Estrin 2000), eventually a single route is established once interest gradients are determined. Our objective in this effort is to compare and evaluate these final routing paths 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 T , rooted at the sink s_q . We now proceed to model the payoffs in this game.

A standard non-cooperative 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 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 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:

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

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

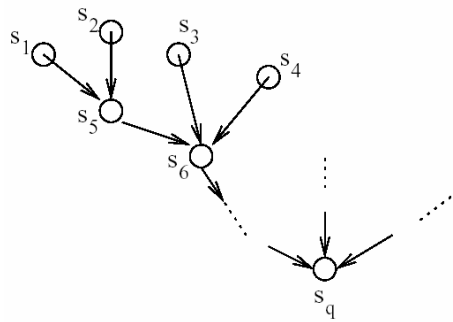


Figure 1: Payoffs with data aggregation.

Consider the data-aggregation tree shown in Figure 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

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

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** \mathbb{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 results 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.

2.2.2.2 Comparison with Current Technology

Current technology on routing in sensor networks focuses primarily on energy-constrained routing by emphasizing the untethered and unattended nature of sensor nodes. Since sensor networks can be deployed in hazardous environments, we consider the problem of routing

under the additional constraint of node survivability. Also, our model is the *first* to consider routing in the context of optimizing individual sensor costs, while taking network wide benefits into account. We therefore classify our approach as sensor-centric, which distinguishes it from other existing models. Our rationale for considering a sensor-centric approach is described below in the technical report section. Current models for communication in sensor networks use protocols like diffusion routing, 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. Furthermore, the lack of an existing theoretical framework in which to analyze such information 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. Other related work such as sensor fusion networks for multiple target tracking using cellular automata models also do not capture the tradeoffs between network reliability, node connectivity and costs. We successfully define a new routing paradigm that explicitly optimizes over both dimensions, i.e., a new model for reliable energy-constrained routing in sensor networks that takes into account all the major constraints of sensor operation as opposed to previous models in this field, which were limited in scope and analysis.

3. Technical Report

3.1 Project Progress

3.1.1 Progress Against Planned Objectives

All three tasks originally proposed have been solved satisfactorily. The formal model for RQR was defined on a sensor network G consisting of N sensors, with one or more querying nodes and the remaining sensors possessing (partial) information related to a specific query. The problem of reporting the results of a query reliably to the querying sensor was abstracted into a graph theoretical problem of embedding reliable, loop free, reverse multicast trees while accounting for the operational constraints of unattended and untethered sensors (with limited communication capacity).

3.1.2 Technical Accomplishments in this Project

We summarize the main results and accomplishments of our effort as follows: 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. We have developed 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 derive some sufficient conditions on communication costs and sensor survival probabilities for well-known routing algorithms such as the most reliable path or least cost neighbor to be congruent to the optimally sensor-centric route. Our analytical model of the abstract RQR problem has set the stage for the future development of practical distributed algorithms for efficiently and reliably routing queries from reporting to querying sensors. While there are many popular routing algorithms for sensor networks that minimize energy consumption, we have shown that routing (or query reporting) must be accomplished within the bounds of *all the four constraints* mentioned above. Ideally sensors should route over the most reliable paths while minimizing their own power/energy consumption rather than some aggregate energy criterion. We note that our paradigm of sensor-centric reliable energy-constrained routing has three benefits:

1. *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.

2. *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.
3. *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 this context, we develop our model for Reliable Query Reporting. 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', 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 intelligent sensors.

We first summarize our results below, followed by a detailed description.

1. We have shown that the solution to the RQR problem lies in designing a routing game with payoff functions such that its Nash equilibrium corresponds to the optimally reliable data aggregation tree [KSI02], [KSI02].
2. We also show that for an arbitrary sensor network G with sensor success probabilities P , communication costs C , and data of value v_i ($0 \leq v_i$), to be routed

from each sensor s_i to the sink s_q , computing the optimally reliable routing path tree P (the maximally reliable energy constrained RQR path) is NP-Hard. Computing the RQR path remains NP-Hard even for the special case when nodes have equal success probabilities. The case when all edges have the same cost is much simpler as described below.

3. Given arbitrary sensor survival probabilities p_i and costs $c_{ij}=c$ for all ij , we have shown that 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.
4. We have derived further bounds on costs and probabilities when the RQR path coincides with easily computable paths such as the most reliable path or the cheapest neighbor path (CNP) from source to sink obtained by each node choosing its successor via the cheapest possible link. In a sense, this path reflects the route obtained when each node has only limited network state information (about neighbor costs and probabilities) and thus minimizes its local communication costs. We have shown 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 the CNP 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 [KSI02].

3.1.2.1 Main 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.

3.1.2.1.1 Complexity Results

Many of the quantities and parameters studied in game theory can at least in principle be computed and approximated. Determining the existence of efficient algorithms for

computing equilibria (and finding such algorithms if they exist) is a problem of common interest to game theory and computer science (Linial 1994). There have been many efforts made to characterize the equilibria of different games in terms of their computational complexity. Gilboa and Zemel (1989) show that finding an equilibrium of a bimatrix game with maximum payoff sum is NP-Hard. Koller and Megiddo (1992) show that finding max-min strategies for two person zero-sum games is NP-Complete in general, but give the first polynomial time algorithm for such games in extensive form. Koller, Megiddo and von Stengel (1996) extend the above result to non-zero sum games, using a complementary pivoting algorithm. Finding optimal strategies for two person games such as chess and go have been shown to be NP-Hard (see Garey and Johnson (1979) for an exhaustive list of known NP-Complete problems). We now address the complexity of finding the equilibrium of the N -person RQR game.

Let $G = (S, E, P, C)$ represent an instance of an information network in which information of value V_r is to be routed from node s_r to 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 player 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 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 T (the RQR tree) is NP-Hard.*

Proof : Given any solution T' to the RQR problem, verifying the optimality of the successor for each node in 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 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 $v_i = 0 \forall i \neq r$. Computing P is NP-Hard.

Proof:

We show that the problem is NP-Hard by considering a reduction from the Hamiltonian Path problem (see Garey and Johnson (1979) for Hamiltonian Path reduction). 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 value³ $V_r = 1$, must be computed as shown in Fig. 2.

Introduce $n+1$ new vertices to form $S = V' \cup T \cup s_q$, where $|T| = n$ and s_q is the other new vertex. The new edge set 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 in E', E_2

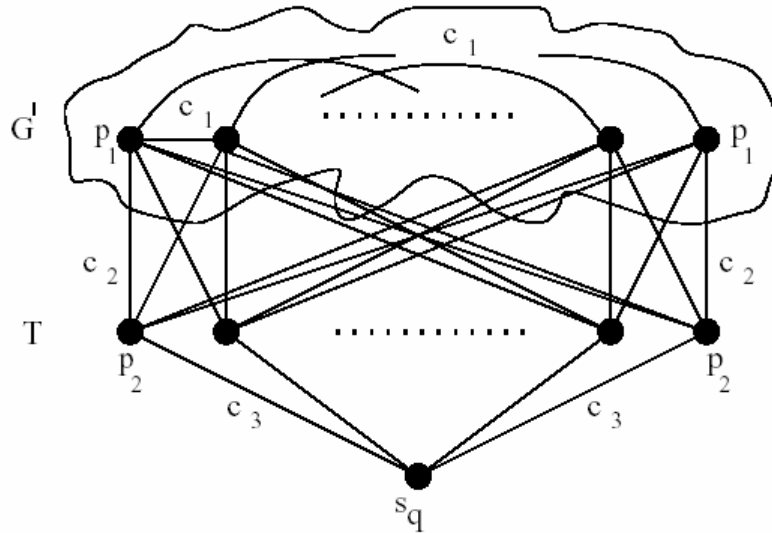


Figure 2: Reduction from Hamiltonian path.

³ We set $V_r=1$ for notational simplicity since results for any V_r can be obtained by scaling edge costs appropriately.

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:

$$p_1 p_2 > \left(\frac{3}{4}\right)^{\frac{1}{n-1}} \quad (1)$$

$$c_1 = \frac{(p_1 p_2)^n}{3} \quad (2)$$

$$c_2 = \frac{2(p_1 p_2)^n}{3} \quad (3)$$

$$c_3 = (p_1 p_2)^n \quad (4)$$

Let s_r and s_t be any two nodes in V' . We claim that there exists an optimal RQR path of reliability $p_1^n p_2$ from s_r to s_q in G if and only if there exists Hamiltonian path from s_r to s_t in G' .

For the first part of the claim, assume there is a Hamiltonian Path $\mathcal{H} = (s_r, \dots, s_t)$ in G' . Consider the path \mathcal{H} followed by the edges (s_t, x) and (x, s_q) in G' , where x is any node in T . This path has reliability $R(\mathcal{H}) = p_1^n p_2$. The payoff of node s_t is $R(\mathcal{H}) - c_2$ obtained by linking to node x , which is optimal since there does not exist any other unvisited node in V' . Similarly the payoff of node x is also optimal since it can only link to s_q . Now consider the k -th node in \mathcal{H} , $1 \leq k \leq n - 1$. The two choices for this node are either to link to some node $x \in T$ or the node in G' that lies on the Hamiltonian path \mathcal{H} . If the first option is chosen, the most reliable alternate path (and hence the maximum possible alternate payoff) is given by $p_1^k p_2 - c_2$ which is less than $R(\mathcal{H}) - c_1$ by conditions (1) - (3). Thus, the second choice is optimal for this node.

For the second part of the claim, we need to show that if no Hamiltonian path exists in G' , there cannot be an optimal RQR path of reliability $p_1^n p_2$. Note that linking to any available node in V' with cost c_2 is always preferable for any node $s_i \in T$. The worst case payoff to s_i via a link of cost c_2 is $p_1^n p_2 - c_2$, which outweighs the best possible payoff via a link of

cost c_3 which is $p_1 p_2 - c_3$. So the optimal path *must* visit all nodes in V' . To maximize payoffs, the optimal path must have the shortest possible length. This will require minimizing visits to T . The optimal path will thus consist of sequences of long paths in V' (the longest possible since any node in V' will always prefer to link to another node in V' , if feasible), interspersed with visits to T . 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 $p_1^n p_2$ as claimed.

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.

3.1.2.1.2 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.

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. For uniform p_i , the equilibrium path is also the path with least overall cost.*

Proof: Consider the most reliable path from the reporting node s_r to the destination node s_q . Clearly, the maximum payoff to s_r is obtained from this path. Given the assumption of uniform costs the payoff to any other sensor $s_i \in S$ on this path must also be maximum.

Otherwise, it would be possible to find a more reliable path from s_r to s_q via s_i .

Note that for uniform p_i , the equilibrium path also coincides with the cheapest path. 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 $c^{\max} = \max_i\{c_i^{\max}\}$ and $c^{\min} = \min_i\{c_i^{\min}\}$.

We use 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 P_r^m .

Proof: Consider an arbitrary node s_i at a distance i from s_r . Since we have uniform p , reliability is now inversely proportional to path length. Let l be the length of the shortest path from s_i to s_q , on which s_{i+1} is the next neighbor of s_i . For s_i , P_a^l is optimal if

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

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 P_r^m , the reliability term above is minimized for $\lambda = l$, whereas the cost term is maximized at $c_{\max}^i - c_{\min}^i$.

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 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 by R_i , the reliability of the most reliable path from s_i to s_q with R'_i being the reliability along any alternative path from s_i . Let $\Delta c_i = c_{i+1} - c_{ij}$ where s_j is any neighbor not on the optimal path and ΔR_i is defined similarly.

Proposition 2 *Given G and $P(s_i) = p_i \in (0, 1]$, the most reliable path from s_r to s_q will be optimal if*

$$\frac{\Delta c_{i+1}}{\Delta c_i} < \frac{\Delta R_{i+1}}{\Delta R_i}$$

for all s_i and s_{i+1} on the most reliable path.

Proof. Let \bar{R}_i represent the reliability on the portion of the most reliable path \bar{P} from s_r to s_i . Since P is optimal, s_i cannot benefit by deviating if

$$\begin{aligned} V_r \bar{R}_i R_i - c_{i+1} &> V_r \bar{R}_i R'_i - c_{ij} \\ \Rightarrow V_r \bar{R}_i &> \frac{\Delta c_i}{\Delta R_i} \end{aligned}$$

It follows that $V_r \bar{R}_{i+1} > \frac{\Delta c_{i+1}}{\Delta R_{i+1}}$. Since $\bar{R}_{i+1} = p_{i+1} \bar{R}_i$, we have $V_r p_{i+1} \bar{R}_i > \frac{\Delta c_{i+1}}{\Delta R_{i+1}}$. This can be

rewritten as $1 \geq p_{i+1} > \frac{\Delta c_{i+1}}{\Delta c_i} \frac{\Delta R_i}{\Delta R_{i+1}}$, which gives us $\frac{\Delta c_{i+1}}{\Delta c_i} < \frac{\Delta R_{i+1}}{\Delta R_i}$ as desired.

The easiest way to interpret this result is by rearranging the terms so that we can write it as

$$\frac{\Delta c_{i+1}}{\Delta R_{i+1}} < \frac{\Delta c_i}{\Delta R_i}$$

Then each fraction can be interpreted as the marginal cost of reliability of deviating from the optimal path. Since each subsequent node on the optimal path has lower expected value of information, this result suggests that the marginal cost of deviation in terms of reliability must be higher for each node's ancestor where the expected value of information is also higher. 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 3 *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*

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

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

Proof: Consider an arbitrary node s_k which is k hops away from s_q on the CNP. Clearly, for the CNP to be optimal s_k should not get higher payoff by deviating to an alternative path. Also, 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 $i = 1, \dots, k-1$, from s_k to s_q to be infeasible, we need

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

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 $c_o = c_{\min}^k$, with c_i being at most $\min\{c^k/c_{\min}^k\}$. Thus, when $\min\{c^k/c_{\min}^k\} - c_{\min}^k > V_r p^l (1 - p^{t-l})$, 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 (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.

3.1.2.2 Technical Accomplishments by Period

1. [June-September 2002] During the initial phase of the project, we derived the basic model for Reliable Query Reporting. We identified the fundamental parameters in the model that correlate to the important constraints on sensor operation, such as independence, energy budget limitations and reliability of paths. This was then used to define a routing game by deriving appropriate payoff functions for sensor nodes that would induce the formation of maximally reliable energy-constrained routing paths from source to destination in the sensor network. Optimal strategies in the routing game lead to optimal adjacent neighbor selection at each node.
2. [October-December 2002] During this period, we first determined that multiple equilibria exist in the routing game. We then evaluated the complexity of computing the maximally reliable energy-constrained routing path among these multiple equilibrium paths. We showed that determining this optimal RQR path was NP-Hard. The publications in [KSI02] and [KSI02] were prepared for submission in February 2003, along with our presentation for the SensIT PI meeting in January.
3. [January-July 2002] During this period, we identified bounds on costs and node survival probabilities that would make the optimal RQR path to practical routing paths obtained by standard routing algorithms such as most reliable path and energy-constrained diffusion routing. During this period, one masters student, Mr. Ysaswi Rachakonda, also worked on the development of a simulation tool for comparing the performance of standard routing algorithms with the optimal RQR path. As part of his Masters thesis (due in May 2003), he will be developing efficient bounded heuristics for optimally reliable energy-constrained routing in large sensor networks. The publications in [GEBO02] and [ICPP02] were prepared and submitted.

3.1.3 Publications

This project led to the following publications. Two papers on reliable query reporting and reliable sensor deployment were presented at the fifth International Conference on Information Fusion in July 2002 ([KSI02], [KSRI02]). Another paper on measuring sensor deployment vulnerability to enemy attack was presented at the International Conference on

Parallel Processing in August 2002 [ICPP02]. An expanded version of this paper is scheduled to appear in the journal ‘Information Processing Letters’ shortly [IPL02]. One paper is currently under review in the journal ‘Games and Economic Behavior’ [GEBO02]. Publication details are listed below.

[KSI02] R. Kannan, S. Sarangi and S. S. Iyengar, “A Simple Model for Reliable Query Reporting in Sensor Networks” in *Fifth International Conference on Information Fusion*, pp. 1025-1030, Annapolis, Maryland, July 2002.

[KSI02] R. Kannan, S. Sarangi, S. Ray and S. S. Iyengar, “Minimal Sensor Integrity in Sensor Grids” in *Fifth International Conference on Information Fusion*, pp. 959-964, Annapolis, Maryland, July 2002.

[GEBO02] R. Kannan, S. Sarangi and S. S. Iyengar, “Strategic Path Reliability in Information Networks” under review in *Games and Economic Behavior*.

[ICPP02] R. Kannan, S. Sarangi, S. Ray and S. S. Iyengar, “Minimal Sensor Integrity: Measuring the vulnerability of Sensor Grids” in *International Conference on Parallel Processing*, Vancouver, B.C, August 2002.

[IPL02] R. Kannan, S. Sarangi, S. Ray and S. S. Iyengar, “Minimal Sensor Integrity: Measuring the Vulnerability of Sensor Deployments” *Information Processing Letters*, to appear.

3.1.4 Meetings and Presentations

The faculty participants in the project traveled to the DARPA SensIT PI meeting held in Santa Fe in January 2002 and presented their results. Additionally, the PI Dr. Kannan attended and presented at the International Conference on Information Fusion in July 2002. The details are as follows:

[1] DARPA SensIT PI meeting; 15-17 January 2002; Santa Fe, New Mexico; Rajgopal

Kannan, S. S. Iyengar and S. Sarangi; Presentation on ‘A model for Reliable Query Reporting’.

[2] International Conference on Information Fusion; 6-8 July 2002; Annapolis, Maryland; Rajgopal Kannan; Presentation on ‘A Simple Model for Reliable Query Reporting in Sensor Networks’.

[3] International Conference on Information Fusion; 6-8 July 2002; Annapolis, Maryland; Rajgopal Kannan; Presentation on ‘Minimal Sensor Integrity in Sensor Grids’.

4. Conclusion

We successfully defined a new paradigm for reliable energy-constrained routing in sensor networks that takes into account all the major constraints of sensor operation as opposed to previous models in this field, which were limited in scope and analysis. Our analytical results have laid the stage for the development of practical distributed routing algorithms under this important paradigm. We believe future work in the area of routing must follow this sensor-centric paradigm. We have obtained complexity results on routing along with bounds on the efficiency of different routing schemes. Three peer reviewed conference papers and one journal paper have been published (or accepted) in this effort with another paper under journal review.