Maximizing the Value of Sensed Information in Underwater Wireless Sensor Networks via an Autonomous Underwater Vehicle

Stefano Basagni⋆, Ladislau Bölöni‡, Petrika Gjanci†, Chiara Petrioli†, Cynthia A. Phillips◇, Damla Turgut‡
⋆ ECE Dept., Northeastern University, Boston, MA, basagni@ece.neu.edu
‡ EECS Dept., University of Central Florida, Orlando, FL, {turgut,lboloni}@eecs.ucf.edu
† Dipartimento di Informatica, Università di Roma "La Sapienza," Roma, Italy, {gjanci,petrioli}@di.uniroma1.it
◇ Sandia National Laboratories, Albuquerque, NM, caphill@sandia.gov

Abstract*—This paper considers underwater wireless sensor networks (UWSNs) for submarine surveillance and monitoring. Nodes produce data with an associated value, decaying in time. An autonomous underwater vehicle (AUV) is sent to retrieve information from the nodes, through optical communication, and periodically emerges to deliver the collected data to a sink, located on the surface or onshore. Our objective is to determine a collection path for the AUV so that the *Value of Information* (VoI) of the data delivered to the sink is maximized. To this purpose, we first define an Integer Linear Programming (ILP) model for path planning that considers realistic data communication rates, distances, and surfacing constraints. We then define the first heuristic for path finding that is fully distributed and adaptive to the occurrence of new events, relying only on acoustic communication for exchanging short control messages. Our *Greedy and Adaptive AUV Path-finding* (GAAP) heuristic drives the AUV to collect packets from nodes to maximize the VoI of the delivered data. We compare the VoI of data obtained by running the optimum solution derived by the ILP model to that obtained from running GAAP over UWSNs with realistic and desirable size. In our experiments GAAP consistently delivers more than 80% of the theoretical maximum VoI determined by the ILP model.

I. Introduction

Underwater exploration and monitoring has emerged as a vital part of the economy and the safety infrastructure of many countries. Its applications range from aquaculture to intrusion detection and prevention, including oil industry deployments, telecommunications, pollution and climate control, search missions and the preservation of cultural heritage [1]. The current, costly approach to underwater monitoring relies on tethered vehicles, cabled monitoring stations, or simply leaving instrumentation on site and then retrieving it periodically. As a consequence, only those applications needed by large corporations (e.g., telecommunications) and governments (e.g., defense) have been implemented. For instance, all continents but Antarctica are interconnected through thousands of kilometers of optical cables, mostly serving the telecommunication industry, the only industry able to spend hundreds of millions of dollars to lay cable on the bottom of the ocean. In recent years, advances in acoustic and optical underwater communication have made new underwater monitoring applications feasible and cost effective. This includes port surveillance and safety, oil platform monitoring, and protection of cultural heritage at underwater sites. These applications rely on sensor nodes mounting cameras and sonar systems that are able to communicate wirelessly among themselves or with passing vessels and underwater vehicles in an *Underwater Wireless Sensor Network* (UWSN). These applications require nodes to transfer a large amount of recorded data in a reasonable amount of time. Acoustic communication, the prevailing underwater communication technology, is insufficient for these applications because acoustic modems' data rate is only a few tens of Kbps. Optical devices allow data rates of several Mbps [2], [3], but only when nodes are close to each other. Thus, this type of communication is most useful in scenarios where statically deployed sensor nodes are visited periodically by mobile data collectors, which can retrieve the sensed data through the high speed optical connection [4].

In this paper we consider a UWSN scenario where sensor nodes are deployed underwater for monitoring purposes. Nodes sense and record data, and wait for an *autonomous underwater vehicle* (AUV) to arrive to collect data. In order to transfer large amount of recorded data, such as videos or high resolution images, the nodes and the AUV are endowed with optical communication devices. Nodes are also equipped with acoustic modems to exchange control information with the AUV. For instance, if a node has important data to deliver, it can send a short control packet over the acoustic channel to the AUV. Periodically, the AUV resurfaces and wirelessly (RF) transmits the collected data to a collection point (a *sink*), located on the surface. Fig. 1 illustrates the networks we consider in this paper.

The information collected by the nodes varies in size, value and urgency, depending on the specific application. It is also unpredictable. It is not possible to know in advance when an event of interest will happen, and which nodes will detect it. In any case, the *value of information* (VoI) of an event is highest at the moment the event is detected, and it then decays with time. Therefore, packets reporting an event should be delivered to the sink as soon as possible: The later the data reaches the sink, the lower its value. Our goal is to investigate the theoretical and practical challenges of planning the path of the AUV to maximize the total VoI delivered to the sink.

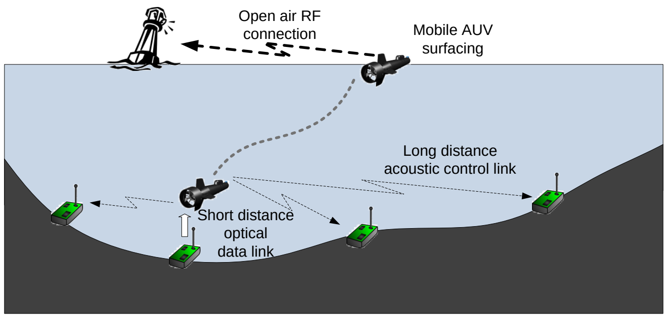

Fig. 1. A UWSN where nodes are visited by an AUV.

We summarize the contributions of this paper as follows:

1) We provide a new Integer Linear Programming (ILP) model for determining AUV paths that maximize the VoI of data delivered to the sink. Our mathematical model provides provable bounds on the best possible network performance (e.g., the best achievable VoI) for benchmarking distributed protocols. The model is independent of sensor deployment strategies, and has parameters for controlling data generation rate, data transmission rates, and AUV speeds. Our solution allows us to compute an upper bound on the maximum VoI retrievable from networks whose size is comparable to that of actual (4 to 6 nodes) and desirable (9 to 12 nodes) sizes. To the best of the authors' knowledge, this is the first model for AUV path planning that takes into account the VoI.
2) We define a realistically deployable distributed heuristic for AUV path planning that adapts to events occurring at unpredictable locations and times. The AUV chooses the next node to be visited based on the VoI it expects to collect at the next location. The information needed to make this decision is propagated to the AUV using short *event packets* transmitted through the acoustic channel. The AUV plans to visit a node that has sent an event packet if and only if visiting that node increases the VoI of the data it will deliver to the sink. Because it makes decisions based on what is best at the moment, and adapts the path planning process to new information, we call our heuristic *Greedy and Adaptive AUV Path-planning*, or GAAP for short.
3) We demonstrate the effectiveness of GAAP in delivering data with high VoI by evaluating its performance in different scenarios. We start by showing the results of a comparative performance evaluation of GAAP against the ILP-based upper bound, termed OPT. The experiments show that GAAP achieves VoI that is remarkably close to that achieved by OPT. In fact, in scenarios with variable number of nodes (from 4 to 12) and events, we observe that GAAP always delivers more than 80% of the theoretical maximum VoI obtained by OPT. We further show how selecting the right starting point impacts the performance of GAAP. Finally, to demonstrate the benefit of VoI awareness, we perform a comparative performance evaluation of GAAP vs. a simple non-VoI-aware heuristic where the AUV travels a pre-established path visiting all nodes. Our results show clearly that, by explicitly considering VoI for path planning, GAAP approaches the VoI-optimum paths provided by OPT.

The paper is organized as follows. Section II defines the problem we consider in detail. In Section III we present an ILP formulation for solving the problem of AUV path planning. A distributed heuristic for AUV routing is presented in Section IV. Performance evaluation and comparisons between GAAP, OPT and a non-VoI-aware solution is presented in Section V. Section VI reviews literature on the topics of this paper. Finally, Section VII concludes the paper.

II. Problem Definition

We consider a scenario where a set S of sensor nodes (or simply, nodes) $S_1, \ldots, S_{|S|}$ are statically deployed in a 3D geographic underwater area. Nodes perform surveillance operations for a given time T. The location of each node is known (from manual deployment or from localization techniques [5]), and therefore, given any pair of nodes, so is the distance between them. The nodes perform continuous sensing (e.g., taking videos), with node S_i storing the sensed data chunk p_{it} at time t, $0 \leq t < T$. The data $p_{i\tau}$ observed by a node S_i at a given time τ has a *value of information* $\mathcal{V}_{p_{i\tau}}(t)$, at time $t \geq \tau$. $\mathcal{V}_{p_{i\tau}}(t)$ is a monotonically decreasing function of t. The VoI of a data chunk is highest at the moment when it was sensed $\mathcal{V}_{p_{i\tau}}(\tau)$; this base value varies depending on the importance of the information captured in the data chunk.

An Autonomous Underwater Vehicle (AUV) travels from node to node to collect the sensed information via high-data-rate optical communication [2], [3], [4]. Specifically, during T time units the AUV visits some nodes at their locations, collects some sensed data, and periodically surfaces to offload what it has collected to a data collection point on the sea surface (a *sink*). Communications between the sink and the AUV happen via high-data-rate wireless communications. The path the AUV follows is therefore a sequence of runs in each of which the AUV visits a number of the nodes to collect some of their data and surfaces to report that data to the sink. Specifically, during the kth run, the AUV makes a set of visits $\{\ldots(S_i, t_i^k)\ldots\}$, with node S_i visited at time t_i^k.

Let $R = \{1, \ldots, r\}$ be the set of runs performed by the AUV within time T. Node S_i is visited by the AUV during a set of runs $R_i \subseteq R$. The AUV might not visit every node in every run, so in general $|R_i| \leq r$. Consider a run $k \in R_i$. Let pred(k, R_i) be the last run the AUV visited node S_i before run k. Thus, if the runs in R_i are sorted by time, pred(k, R_i) is the largest run in R_i less than $k \in R_i$. By definition $t_i^{pred(k,R_i)} = 0$ when k is the smallest element of R_i. The value of information for node S_i given a path P of the AUV will be:

$$\mathcal{V}(S_i, P) = \sum_{k \in R_i} \sum_{h=t_i^{pred(k,R_i)}}^{t_i^k} \mathcal{V}_{p_{ih}}(t_{ih}),$$

where t_{ih} is the time the AUV delivers the packet collected from node S_i at time h to the sink. In other words, the value of information of the data sensed by node S_i and delivered to the sink by the AUV traveling path P during time T is given by summing the values of information collected by S_i between two consecutive visits of the AUV when each packet is delivered to the sink. The AUV must be on the surface to deliver packets to the sink. After each run, it delivers the packets it has collected during the run.

The AUV path planning problem is then stated as follow.

Given $|S|$ nodes and their locations, given a set of surfacing locations, and given the value of information of the sensed data, determine the path P_{opt} (sequence of nodes and surfacing locations) of the AUV so that the value of information is maximized:

$$P_{opt} = \max_P \left(\sum_{i=1}^{|S|} \mathcal{V}(S_i, P) \right).$$

We assume the AUV begins on the surface and it ends on the surface, since any collected, but undelivered packets will be worth nothing.

III. A Mathematical Model for AUV paths

We present a mathematical model for the problem defined above. We start by defining the sets, the parameters and the variables used for formalizing our problem.

Definitions, sets and parameters

- T is the length of network operations, in time units (numbered from 0 to $T-1$). Nodes produce data and the AUV travels to collect and deliver data during this time. The last time the sink can receive data is T, the data sent by the AUV at time $T-1$.
- S is the set of nodes (and their locations).
- W is the set of surfacing locations.
- $N = S \cup W$ is the set of all locations to which the AUV can travel and sojourn to either receive or transmit packets, respectively.
- ω is a fictitious location, indicating in fact that the AUV is in transit, i.e., it is not at any actual location in N.
- d_{ij} is the time it takes for the AUV to travel between any two locations i and j in N, in time units. Time distance d_{ij} is easily derived from the known position of nodes and surfacing locations and from the known AUV speed.
- τ_i is the shortest travel time for the AUV to go from node $S_i \in S$ to some location $w \in W$ on the surface and vice versa (assumed symmetrical), in time units.
- A *packet* is the amount of data that a node produces in a time unit, in bits.
- u_c indicates the number of packets that the AUV can collect from a node in a time unit.
- u_b indicates the number of packets that the AUV can transmit (broadcast) to the sink in a time unit.
- V_{i,t_1,t_2} is the value of information captured by node $S_i \in S$ at time t_1 and delivered at time t_2.

We introduce some notation to make the model easier to understand and write. For each node $S_i \in S$ we call the set of feasible data capture times $\mathcal{T}_{i1}$. Given a node S_i and a capture time $t_1 \in \mathcal{T}_{i1}$, the set of possible delivery times is $\mathcal{T}_{i12}$. Let F be the set of all feasible tuples (i, t_1, t_2) for a node S_i, capture time $t_1 \in \mathcal{T}_{i1}$ for a packet at S_i and delivery time $t_2 \in \mathcal{T}_{i12}$ for the packet captured at time t_1. Let $F(t_2)$ denote the set of packet (sensor source, capture-time) pairs that can be legally delivered at time t_2. Let $F(i, t_2)$ denote the set of data capture times t_1 of packets from node S_i that can be delivered at time t_2. For any location $n \in N \cup \{\omega\}$, let $L(n)$ denote the times where it would make sense for the AUV to be at location n. Finally, we denote with $C(\tau, i, t)$ the set of all location-time pairs (j, τ) such that the AUV cannot be at position j at time τ if at time t it is at location i. (Detailed definitions of these sets are provided in the Appendix.)

Variables

- $x_{it_1t_2}$: Binary variable taking the value 1 if a packet captured by node $S_i \in S$ at time t_1 is delivered at time t_2; 0 otherwise.
- $c_{it_1t_2}$: Binary variable taking the value 1 if the AUV collects at time t_2 a packet captured by node $S_i \in S$ at time t_1; 0 otherwise.
- z_{nt}: Binary variable taking the value 1 if the AUV is at location n at time t, $n \in N \cup \{\omega\}$; 0 otherwise.
- y_t: Binary variable taking the value 1 if the AUV is on the surface (at one of the locations in W) at time t; 0 otherwise. Note that y_t is used to make the model description more succinct, since it is $y_t = \sum_{w \in W} z_{wt}$.

ILP formulation

The objective function maximizes the value of information collected from all nodes and delivered by time T.

$$\text{maximize} \sum_{(i,t_1,t_2) \in F} \mathcal{V}_{(i,t_1,t_2)} x_{it_1t_2}$$

subject to the following constraints.

$$\sum_{w \in W} z_{w0} = 1 \tag{1}$$

$$\sum_{w \in W} z_{wT} = 1 \tag{2}$$

$$\sum_{n \in N \cup \{\omega\}} z_{nt} = 1, \ \forall t = 1 \ldots T-1 \tag{3}$$

$$\sum_{(j,\tau) \in C(\tau^*, i, t)} z_{j\tau} \le 1 - z_{it}, \forall i \in N, t \in L(i), \tau^* \in T^* \tag{4}$$

These first four sets of constraints concern the AUV locations. Its path should start and end at a location on the surface (constraints (1) and (2), respectively). At a given time the AUV should be either at one location or it should be in transit (constraints (3)). Constraints (4) enforce travel time. They state that if the AUV is at location i at time t, it cannot be in any other location j during the time needed to travel from location j to i. (Detailed definitions for $C(\tau^*, i, t)$ and T^* are given in the Appendix.)

$$x_{it_1t_2} \leq y_{t_2},\ \forall (i, t_1, t_2) \in F \tag{5}$$

Constraints (5) require that if the AUV is delivering data at time t_2, at that time it must be on the surface.

$$x_{it_1t_2} \leq \sum_{\tau=\max\{\tau_i+1, t_1+2\}}^{t_2-\tau_i-1} c_{it_1\tau},\ \forall (i, t_1, t_2) \in F \tag{6}$$

Constraints (6) force the AUV to collect data from a node after the node has captured the data, and early enough to get to the surface in time to deliver it.

$$\sum_{t_2 \in \mathcal{T}_{i12}} x_{it_1t_2} \leq 1,\ \forall S_i \in S, t_1 \in \mathcal{T}_{i1} \tag{7}$$

With constraints (7) we force a packet to be delivered at most once.

$$c_{it_1t_2} \leq z_{it_2},\ \forall (i, t_1, t_2 + \tau_i + 1) \in F \tag{8}$$

Constraints (8) say that in order to collect packets from node $S_i \in S$ the AUV must be at S_i. If $t_2+\tau_i+1$ is a legal delivery time, then t_2 is a legal collection time.

$$\sum_{t_1 \in F(i,t_2)} c_{it_1t_2} \leq u_c z_{it_2-1}$$
$$\forall S_i \in S,\ \forall t_2 : \exists (i, t_1, t_2 + \tau_i + 1) \in F \tag{9}$$
$$\sum_{(i,t_i) \in F(t_2)} x_{it_1t_2} \leq u_b y_{t_2-1},\ \forall t_2 : \exists (i, t_1, t_2) \in F \tag{10}$$

Constraints (9) and (10) enforce the capacity constraints on data collection from a sensor and on data transmission to the sink, respectively.

The remaining constraints concern the domain and definition of the variables:

$$x_{it_1t_2} \in \{0, 1\},\ \forall (i, t_1, t_2) \in F$$
$$c_{it_1t_2} \in \{0, 1\},\ \forall (i, t_1, t_2 + \tau_i + 1) \in F$$
$$z_{nt} \in \{0, 1\},\ \forall n \in N \cup \{\omega\}, t \in L(n)$$
$$y_t = \sum_{w \in W} z_{wt},\ \forall t = 0 \ldots T-1$$

IV. GAAP: Greedy and Adaptive AUV Path-planning

We describe a path planning mechanism for the AUV that is both distributed and adaptive to the occurrence of new events, thus allowing for realistic utilization in UWSNs.

We assume that the AUV receives small control packets from any node in the network through acoustic communications, directly (single hop communication), via the sink, or through a simple flooding mechanism. The packets, called *event packets*, are meant to convey the value of the event currently being sensed at a node. The starting point of the AUV path is set, and so are the possible surfacing locations. By the end of network operations (i.e., at time T), the AUV will be on the surface.

The AUV follows a greedy strategy for visiting the nodes. Upon receiving an event packet from a node A, it determines whether changing its current path and visiting node A, collecting and delivering its data, would produce a higher VoI or not, and bases its decision on that information. Because the approach to path planning follows a greedy strategy and it adapts to the dynamic occurrences of events, we named our heuristic *GAAP*, for Greedy and Adaptive AUV Path-planning.

More specifically, GAAP makes the AUV move as follows. Initially the AUV travels from the surface to the closest node, say A. If while moving towards A it receives an event packet from another node B, the AUV decides whether to keep going to A or to plan to visit B depending on the highest VoI that it would be able to deliver to the sink. In particular, the AUV considers the times that it takes to travel to A and B, respectively, and the VoI of the data that A and B produce in that time (as if both nodes would keep producing data with that values for that time). It also considers the time it would take to resurface and to transmit the collected packets to the sink. Let t_A and t_B be the total time it would take to deliver data from A and B, respectively, and let $\mathcal{V}_A$ and $\mathcal{V}_B$ be the VoI delivered to the sink. Then the AUV moves to A if and only if $\frac{\mathcal{V}_A}{t_A} \geq \frac{\mathcal{V}_B}{t_B}$. When it arrives at the selected node, say, B, the AUV collects all packets that B generated since the AUV received the event packet from B. In fact, it collects all packets available at B that maximize the VoI, considering the time to surface and deliver those packets to the sink. It then surfaces and offloads the collected packets to the sink. Upon completing the transmission to the sink, the AUV moves to the node that, according to the event packets received so far, can provide the highest VoI, and so on.

The selection of which data packets are to be collected first from a node or to be transmitted first to the sink is driven by their VoI: Packets that obtain the highest VoI are transmitted first.

V. Performance Evaluation

In this section we discuss the results of a simulation-based performance evaluation of the solutions proposed in this paper. We first introduce the simulation scenarios and their parameters. We have organized the experiments into three parts. We start by comparing the performance of GAAP to the upper bound provided by running the ILP model (termed OPT). We then show the impact of selecting the AUV starting point on the performance of GAAP, comparing how GAAP behaves when starting from different surfacing points, including the optimal point determined by OPT. Finally, we compare the performance of GAAP to that of a simple heuristic where the AUV movements are not VoI aware. This last set of experiments is performed in networks with varying density to show how well GAAP scales with increasing numbers of nodes deployed in the same geographic area.

A. Simulators and simulation scenarios

The results for OPT have been obtained by solving the ILP model defined in Section III with the freely available solver Gurobi (http://www.gurobi.com/) run on Linux-based 64-bit multiple-core servers with default settings. Each of the three servers we used has 16 cores, clocked at 2.8GHz, and 64GB of RAM. The various runs took from a few hours to a few days to produce the optimal solutions. We implemented GAAP (Section IV) in a home-grown software framework written in Java. In all the scenarios we considered, GAAP executions were extremely fast, never lasting more than few seconds.

All our experiments consider realistic parameters of UWSNs. We consider topologies with $|S| = 4$, 5, 9 and 12 wireless underwater sensor nodes deployed over a rectangular area 2km×3km. Nodes are deployed at a depth chosen at random between 50m and 100m. Fig. 2 shows the layout of the topologies we considered as seen from above. We consider

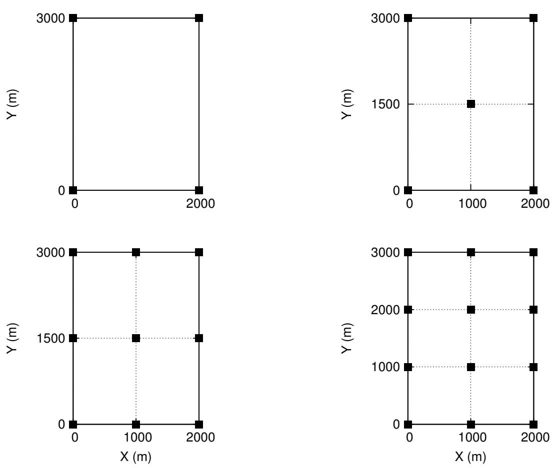

Fig. 2. The topologies considered in our experiments.

scenarios with $|S|$ surfacing points, located directly above each of the $|S|$ nodes. We set the AUV cruise speed to 1.8m/s, to match an Odyssey-IV class vehicle [6].

Each node sends event packets to the AUV over the acoustic data channel. Statistical data on the transmission delays of event packets from a node to the AUV are considered, consistent with their distance and an acoustic channel data rate of 10Kbps. We gathered this data by running the SUNSET flooding algorithm, a framework for underwater emulation/simulation [7], multiple times. We set the optical and wireless data transfer rate to 10Mbps. Optical communication can reach this data rate when the AUV hovers within 100m of a node [2], [3]. The muddier the water, the closer the AUV must be to the node.

We simulate a scenario where the nodes use cameras to take videos for intrusion detection. Surveillance data are stored as 720p high-definition videos, with a resolution of

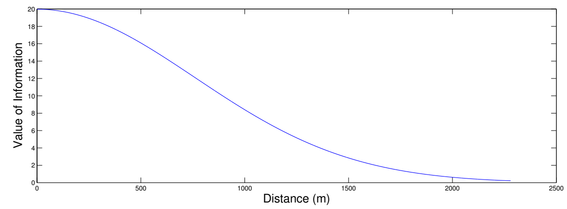

Fig. 3. Event value vs. distance.

1280×720 pixels at 3 frames/s. Each node produces packets corresponding to 5 minute of recorded video. Assuming a video encoded using the standard H.264 codec, a five minute recording produces a 9MB packet. Timeliness of delivery of these data is considered very important. For this reason, we stipulate that the VoI of video packets decays exponentially in time.

We model event arrival as a Poisson process with arrival rate parameter $\lambda = 1$ hour. Once generated, an event is assigned a location within the 3D network deployment area randomly and uniformly. It is also assigned a random duration that is exponentially distributed, with 1 hour average. We stipulate that the event is sensed, and therefore recorded and reported, by only one node, namely, the one geographically closest to the event location. (Unlikely ties are broken by using the nodes unique ID.) Events have values varying between 0.4 and 20. The actual value perceived and reported by the sensing node depends on the distance from the node and the location of the event according to a Gaussian-like distribution (Fig. 3). For instance, an event that happens a few meters from a node will be reported with its full value (20). If the event happens instead farther away from the node, its reported value will be lower, according to the selected distribution curve. For example, at 1000m the value is 9 according to the curve of Fig. 3). The duration of network operation T is set to 12 hours, and each time unit lasts 5 minutes.

B. Results: GAAP vs. OPT

Fig. 4 shows the values of the VoI for data delivered by the AUV in networks with increasing number of nodes. The results we present here are obtained by averaging over 18 experiments for each displayed bar for both GAAP and OPT. In each experiment we randomly vary the set of the events and their location. For the GAAP heuristic, the AUV always starts at the most central surface location. For the scenario with 4 nodes we choose one of the four points randomly as a starting point. As expected, OPT outperforms GAAP because the ILP instance knows in advance where and when, and for how long, the events are going to happen and ILP solvers use a provably complete intelligent search algorithm. The gap between the two, however, is reasonably low: GAAP obtains a total VoI 20% lower than that of OPT in networks with 4 nodes, 16.12% lower than that of OPT in networks with 5 nodes, and only 15.84% and 12.5% lower than that of OPT in networks with 9 and 12 nodes, respectively. Overall, the VoI

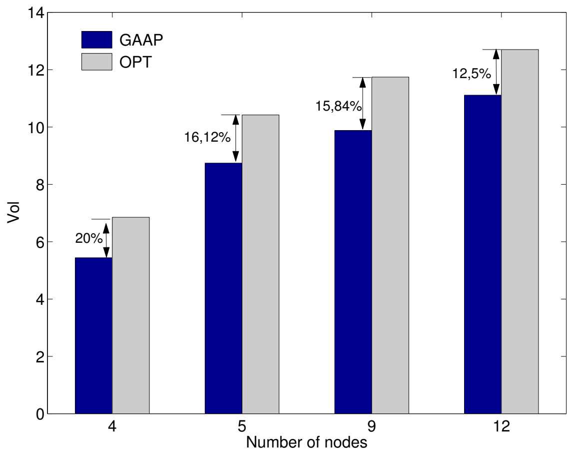

Fig. 4. GAAP vs. OPT.

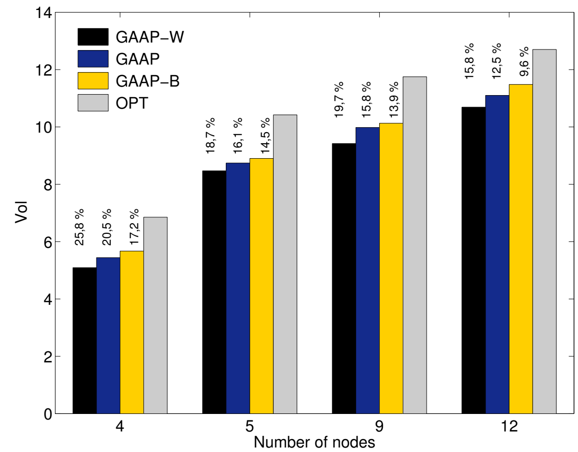

Fig. 5. VoI obtained when the AUV starts from different surface points.

provided by GAAP is never more than 20% lower than that of OPT.

We observe that the higher the number of nodes, the higher the retrieved VoI. This is because the higher network density decreases the distance between a sensing node and an event, and therefore the event value as perceived by the node is higher. Moreover, since the nodes are closer to each other, the AUV needs to travel less between pairs of nodes, and it is able to make better "predictions" about the VoI it could retrieve by visiting a node reporting a new event.

C. Results: Impact of the starting point

The first move of the AUV according to GAAP is to travel to the closest node. So in our experiments we have chosen to start the AUV at the surface location closest to the center of the deployment area. All points are equivalent in the 4 node scenarios, so in that case we choose one of them randomly. As mentioned above, this choice is quite acceptable, being a reasonable starting point in the absence of other data and providing results that are reasonably close to those produced by OPT. We ran a set of experiments to show the impact of different choices of the AUV starting point on the VoI performance of GAAP. In particular, in addition to the VoI obtained by GAAP and OPT as shown above, we consider the VoI of data delivered by GAAP to the sink in the following cases:

- GAAP-B: The AUV starts from the best starting point on the surface, which is the one determined by OPT.
- GAAP-W: The AUV starts from the worst starting point on the surface, which is the one that obtains the lowest possible VoI.

Fig. 5 shows the average VoI for GAAP-W, GAAP, GAAP-B and OPT in the considered scenarios. The percentages on top of the GAAP bars indicate the percentage reduction in VoI compared to that achieved through OPT. We observe that the performance of GAAP-B is closer to that of OPT than that of GAAP, as expected. Especially in denser scenarios (12 nodes and 12 surfacing points), the improvement of GAAP-B over GAAP is non-negligible: The gap between GAAP-B and OPT is now 9.6%, a reduction from the 12.5% gap observed between GAAP and OPT. These results show that a non-optimal starting point can impose a penalty on the retrieved VoI from the start, a penalty that GAAP is not able to recover from as time goes on. This is particularly true in denser networks where, because of the closer proximity of events to nodes, the reported VoI is higher, and so is the toll to pay for starting from a non-optimal starting point.

Starting from the wrong surfacing points leads to noticeable decreases in performance, as shown by the gaps between GAAP-W and OPT. The VoI obtained at the sink when the AUV starts from the worst possible starting point is always at least 15.8% lower than that obtained by OPT. The gap is particularly high in small networks (4 nodes and 4 surfacing points), where it nears 26%. This is because of the low number of nodes, their locations, and of the greater distances that the AUV has to travel in the network. Going to the first node from the farthest location takes a long time, and also takes a lot of VoI away from the final total.

D. Results: Impact of VoI awareness

The third set of experiments shows that the VoI-aware approach taken by GAAP for AUV path planning provides a remarkably good design choice. We have already seen that GAAP delivers VoI that is never more than 20% lower than that obtained by OPT. Here we perform a comparative performance evaluation between GAAP and a simple non-VoI-aware heuristic, where the AUV travels a pre-established path visiting all network nodes. The heuristic is modeled as a traveling salesman problem (TSP) based on the distance between nodes, so we name it AUV-TSP. According to this heuristic, the AUV starts from a surfacing point (the same as GAAP) and visits the closest node, say A. Then, regardless of the data value, it collects all data produced by node A

so far, resurfaces, and delivers them to the sink. The AUV then proceeds to the following node according to the pre-established path, till the time limit T for network operations is reached. Our experiments are run over the topologies of Figure 2 and also on two new kinds of topologies with 18 and 35 nodes deployed as a grid in the 2km$\times$3km area (Fig. 6). Therefore, our experiments also show that GAAP scales well

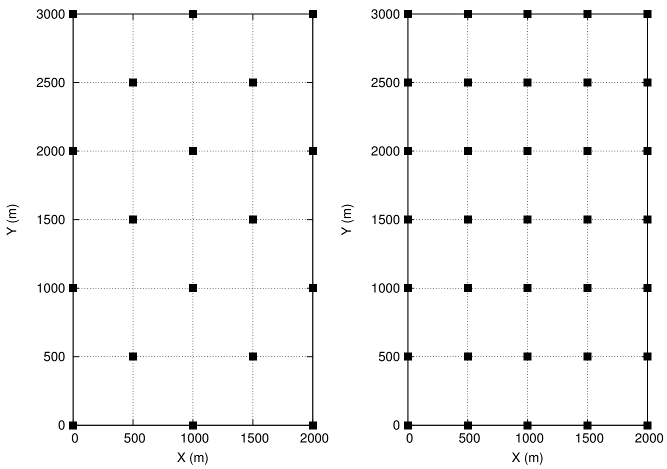

Fig. 6. Topologies with 18 (left) and 35 (right) nodes.

with increasing network densities (number of nodes). The results we present here are obtained by averaging over 200 experiments for each displayed bar for both GAAP and AUV-TSP. In each experiment we randomly vary the set of the events and their location.

Fig. 7 shows quantitative evidence that the VoI-aware approach followed by GAAP is effective in retrieving higher values of VoI. The figure shows the average VoI of data delivered by the AUV to the sink in networks with increasing number of nodes when the AUV follows paths found by GAAP (darker color bars) and AUV-TSP (lighter color bars). For all

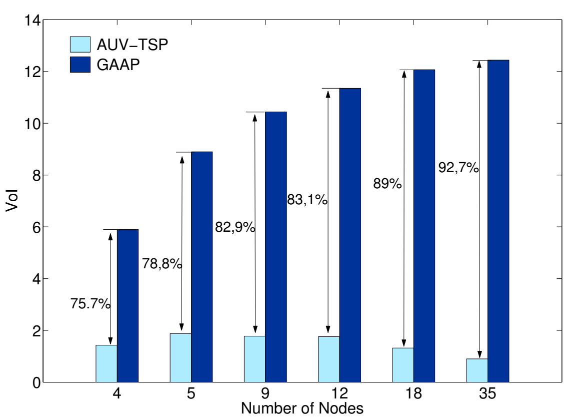

Fig. 7. GAAP vs. a TSP-like heuristic.

network sizes we tested, AUV-TSP delivers data whose VoI is always at least 75% lower than that delivered by GAAP. The highest gap occurs in networks with a larger number of nodes, where AUV-TSP performs poorly because following a pre-established path forces the AUV to visit nodes sensing events with no relevant value.

The good performance scalability of GAAP is confirmed by the increasing values of VoI it produces in networks with 18 and 35 nodes. The VoI for data delivered to the sink by the AUV moving according to GAAP keeps growing, since the higher network density makes events closer to the nodes, and therefore their perceived value is higher. As noted before, the nodes are also closer to each other, which makes the AUV traveling time shorter.

The performance of AUV-TSP grows from networks with 4 nodes to networks with 5 nodes, going from an average 1.4 of retrieved VoI to an average 1.8 This is because of the presence of a central node in the topology of networks with 5 nodes (Fig. 2), which shortens the distances traveled by the AUV. Increasing the network size from 5 to 35 nodes, the performance of AUV-TSP monotonically decreases, from 1.8 down to 0.9 of retrieved VoI. This is because of the larger number of nodes sensing events with basically no value that the AUV is forced to visit. Furthermore, for networks with more nodes (18 and 35 nodes) we observe that by the end of the network operations, i.e., by time T, the AUV could not visit all nodes. As a consequence, potentially highly valuable data were never retrieved at all.

The results shown in this section demonstrate the effectiveness of the choice performed by GAAP to consider VoI for path planning. On one side, GAAP obtains a VoI that is remarkably close to the theoretical optimum provided by OPT, almost independently of the starting point and on the knowledge of the occurrence of future events. On the other side, it easily outperform strategies, such as AUV-TSP, that are not VoI aware, and scales well to UWSNs with increasing number of nodes, thus being a viable option for use in real-life applications.

VI. Related Works

UWSNs have received significant attention in recent years, as a challenging application scenario with important practical applications, and as a powerful motivator for new theoretical advances. One of the challenges is that each of the alternatives of underwater communications, acoustic and optical, have serious drawbacks. Because surveillance applications, especially those involving large production of data (e.g., video, sonar-generated files), require swift data transfers, many research projects investigate the combination of different approaches to data transfer, such as acoustic networking, short-distance optical transport and physical transport of data with an AUV. The majority of these projects assume long-distance acoustic communications supplemented by an AUV collecting data from the nodes through a free-space optical transmission and then physically transporting it to the sink. For instance, Vasilescu et al. [8] and Detweiller et al. [9] describe a system for long-term monitoring of fisheries and coral reefs. The acoustic and optical communication protocols are integrated into the TinyOS operating system of sensor nodes and very simple AUVs. A more recent application of a similar scenario is the system designed at the Woods Hole Oceanographic institute described in Farr et al. [10], [11]. In particular, [10] describes the system of integrated optical and acoustic capabilities. In [11] the authors describe an application where an AUV is sent to offload data from a seafloor borehole observatory. Hollinger et al. [12] consider an approach where an AUV visits various neighborhoods in the UWSN. The AUV can communicate with more than one nearby sensor node using acoustic communication. However, the probability of packet loss in acoustic networks increases with distance. This leads to a formulation of a problem where the path of the AUV must be determined so that it optimizes the balance between the cost of the path and the sum of the probabilistically determined quality of information of the collected data. A variation of this model that does not consider the presence of the AUV is the MURAO protocol described by Hu and Fey [13]. MURAO is a protocol where nodes are divided into two classes: The group leaders have acoustic and optical modems and coordinate the group members that have bidirectional optical capabilities and acoustic receivers, but no acoustic transmitters. The group leaders play only a coordination role in the data transmission. The MURAO protocol uses multi-level Q-learning (a variant of reinforcement learning) to determine the paths. The node values and reward from the environment is transferred by dedicated control packets over the optical channels.

One of the characteristics of the UWSN scenario is that the data rate of modern sensors greatly exceeds the capabilities of underwater networking channels. It is simply impossible to transfer all the sensed data in near-real time. Therefore, if the value of a data chunk varies in time, the system needs to make decisions about which data chunks are to be transferred preferentially. We argue that the intellectual framework of *value of information* as considered here is an appropriate, disciplined way to make these decisions. The concept of VoI was originally proposed in game theory [14]. The intuition behind the game theoretical definition of VoI is the price an optimal player would pay for a piece of information. In artificial intelligence, autonomous agents often need to balance actions taken to acquire information with actions that actually further their goals. In these settings, the concept of VoI might help balance the action of an agent, by preventing it from wasting its resources pursuing the collection of information of low value [15]. In the context of sensor networks, a number of recent projects have introduced similar metrics to model situations where one either needs to select a subset of the collected data or choose between transmitting a piece of information or not. Bisdikian et al. [16], [17] define "Quality of Information" (QoI) as the degree to which a piece of information is (or is perceived to be) fit-to-use for a particular purpose. In contrast to definitions originating from game theory, this definition has its origins in the enterprise/database/data quality community. The QoI is usually conceived as a vector of quality attributes that include accuracy, latency, and spatiotemporal relevance. Turgut et al. [18], [19] defined a variant called "pragmatic VoI" as the support the information gives to the decisions and actions of the operator (without assuming an optimal decision-maker). In a UWSN setting similar to the one considered in this paper, Bölöni et al. [20] discuss a scenario where the VoI is used to balance between the direct acoustic transmission of digests of large data chunks to the sink and the optical transfer of the whole chunk to an AUV moving according to a fixed trajectory. This paper considers the decay of the VoI in time (in form of an exponential decay function), the fraction of VoI that is retained by digests of the original data, and introduces the concept of conditional VoI, which captures the novelty of a data chunk in the context of previously transferred data.

None of these previous works, namely, those concerned with path planning for AUVs or those on VoI in networking, tackle the problem of planning paths for an AUV to deliver the sensed data with the highest VoI.

VII. Conclusions

We presented a mathematical model and a distributed heuristic for path finding for a AUV collecting data with decaying value from nodes of a UWSN. The heuristic drives the AUV to visit the node that greedily maximizes the Value of Information of the data delivered to the sink. Our ILP model considers realistic data communication rates, distances and surfacing constraints. Our Greedy and Adaptive AUV Path-finding (GAAP) heuristic successfully mimics the optimal paths and obtains VoI of the delivered data that is at most 20% lower than that obtained by the ILP model, as shown by simulations over networks with increasing number of nodes.

References

[1] J. Heidemann, M. Stojanovic, and M. Zorzi. Underwater sensor networks: Applications, advances and challenges. *Philosophical Transactions of the Royal Society A*, 370:158–175, August 2 2012.

[2] S. Harnon. Underwater optical wireless communication network. *Optical Engineering*, 49(1):1–6, January 2010.

[3] F. Hanson and S. Radic. High bandwidth underwater optical communication. *Applied Optics*, 47(2):277–283, 2008.

[4] M. Doniec, C. Detweiler, I. Vasilescu, and D. Rus. Using optical communication for remote underwater robot operation. In *Proceedings of IEEE/RSJ Intelligent Robots and Systems, IROS 2010*, pages 4017–4022, Taipei, Taiwan, October 18–22 2010.

[5] M. Erol-Kantarci, H. T. Mouftah, and S. F. Oktug. A survey of architectures and localization techniques for underwater acoustic sensor networks. *IEEE Communications Surveys & Tutorials*, 13(3):487–502, third quarter 2011.

[6] S. Desset, R. Damus, F. Hover, J. Morash, and V. Polidoro. Closer to deep underwater science with ODYSSEY IV class hovering autonomous underwater vehicle (HAUV). In *Proceedings of Oceans 2005—Europe*, volume 2, pages 758–762, Brest, France, June 20–23 2005.

[7] C. Petrioli and R. Petroccia. SUNSET: Simulation, Emulation and Real-life Testing of Underwater Wireless Sensor Networks. In *Proceedings of IEEE UComms 2012*, Sestri Levante, Italy, September, 12–14 2012.

[8] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke. Data collection, storage, and retrieval with an underwater sensor network. In *Proceedings of ACM SenSys 2005*, pages 154–165, San Diego, CA, November 2–4 2005.

[9] C. Detweiler, I. Vasilescu, and D. Rus. An underwater sensor network with dual communications, sensing, and mobility. In *Proceedings of OCEANS 2007—Europe*, pages 1–6, Aberdeen, Scotland, June 18–21 2007.

[10] N. Farr, A. Bowen, J. Ware, C. Pontbriand, and M. Tivey. An integrated, underwater optical/acoustic communications system. In *Proceedings of the MTS/IEEE OCEANS 2010*, pages 1–6, Sydney, Australia, May 24–27 2010.

[11] N. Farr, J. Ware, C. Pontbriand, T. Hammar, and M. Tivey. Optical communication system expands CORK seafloor observatory's bandwidth. In *Proceedings of the MTS/IEEE OCEANS 2010*, pages 1–6, Seattle, WA, September 20–23 2010.

[12] G. A. Hollinger, S. Choudhary, P. Qarabaqi, C. Murphy, U. Mitra, G. S. Sukhatme, M. Stojanovic, H. Singh, and F. Hover. Underwater data collection using robotic sensor networks. *IEEE Journal on Selected Areas in Communications*, 30(5):899–911, 2012.

[13] T. Hu and Y. Fei. MURAO: A multi-level routing protocol for acoustic-optical hybrid underwater wireless sensor networks. In *Proceedings of SECON 2012*, pages 218–226, Seul, Korea, June 18–21 2012.

[14] R. A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

[15] E. Kamar and E. Horvitz. Light at the end of the tunnel: A Monte Carlo approach to computing value of information. In *Proceedings of AAMAS 2013*, pages 571–578, Saint Paul, MN, May 6–10 2013.

[16] C. Bisdikian, J. Branch, K. K. Leung, and R. I. Young. A letter soup for the quality of information in sensor networks. In *Proceedings of IQ2S 2009*, pages 1–6, Galveston, TX, March 9 2009.

[17] C. Bisdikian, L. M. Kaplan, M. B. Srivastava, D. J. Thornley, D. Verma, and R. I. Young. Building principles for a quality of information specification for sensor information. In *Proceedings of IEEE FUSION 2009*, pages 1370–1377, Seattle, WA, July 6–9 2009.

[18] D. Turgut and L. Bölöni. A pragmatic value of information approach for intruder tracking sensor networks. In *Proceedings of IEEE ICC 2012*, pages 4931–4936, Ottawa, Canada, June 10–15 2012.

[19] D. Turgut and L. Bölöni. IVE: Improving the value of information in energy-constrained intruder tracking sensor networks. In *Proceedings of IEEE ICC 2013*, pages 4953–4957, Budapest, Hungary, June 9–13 2013.

[20] L. Bölöni, D. Turgut, S. Basagni, and C. Petrioli. Scheduling data transmissions of underwater sensor nodes for maximizing value of information. In *Proceedings of IEEE Globecom 2013, Ad Hoc and Sensor Networking Symposium*, Atlanta, GA, December 9–13 2013.

APPENDIX

We provide details of the sets used for the description of the ILP model of Section III, omitted in that section to ease readability.

For each node $S_i \in S$ the set of feasible data capture times are $\mathcal{T}_{i1} = 0 \ldots T - \tau_i - 3$. The soonest a packet from node S_i can arrive at the sink is $\tau_i + 3$ time units after node S_i starts to capture it. This is the minimum time for the AUV to go to the surface and then a unit each for capture, transmission from S_i to the AUV and for the AUV to transmit the packet to the sink. The final time to transmit a packet is $T - 1$, for arrival at time T. There is no reason to model packets that are captured too late to be delivered. Given a node S_i and a capture time $t_1 \in \mathcal{T}_{i1}$, the set of possible delivery times are $\mathcal{T}_{i12} = \max\{t_1 + 1, \tau_i\} + \tau_i + 2 \ldots T$. The earliest a packet can be collected by the AUV is τ_i, since the AUV starts on the surface and must travel to the node. So, packets captured before this time have the same minimum delivery time as the packet captured at time $\tau_i - 1$. Any packet captured after time $\tau_i - 1$ can be delivered in the minimum time after capture, which is one unit to transmit the data to the AUV plus the time for the AUV to travel to the surface plus one time unit for transmission.

Let F be the set of all feasible tuples (i, t_1, t_2) for a node S_i, capture time $t_1 \in \mathcal{T}_{i1}$ for a packet at S_i and delivery time $t_2 \in \mathcal{T}_{i12}$ for the packet captured at time t_1. Let $F(t_2) = \{(i, t_1) | (i, t_1, t_2) \in F\}$ be the set of packet (sensor source, capture-time) pairs that can be legally delivered at time t_2. Let also $F(i, t_2) = \{t_1 | (i, t_1, t_2) \in F\}$ the set of data capture times t_1 of packets from node S_i that can be delivered at time t_2. For any $n \in N \cup \{\omega\}$, let $L(n)$ be the times where it would make sense for the AUV to be at location n. For $n \in S$, we have $L(n) = \tau_n \ldots T - \tau_n - 1$. For $n \in W$, we have $L(n) = 0 \ldots T$. Finally, $L(\omega) = 1 \ldots T - 1$.

The set $C(\tau, i, t)$, i.e., the set of all pairs location-time (j, τ) such that the AUV cannot be at position j at time τ if at time t it is at location i, is defined in detail as follows: We first define $G(i, t) = \{(j, \tau) \mid \forall j \in N \setminus i,\ \tau \in [max(t - d_{ij} + 1, 0)...t]\}$ to be the set of all pairs location-time (j, τ) where and when the AUV cannot be if it is at location i at time t. We then define $\tau_{it}^{min} (\tau_{it}^{max})$ as the minimum (maximum) time in the set $G(i, t)$. Finally, the set $C(\tau, i, t)$ is defined as $C(\tau, G(i, t)) = \{(j, \tau) \mid (j, \tau) \in G(i, t)\}$ as the set of all pairs location-time (j, τ) such that the AUV cannot be at position j at time τ if at time t it is at location i. The set T^* of appropriate values for τ^* as needed in Constraints (4) is defined as the interval $[\tau_{it}^{min} \ldots \tau_{it}^{max}]$.