Communication protocols for underwater data collection using a robotic sensor network

Communication Protocols for Underwater Data Collection using a Robotic Sensor Network

Geoffrey A. Hollinger, *Member, IEEE,* Sunav Choudhary, *Student Member, IEEE,*
Parastoo Qarabaqi, *Student Member, IEEE,* Christopher Murphy, *Student Member, IEEE,*
Urbashi Mitra, *Fellow, IEEE,* Gaurav S. Sukhatme, *Fellow, IEEE,* Milica Stojanovic, *Fellow, IEEE,*
Hanumant Singh, *Member, IEEE,* and Franz Hover, *Member, IEEE*

***Abstract*—We examine the problem of collecting data from an underwater sensor network using an autonomous underwater vehicle (AUV). The sensors in the network are equipped with acoustic modems that provide noisy, range-limited communication to the AUV. One challenge in this scenario is to plan paths that maximize the information collected and minimize travel time. While executing a path, the AUV can improve performance by communicating with multiple nodes in the network at once. Such multi-node communication requires a scheduling protocol that is robust to channel variations and interference. To solve this problem, we develop and test a multiple access control protocol for the underwater data collection scenario. We perform simulated experiments that utilize a realistic model of acoustic communication taken from experimental test data. These simulations demonstrate that properly designed scheduling protocols are essential for choosing the appropriate path planning algorithms for data collection.**

***Index Terms*—path planning algorithms, acoustic communication, underwater robotics, sensor networks**

I. Introduction

THE use of sensor fields to monitor phenomena in underwater environments is of growing interest. Examples include algal blooms [1], seismic activity, and intrusion of enemy submarines [2]. In underwater monitoring scenarios, many standard methods of communication are no longer feasible (e.g., WiFi, cellular, satellite). Acoustic modems can provide communication underwater, but they suffer from severe range limitations and channel variations [3].

Without reliable communication, collecting data from underwater sensor networks becomes a challenging problem. A potential solution is the use of a mobile autonomous underwater vehicle (AUV) equipped with an acoustic modem to gather data from the sensors [4]. In the applications of interest, sensors are deployed for long-term monitoring and are fixed to the ocean floor.[1] Hence, we have a robotic sensor network that includes stationary measurement nodes and an AUV that gathers data from these nodes. The problem now becomes one of planning the AUV's path to minimize its travel time and maximize information gathered. We will refer to this as the Communication-Constrained Data Collection Problem (CC-DCP).

In our prior work, we showed that the CC-DCP is closely related to the classical Traveling Salesperson Problem (TSP) [5]. The key difference is that information is gathered from sensors through a noisy channel, whose reliability decreases with distance and can be modeled probabilistically. We previously showed that the CC-DCP can be modeled as a TSP with probabilistic neighborhoods, and we provided algorithms that solve the problem approximately [6].

Related problems have been studied in the context of robotic data mules. Bhadauria and Isler developed approximation algorithms for multiple data mules that must traverse a sensor field and download data [7]. In their work, downloading time is considered as part of the tour, and the communication radii are assumed to be uniform, fixed, and deterministic (i.e., data from a sensor is known to be accessible at a given location). Vasilescu et al. demonstrated a system of mobile and stationary nodes for underwater data collection based on the use of both optical and acoustic communication [4]. They described the networking architecture and sensor specifications necessary for underwater data collection, and they presented experiments in the field on a mobile network. These experiments demonstrate the feasibility of utilizing AUVs for underwater data collection, but the authors leave open the problem of both path planning and communication scheduling.

Algorithms in prior work were designed under the assumption that the AUV would only communicate with a single node at a time. To overcome this limitation, we develop a scheduling protocol that allows the AUV to communicate with multiple nodes at once while performing the tour. In this paper, we design and test a Time Division Multiple Access (TDMA) based protocol and use the results to select parameters for an AUV path planning algorithm. The key novelty of this paper

G. Hollinger and G. Sukhatme are with the Computer Science Department, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089 USA, (e-mail: {gahollin,gaurav}@usc.edu).

S. Choudhary and U. Mitra are with the Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: {sunavcho,ubli}@usc.edu).

P. Qarabaqi and M. Stojanovic are with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA, (e-mail: {qarabaqi,millitsa}@ece.neu.edu).

C. Murphy and H. Singh are with the Deep Submergence Laboratory, Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: {cmurphy,hsingh}@whoi.edu).

F. Hover is with the Center for Ocean Engineering, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: hover@mit.edu).

This research has been funded in part by the following grants: ONR N00014-09-1-0700, ONR N00014-07-1-00738, NSF 0831728, NSF CCR-0120778, and NSF CNS-1035866.

[1] In some cases the nodes may move slowly over time. If we make the assumption that the nodes are nearly stationary for a given data collection interval, our methods apply to these cases as well.

is the use of scheduling protocols to inform path planning methods for AUVs. The proposed methods are validated through simulated experiments that utilize models built on experimental data from an AUV deployment.

II. Problem Setup

We are given a pre-deployed network of N sensors located in $\mathbb{R}^{dim}$. For this paper, we limit analysis to $dim \in \{2,3\}$, which yields the 2D and 3D problems respectively. We assume that the location $x_n \in \mathbb{R}^{dim}$ is given for each sensor $n \in \mathcal{S}$, where $\mathcal{S}$ is the set of deployed sensors. Each sensor n contains data for retrieval, which we denote as Y_n. The data consists of packets, and the number of packets is denoted as $|Y_n| = N_p$. We define the *information quality* of the data as $I(Y_n)$, which corresponds to the expected value of information (e.g., information gain in an inference problem [8], or variance reduction in a regression problem [9]). In the general case, coupling between the sensor measurements can lead to information being subadditive or superadditive. In the context of data collection, we will assume that information is additive (i.e., $I(Y_i, Y_j) = I(Y_i) + I(Y_j)$ for all $i \neq j$). Extensions to subadditive information is a subject of ongoing work.

The sensors are assumed to have limited capabilities. Each sensor is capable of transmitting packets of data over a noisy channel. A single mobile vehicle has the capability to communicate with the sensors. The location $x_v \in \mathbb{R}^{dim}$ of the vehicle is controlled and may be subject to constraints, such as obstacles or vehicle kinematics. Based on these constraints, a *traversal cost* $c(x_1, x_2)$ is defined for all pairs of points $x_1, x_2 \in \mathbb{R}^{dim}$. We assume that the traversal cost obeys the triangle inequality and that the location of the AUV is known. Example traversal costs include Euclidean distance and time to arrival. The *communication quality* of a location degrades with distance: $\mathcal{C}(x_v, x_n) = f(D(x_v, x_n))$, where $D(x_v, x_n) = |x_n - x_v|$, and f decreases monotonically with distance.

The optimization problem is to plan a path $\mathcal{P} = [x_v(1), x_v(2), \ldots, x_v(T)]$ for the vehicle that retrieves data from the sensors and minimizes the traversal cost of the path. In prior work, we assumed that the AUV communicates with a single sensor at a given time, which allow for simple methods to calculate the expected information $R(\mathcal{P})$ along a path [6]. Relaxing this assumption requires the development of more sophisticated techniques to calculate the information quality at a given AUV location (see Section IV). Given an expression for $R(\mathcal{P})$, we can write the Communication-Constrained Data Collection Problem (CC-DCP) formally.

Problem 1: Given path costs c, expected information quality R, and a set of possible AUV paths Ψ, find

$$\mathcal{P}^* = \operatorname*{argmin}_{\mathcal{P} \in \Psi} \sum_{t=2}^{T} c(\mathcal{P}(t-1), \mathcal{P}(t)) \text{ s.t. } R(\mathcal{P}) \geq \beta, \quad (1)$$

where T is the index of the last location on the path, and β is a threshold on information quality. The value of β can be tuned depending on the desired weighting of information quality and cost. Higher information quality thresholds will require additional cost (communication cost and/or traversal cost).

III. Planning Algorithms

While it is possible to calculate a single best placement for the AUV to maximize information gathered, it is significantly more difficult to find a maximally informative information gathering path. In fact, the resulting CC-DCP problem is a generalization of the TSP, which makes it NP-hard. Due to the computational intractability of finding an optimal tour for networks with many nodes, heuristics are necessary to solve the CC-DCP approximately. In prior work, we introduced the idea of generating contours of equal probability around the sensors and utilizing these contours as if they were deterministic neighborhoods. We give a brief overview of the algorithm here and direct the reader to our prior work for additional detail [6].

We define a *probabilistic neighborhood* $\mathcal{G}_n \subset \mathbb{R}^{dim}$ as all locations x_v where the probability of successful data transfer $P(x_v, x_n)$ is greater than p. The value of $p \in (0, 1)$ determines how conservative the probabilistic neighborhood is. As $p \to 1$, it will be near certain that information will be received from sensor n if the AUV is within the neighborhood. As $p \to 0$, the AUV may need to query a sensor multiple times before receiving data from it. In Section V, we run experiments to determine the value of p that maximizes the information to cost ratio.

Once the probabilistic neighborhoods are defined, we can generate a covering set of neighborhoods by greedily choosing sensors and removing adjacent sensors within the resulting neighborhood. A valid tour can then be found by visiting the neighborhoods in the covering set [10]. The resulting algorithm requires a TSP solver for calculating a neighborhood tour. We utilize the Concorde solver for this task [5].

This path planning algorithm was shown in prior work to outperform existing methods, including a simple reactive strategy and a standard TSP solution [6]. These prior results did not consider multiple access communication. In the present paper, we allow the vehicle to execute a multiple access protocol to all nodes in the neighborhood once it reaches the center of the neighborhood. We assume that no communication occurs while the vehicle is moving between neighborhoods, which allows the neighborhoods to remain static during the information exchange. Relaxing this assumption is an avenue for future work.

IV. Acoustic Communication

Acoustic propagation is characterized by energy spreading and absorption that occur in an unobstructed medium over a single propagation path, as well as by additional distortions caused by multipath propagation (i.e., surface-bottom reflections and refraction due to sound speed variation with depth [11]). Ray tracing offers an accurate picture of the resulting sound field at a given frequency and a given location in a frozen ocean, and tools such as the Bellhop code [12] are typically used to predict the signal strength prior to system deployment. However, the actual signal strength, observed in a finite bandwidth and over finite intervals of time during which the transmitter/receiver become slightly displaced around their nominal locations or the surface conditions change, deviates from the so-obtained value. These variations appear as random, and our goal is to describe them statistically.

A. Data from AUV Deployment

We utilize data acquired by the AUV *Lucille*. *Lucille*, a SeaBED-class AUV [13] operated by the NOAA Northwest Fisheries Science Center, is equipped with a WHOI Micro-Modem and 12.5 kHz ITC-3013 hemispherical transducer for acoustic communications [14]. In September of 2010, *Lucille* assisted in mapping the submerged portion of the San Andreas Fault off Northern California, at approximately $39\,°50'\mathrm{N}, 124\,°\mathrm{W}$. During this survey, the AUV's onboard networking stack, capable of handling data fragmentation and image compression [15], transmitted one three-second packet every five seconds. These packets were encoded using both Frequency-Hopping Frequency Shift Keying (FH-FSK) and Phase Shift Keying (PSK), and transmitted using 4-5 kHz bandwidth around a center frequency of 10 kHz.

Throughout the course of the dive, the vehicle maintained a constant altitude above the seafloor of 3 m, at a depth of approximately 130 m. The surface ship, the R/V Pacific Storm, varied in slant range from 200 m to 1 km from the vehicle. The surface ship remained underway with the hydraulics running during this experiment, resulting in significant noise being generated across all frequencies, including those used for communication. These conditions are typically experienced by AUVs operating from near-shore vessels on the continental shelf.

B. Acoustic Channel Model

To specify a propagation model, we represent the gain as

$$g(d,t) = \overline{g}(d) + y(t), \tag{2}$$

where $\overline{g}(d)$ is the mean value of the gain at a distance d,[2] and $y(t)$ is a random process. In this model, the gain $\overline{g}(d)$ represents the expected communication quality $C(x_v, x_n)$ when $d = D(x_v, x_n)$ (see Section II). We do not consider changes in water pressure with depth, which would affect the propagation speed. Such changes could be accounted for in the signal processing layer by inserting guard time slots to account for slight variation in propagation speed.

We now proceed to establish two models based on our experimental data: one that relates the mean value $\overline{g}$ to the distance d, and another that specifies the probability distribution function (pdf) of the random component y. We utilize a channel model similar to prior work [3] that identifies log-distance parameters. We also add an additional random component, and we specify the overall power loss, including all frequencies and all propagation paths. These models will be valid for the chosen operating conditions (frequency band and transmission distances). Specifically, we make the following conjectures:

(i) the mean value obeys a log-distance model

$$\overline{g}(d) = g_0 - k_0 \cdot 10 \log d \tag{3}$$

[2]The distance is varying with time, i.e. $d = d(t)$.

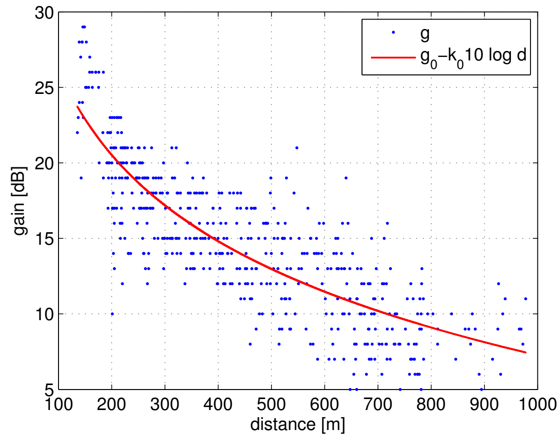

Fig. 1. Gain (normalized) vs. transmission distance. Dots show measured values; solid curve shows an estimated trend (a first-order logarithmic-scale polynomial fit to the ensemble mean at each distance yields k_0=1.9).

(ii) the random component obeys a Gaussian distribution, $y \sim \mathcal{N}(0, \sigma^2)$.

Figure 1 summarizes the recorded values (from the deployment described above) of the gain as a function of distance. The solid curve represents the log-distance model (3), whose parameters g_0 and k_0 were obtained by first-order polynomial fitting.[3] We emphasize again that the model parameters will in general depend on the operational conditions, i.e. that the values indicated in the figure are representative of the 8-12 kHz acoustic band and transmission distances on the order of several hundreds of meters.

Shown in Figure 2 is the histogram of the random component $y = g - \overline{g}$. This figure motivates our second conjecture, i.e. the Gaussian model for y. The variance σ^2 is calculated from the data at hand. We note that its value appears to be invariant for the range of distances considered, although greater distance spans could require sectioning. We also note that the variance will depend on the bandwidth, decreasing as the bandwidth increases. Similar conclusions have been found using different data sets [16].

C. Packet Error Modeling

We utilize an underwater acoustic noise model developed in prior work [2], [11]. This model accounts for noise factors in the environment, such as wind and shipping activity, as well as thermal noise and turbulence. We also assume a block log-normal fading model for the received signal-to-noise ratio (SNR) based on section IV-B. Let $\overline{P_S}$ be the probability of symbol error averaged over the SNRs. For a packet with Q symbols encoded with a code of rate r, the average packet-error-rate is given by

$$\overline{P_D} = 1 - (1 - \overline{P_S})^{rQ}. \tag{4}$$

There is no known simple approximation to $\overline{P_S}$ when SNR is log-normally distributed, and we employ Monte-Carlo methods to perform simulations. In this model, the packet success

[3]Logarithms are taken with base 10.

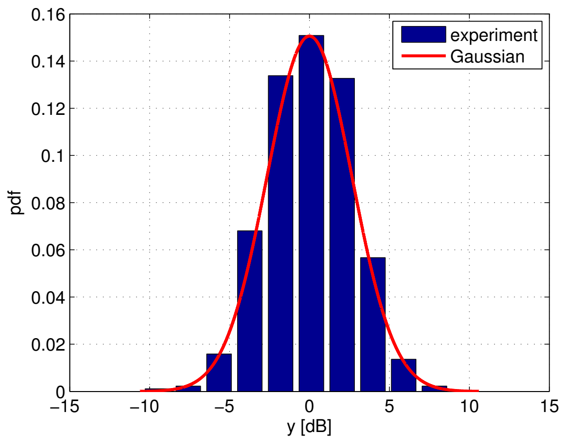

Fig. 2. Histogram of the measured deviation y and the theoretical p.d.f. of a zero-mean Gaussian random variable with σ^2=6.7 dB.

rate of $1 - \overline{P_D}$ between a vehicle at x_v and a sensor at x_n represents the expected communication quality $\mathcal{C}(x_v, x_n)$ (see Section II).

D. Scheduling Protocol

We assume a set of sensor nodes with fixed locations and that synchronization amongst the nodes has been accomplished. Synchronization among sensors is a hard problem in general, but is relatively easy if the locations are fixed and known. Thus, we do not address synchronization protocol specific issues in this paper. We further assume a single carrier, half-duplex communication system. We describe below a three phase multiple access control protocol based on Time-Division-Multiple-Access with Acknowledgement (TDMA-ACK):

1) *Initiation:* The sensors begin in a sleep state. The AUV sends a high power broadcast wake-packet which brings the sensors into an active state and contains initial communication schedules for all sensors.
2) *Scheduling:* The functional sensors, which received the broadcast correctly, reply with an acknowledgement according to the schedule. The AUV selects a subset of the functional sensors and sends out the next round of scheduling information to this subset.
3) *Data Transfer:* The sensors reply with data packets. After all sensors have completed their transmissions, if any packet fails, the AUV re-schedules the corresponding sensors with an Automated-Repeat-Request (ARQ) for the failed packets.

The number of sensors in a neighborhood has an upper bound, which is assumed to be known at the AUV for the initiation phase. Replies to the broadcast wake-packet are assumed to include a sensor identification header for further rounds of scheduling as demanded by the protocol. The actual broadcast wake sequence and corresponding reply sequences can be customized in order to provide an estimate of the average quality for each sensor to AUV link. Such information helps the AUV perform sensor selection in the scheduling phase. As part of the implementation, the number of ARQs executed in the data transfer phase would also be upper bounded.

E. Protocol Analysis

We make a few simplifying assumptions in order to compute the communication throughput. In particular, we let fading be independent across all distinct sensor to AUV links as well as across retransmissions over the same link. Spatial independence of fading is only assumed as a first approximation, and addressing correlated fading is an avenue for future work. We assume that all sensors are equally informative, and thus each sensor transmits one unit of information in a single packet. While the transmissions from the sensors to the AUV will incur errors, we assume that transmissions from the AUV to the sensors are perfectly decoded. Information across sensors is assumed to be independent. The case of correlated information is a subject of ongoing work.

In a given neighborhood we assume a total of M sensors, where $M \leq N$ total sensors. Let L_i denote the location of sensor i and $|L_i|$ denote its distance from the AUV. Let $\{s(1), \ldots, s(M_s)\}$ be the set of M_s functional sensors selected in step 1 of the protocol. We assume that the sensors are indexed to satisfy $|L_i| \leq |L_j|$ and $|L_{s(i)}| \leq |L_{s(j)}|$ whenever $i < j$. Let B_A and B_D respectively denote the sizes of ACK and DATA packets and $B_S^{(M)}$ and $B_S^{(M_s)}$ quantify scheduling packet sizes for M and M_s sensors respectively. Let M_A denote the number of distinct scheduling slots required for one round of ACK transmission in step 1. Let $P_D(\gamma)$ denote the probability of data packet error given an instantaneous SNR of γ, and $\overline{P_D}$ be the packet-error-rate averaged over γ. Let N_p denote the total number of data packets per sensor.

We next compute the expected information transferred and the expected cost of communication (in seconds) for a maximum of K transmission rounds. If a packet fails, a retransmission occurs. Thus a packet is not transferred only if it fails in K rounds of communication. We define the *information gain* from sensor $s(i)$ after K rounds of transmission as:

$$\begin{aligned} I_{s(i)} &= \#\text{ of packets} \times \text{prob. of success} \\ &= N_p \cdot \left(1 - \prod_{k=1}^{K} P_D\left(\gamma_{s(i)}^{(k)}\right)\right), \end{aligned} \tag{5}$$

where $\gamma_{s(i)}^{(k)}$ denotes the instantaneous SNR for sensor $s(i)$ during the k^{th} round. The total information gain is then:

$$\begin{aligned} I &= \sum_{i=1}^{M_s} I_{s(i)} \\ &= N_p \cdot \left[M_s - \sum_{i=1}^{M_s} \prod_{k=1}^{K} P_D\left(\gamma_{s(i)}^{(k)}\right)\right]. \end{aligned} \tag{6}$$

Given the set $\{s(i)\}$ of selected sensors, let $\Gamma = \{\gamma_{s(i)}^{(k)}\}$ be the set of all random SNRs in (6). By our independent fading assumptions, the expectation, with respect to Γ, of the total information gain is computed as:

$$\mathbb{E}_{\Gamma}\left[I\right] = N_p \cdot \left[M_s - \sum_{i=1}^{M_s} \overline{P_D^{s(i)}}^{K}\right]. \tag{7}$$

For a data collection path $\mathcal{P}$ that visits a number of neighborhoods, we can sum the information gain I across the entire path to calculate a value for the expected information quality $R(\mathcal{P})$ (see Section II). The expected information quality provides a metric for evaluating that path.

Next we calculate the cost of communication. The initiation phase has a broadcast of size $B_S^{(M)}$. This must reach the farthest sensor, so round-trip propagation delay is $C_1 \cdot 2 \cdot |L_M|$ and the transmission cost is $C_2 \cdot B_S^{(M)}$. In the scheduling phase, the reception time for all ACK packets is $C_2 \cdot B_A \cdot M_A$ and each subsequent scheduling broadcast takes $C_2 \cdot B_S^{(M_s)}$ transmission time and the worst case round-trip propagation time of $C_1 \cdot 2 \cdot |L_{s(M_s)}|$. After the k^{th} round, the number of data packets left to transmit at sensor $s(i)$ is $N_p \cdot \prod_{l=1}^{k} P_D(\gamma_{s(i)}^{(l)})$, which determines the duration of the next scheduling slot. If τ_{max} is the maximum delay spread, we need a guard interval of $2 \cdot M_s \cdot \tau_{max}$ for each transmission round and $2 \cdot M \cdot \tau_{max}$ for the initiation phase. Summing over all transmission rounds across all sensors, we have the communication cost as:

$$\begin{aligned} t &= \text{Initiation Cost} + K \cdot \text{Scheduling Cost} \\ &\quad + \text{Guard Interval} + \text{Data Transfer Cost} \\ &= 2 \cdot C_1 \cdot |L_M| + C_2 \cdot B_S^{(M)} + C_2 \cdot B_A \cdot M_A \\ &\quad + K \cdot \left(2 \cdot C_1 \cdot |L_{s(M_s)}| + C_2 \cdot B_S^{(M_s)}\right) \\ &\quad + 2 \cdot M \cdot \tau_{max} + 2 \cdot K \cdot M_s \cdot \tau_{max} \\ &\quad + C_2 \cdot B_D \cdot N_p \cdot \sum_{k=0}^{K-1} \sum_{i=1}^{M_s} \prod_{l=1}^{k} P_D(\gamma_{s(i)}^{(l)}). \end{aligned} \quad (8)$$

From the independent fading assumptions and (8), the expectation, with respect to Γ, of the total cost of communication becomes:

$$\begin{aligned} \mathbb{E}_\Gamma[t] &= 2 \cdot C_1 \cdot |L_M| + C_2 \cdot B_S^{(M)} + C_2 \cdot B_A \cdot M_A \\ &\quad + K \cdot \left(2 \cdot C_1 \cdot |L_{s(M_s)}| + C_2 \cdot B_S^{(M_s)}\right) \\ &\quad + 2 \cdot M \cdot \tau_{max} + 2 \cdot K \cdot M_s \cdot \tau_{max} \\ &\quad + C_2 \cdot B_D \cdot N_p \cdot \sum_{i=1}^{M_s} \left(\frac{1 - \overline{P_D^{s(i)}}^K}{1 - \overline{P_D^{s(i)}}} \right). \end{aligned} \quad (9)$$

When evaluating the total cost of a data collection tour, the cost of communication is added to the traversal cost to calculate a total mission time.

V. Simulations

A simulation environment was implemented in C++ running on Ubuntu Linux to test our proposed CC-DCP algorithms. The simulated experiments were run on a 3.2 GHz Intel i7 processor with 9 GB of RAM. We examine the performance of the proposed scheduling protocol integrated with the contour-based TSP path planning algorithm. One-hundred random 2D deployments of 100 sensors were generated in a 2D 1 km $\times$ 1 km area, and a simulated AUV was added to the environment that moves at a speed of 1 m/s. The AUV executes a plan found using the path planning algorithm described in Section III. The packet error modeling described in Section IV was used to determine which packets are successfully received by the AUV.

Simulations were run with varying values of the parameter p, which represents the size of the probabilistic neighborhoods (see Section III). The number of automated repeat requests (ARQs) was also varied. These two parameters represent design decisions when implementing the contour-based TSP algorithm. For link quality simulation, each link is assigned a distance dependent loss as well as a random log-normally distributed loss. Each random loss is selected independently, and the average packet error rate (APER) is numerically calculated. The APER is used directly to calculate information gain and communication cost.

Figure 3 shows the results of these simulations. We note that, since the sensors are equally informative and received information is additive, information gain is equivalent to the number of distinct packets received. The AUV speed was used to calculate a traversal time between points in the 2D space, which was used as the traversal cost. The total cost of the mission is the sum of the traversal time and the communication time. As expected, both information gain and communication cost increase as the number of ARQs is increased. In addition, increases in contour probability (corresponding to decreased neighborhood size) lead to longer paths for the AUV and increased cost.

More interesting observations arise when we examine the gain to cost ratio in Figure 3. We see that the gain to cost ratio first increases with increasing ARQs and then decreases. Additionally, the gain to cost ratio maximum appears at a different ARQ value for varying probabilistic neighborhood size. The highest gain to cost ratio occurs with $p = 0.7$ and $ARQ = 6$, which provides optimized performance for the path planning algorithm. Additionally, if we look at the gain/cost frontier, we see that we can tune the solution based on different weightings of cost and gain. By varying the value of p and the ARQs, we have built up a frontier of solutions that tradeoff between mission time and information gain. These simulations provide an empirical method for selecting the value of p that maximizes the information to cost ratio.

VI. Conclusions and Future Work

This paper has demonstrated the benefit of utilizing scheduling protocols to design path planning algorithms for autonomous underwater data collection. We have shown that simulated analysis with varying parameters can be used to build up a frontier of solutions that tradeoff between mission time and information gain. Without such analysis, it would not be possible to generate this frontier of solutions, and the path planning algorithm would need to execute blindly. Thus, improved scheduling protocols and analysis of communication provide powerful tools for optimizing path planning algorithms in data collection scenarios.

A number of interesting extensions provide avenues for future research. This paper assumes that communication does not occur while the vehicle is moving between neighborhoods. Incorporating this functionality would require more complex modeling of the (time-varying) information gain. In addition,

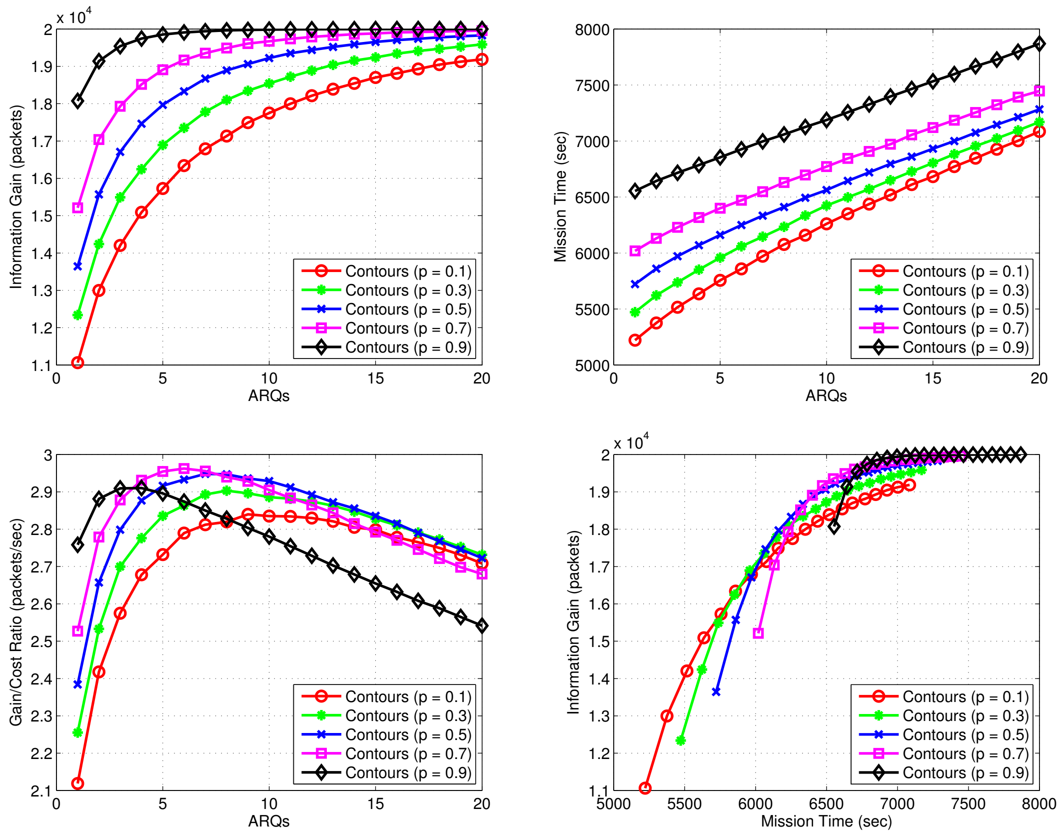

Fig. 3. Simulations of an AUV collecting data from an underwater sensor network. Averages are over 100 random deployments in a 1 km $\times$ 1 km area with 100 nodes. The AUV executes a data collection tour found using a TSP with neighborhoods. The simulations are performed with varying neighborhood size and number of ARQs.

we are in the process of deriving equations to calculate information gain when correlations exist between sensors, which causes the the value of information to become subadditive.

Acknowledgment

The authors gratefully acknowledge Jonathan Binney, Arvind Pereira, Hordur Heidarsson, and Srinivas Yerramalli at the University of Southern California for their insightful comments. Thanks also to Chris Goldfinger of Oregon State University, the captain and crew of the R/V Pacific Storm, and Elizabeth Clarke of the NOAA Northwest Fisheries Science Center for their support of this work.

References

[1] R. N. Smith, Y. Chao, P. P. Li, D. A. Caron, B. H. Jones, and G. S. Sukhatme, "Planning and implementing trajectories for autonomous underwater vehicles to track evolving ocean processes based on predictions from a regional ocean model," *Int. J. Robotics Research*, vol. 29, no. 12, pp. 1475–1497, 2010.

[2] G. Hollinger, S. Yerramalli, S. Singh, U. Mitra, and G. S. Sukhatme, "Distributed coordination and data fusion for underwater search," in *Proc. IEEE Conf. Robotics and Automation*, 2011, pp. 349–355.

[3] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007.

[4] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, "Data collection, storage, and retrieval with an underwater sensor network," in *Proc. Int. Conf. Embedded Networked Sensor Systems*, 2005, pp. 154–165.

[5] D. L. Applegate, R. E. Bixby, V. Chvátal, and W. J. Cook, *The Traveling Salesman Problem: A Computational Study*. Princeton Univ. Press, 2006.

[6] G. Hollinger, U. Mitra, and G. Sukhatme, "Mobile underwater data collection using acoustic communication," in *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2011.

[7] D. Bhadauria and V. Isler, "Data gathering tours for mobile robots," in *IEEE Int. Conf. Intelligent Robots and Systems*, 2009, pp. 3868–3873.

[8] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," in *Proc. Uncertainty in Artificial Intelligence*, 2005.

[9] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communication cost," in *Proc. Information Processing in Sensor Networks*, 2006, pp. 2–10.

[10] A. Dumitrescu and J. Mitchell, "Approximation algorithms for TSP with neighborhoods in the plane," *J. Algorithms*, vol. 48, no. 1, pp. 135–159, 2003.

[11] L. Berkhovskikh and Y. Lysanov, *Fundamentals of Ocean Acoustics*. Springer, 1982.

[12] M.Porter, "Bellhop code," available online at http://oalib.hlsresearch.com/Rays/index.html.

[13] H. Singh, A. Can, R. Eustice, S. Lerner, N. McPhee, O. Pizarro, and C. Roman, "Seabed AUV offers new platform for high-resolution imaging," *EOS Trans. AGU*, vol. 85, no. 31, 2004.

[14] L. Freitag, M. Grund, S. Singh, J. Partan, P. Koski, and K. Ball, "The WHOI micro-modem: An acoustic communcations and navigation system for multiple platforms," in *Proc. IEEE Oceans Conf.*, 2005, pp. 1086–1092.

[15] C. Murphy and H. Singh, "Wavelet compression with set partitioning for low bandwidth telemetry from AUVs," in *Proc. ACM Int. Wkshp. UnderWater Networks*, 2010.

[16] P. Qarabaqi and M. Stojanovic, "Adaptive power control for underwater acoustic channels," in *Proc. IEEE Oceans Conf.*, 2011.