A Cluster-Based Mobile Data-Gathering Scheme for Underwater Sensor Networks

Seyed Mohammad Ghoreyshi, Alireza Shahrabi, and Tuleen Boutaleb
School of Engineering and Built Environment
Glasgow Caledonian University, Glasgow, UK
{Seyed.MohammadGhoreyshi, A.Shahrabi, T.Boutaleb}@gcu.ac.uk

Abstract—In this paper, a new distributed Cluster-based Mobile Data-Gathering scheme (CMDG) for large-scale Underwater Sensor Networks (UWSNs) is proposed to improve the system performance in terms of routing scalability, energy saving, and data gathering latency. In this scheme, a subset of underwater sensors is selected as cluster heads to collect data from affiliated sensors and transmit the data to an Autonomous Underwater Vehicle (AUV). Then, the AUV tour is planned such that all cluster heads are visited while shortening the tour length of the AUV. Using extensive simulation study, we analyse the performance of CMDG and show that it can effectively reduce the tour length while prolonging the network lifetime compared to another existing mobile data-gathering approach.

I. INTRODUCTION

In recent years, underwater sensor networks have emerged as a new information-gathering paradigm in a wide-range of aquatic applications, such as environmental monitoring, battlefield surveillance, pollution monitoring, etc [1], [2]. Underwater sensors are usually distributed in a large-scale marine environment to collect data and transfer them to a destination, which may be a static sink, or a mobile sink (e.g. AUV) [1], [3], [4].

Underwater sensor nodes use the acoustic transmission due to restrictions on the use of radio waves in the underwater environment. However, to cope with the path loss and high bit error rate in underwater, sensors need to transfer data packets with higher transmission power which makes the energy saving as a critical issue [5], [6]. In a static sink model, underwater sensors close to the sink consume much more energy than other sensors because of relaying more data packets. Thus, these node may fail sooner and network become disconnected. This problem can be exacerbated in large-scale UWSNs. However, data-gathering using AUV is more suitable for large-scale networks due to reducing the number of transmissions and balancing the energy consumption [3]. AUV is a mobile sink equipped with a powerful transceiver, moving through the underwater area and continuously collects data packets from sensor nodes [7].

The data-gathering tour is periodically initiated from a static base station, followed by collecting data packets from sensor nodes, and completed by transferring data packets to the static base station [8]. Using a mobile sink contributes to prolong the lifetime of sensors since any packet relay is bounded within a given number of hops. In AUV tour planning, there is a trade-off between energy saving and data gathering latency. If AUV traverses the transmission range of each sensor directly to collect data without any relay, the maximum energy saving can be achieved. However, data gathering latency is increased due to increase in the tour length. Thus, it is more appropriate to decrease the data gathering latency by performing the local aggregation in a subset of nodes as cluster heads and transferring the aggregated data to AUV. It should be noted that the local data aggregation should be bounded to few number of transmission hops to cope with increased energy consumption and packet loss.

In this paper, a new distributed Cluster-based Mobile Data-Gathering scheme (CMDG) is presented to minimise tour length, energy consumption, and latency. In this scheme, a subset of underwater sensors is selected as cluster heads to collect data from affiliated sensors and transfer the data to AUV when it arrives. Furthermore, by affiliating sensors with cluster heads, any packet relay is bounded within a few number of hops which decreases the chance of collisions and packet loss. Limiting multi-hop relay to a particular level also reduces the energy consumption at sensor. In CMDG, underwater sensors compete to be a cluster head based on their priority in a distributed manner. The effectiveness of CMDG is verified by comparing with another existing mobile data-gathering approach.

The rest of this paper is organized as follows. In Section II, we review the related work on mobile data gathering in UWSNs. In Section III, we provide a detailed description of the system model. In Section IV, CMDG is presented in details. Section V gives simulation results and some discussions. In Section VI, we conclude the paper.

II. RELATED WORK

During these years, some mobile data-gathering techniques have specifically been proposed for UWSNs.

In AEERP (AUV aided Energy Efficient Routing Protocol) [7], AUV traverses a predetermined elliptical trajectory in each cycle. The sensors are divided into two categories: gateways and members. The gateway sensors only can communicate with an AUV, and they are selected based on the nearness to AUV trajectory and their remaining energy. Members are then allocated to the gateways sensors using a Shortest Path Tree (SPT). However, there is no bound on the hop distance from members to a gateway sensor which causes an increase in energy consumption in a broader network.
Mobicast [9] is a mobile data-gathering protocol in which AUV traverses a predetermined circle path to collect data packets from sensors in different geographic regions called 3-D zone of references (3-D ZOR). In Mobicast, there is no use of clustering mechanism and sensors should relay the data packets to AUV in a single hop or multi-hop transmissions. To deal with the presence of various water currents and void areas, a larger covering area surrounding the 3-D ZOR is considered to find alternative paths to deliver the packet to all sensors. This covering area involves more sensors resulting in more energy consumption. Furthermore, collecting data from all sensors in the sensing field is not possible because only sensors within 3-D ZORs can be investigated.

In AURP (AUV-aided underwater routing protocol) [8], multiple AUVs are used as a relay to collect data packets from gateway nodes and then forward to a sink. Thus, the sink and gateways should periodically broadcast their interest in receiving data to be used by sensors for choosing the next hop such that the path length is minimised. In AURP, it is mentioned that gateways and the trajectory of AUVs can be determined dynamically or before deployment; however, this procedure and the resulting overhead has not been investigated thoroughly. Moreover, an AUV trajectory is a fixed elliptical path which reduces the flexibility in confront of different sensor deployment strategies.

AUV_PN [3] is a mobile data-gathering protocol in which underwater sensors are clustered around several CHs while an AUV is employed to visit some identified locations, called Path Nodes (PNs), to collect the aggregated data. During data-gathering phase, AUV travels to each CH to obtain the list of PNs and then visits each PN to collect data. However, there are some constraints which can confine the AUV_PN performance. The AUV tour is not optimal, and it crosses over itself. By broadening the network size, sensors should transmit the data packets with higher power because they are placed farther away from PNs. The network-partitioning has a complicated procedure, and it is performed with substantial overhead in energy and communication.

III. SYSTEM MODEL

The network architecture and acoustic propagation model are described in this section. Fig. 1 shows different components of a typical mobile data-gathering UWSN widely used in the literature.

A. Network architecture

In this work, we consider a UWSN consisted of underwater sensors, a single data sink, and a mobile data sink (AUV) [3], [9], [10]. Underwater sensors are distributed in a two-dimensional plane with a fixed depth. Each sensor can control its depth using a pressure gauge and fish-like bladder apparatus [11], [12]. It is assumed that sensors are static or anchored to the bottom of the ocean. The static sink is located on the water surface, which can communicate with AUV and monitoring centre using an acoustic and radio modem, respectively. The AUV operates at a fixed depth above the underwater sensors, and it has the freedom to move in all directions.

Underwater sensors are not required to know their full geographical coordinates; however, AUV can obtain sensors coordinates by marking the locations where it receives data from them. All sensors are homogeneous regarding transmission range and power. Each sensor can measure its distance to other neighbouring sensors by using Received Signal Strength Indicator (RSSI) [11].

B. Acoustic propagation model

In this section, we discuss the channel model for acoustic communication. The Thorp model [13] is used for describing the underwater acoustic channel model. The path loss or acoustic channel attenuation over distance \( d \) can be represented as [13]:

\[
A(d, f) = A_0 d^k \alpha(f)^d
\]  

where \( f \) is the signal frequency and \( \alpha(f) \) is the absorption coefficient which is determined by the Thorp model. Furthermore, \( A_0 \) denotes a unit-normalizing constant, and \( k \) is the geometric spreading factor which is set to 1.5 for practical scenarios. The underwater noises are dominant in the different frequency regions and are composed of four main components of turbulence \( PN_t(f) \), shipping \( PN_s(f) \), waves \( PN_w(f) \) and thermal energy \( PN_{th}(f) \) which can be represented as [5]:

\[
PN(f) = PN_t(f) + PN_s(f) + PN_w(f) + PN_{th}(f)
\]

The signal-to-noise ratio (SNR) over distance \( d \) with the signal frequency \( f \) can be expressed as [13]:

\[
SNR(d, f) = \frac{PR(f)}{A(d, f)PN(f)}
\]

where \( PR(f) \) denotes the transmission power with frequency \( f \) at the forwarding node. To receive the data packet without any error, \( SNR \) at the receiver should be greater than the detection threshold.

The bit error probability over distance \( d \) can be computed by [11]

\[
P_e(d) = \frac{1}{2} \left(1 - \sqrt{\frac{SNR_{avg}(d, f)}{1 + SNR_{avg}(d, f)}}\right)
\]

where \( SNR_{avg}(d, f) \) is the average signal-to-noise ratio over distance \( d \) with frequency \( f \) which can be calculated using Eq.
3. Accordingly, the delivery probability of a data packet with size $n$ bits over distance $d$ can be expressed as follows [11]:

$$P = (1 - P_e(d))^n$$  \(5\)

IV. CMDG DETAILS

This section presents CMDG scheme in detail.

A. Overview

In CMDG, as a cluster-based scheme, our primary goal is to find a subset of sensors as Cluster Heads (CHs) to cover affiliated sensors within limited number of hops and then to find a short tour for AUV to visit each CH in a particular sequence. Each cluster head can buffer all collected data from its cluster members and upload them to AUV when it arrived within a single-hop vicinity of the CH. In order to perform this task, our scheme is divided into three phases: updating phase, clustering phase, and data gathering phase.

The data-gathering phase has the largest share of the network energy consumption [3]. However, the performance of this phase is mostly dependent on the efficiency of clustering phase when nodes are clustered around several cluster heads. During the initial phases of updating and clustering, underwater sensors should discover neighbouring nodes and organise themselves into different clusters. The number of transmission hops that connects nodes to the cluster heads has a crucial role in the energy consumption [14].

In UWSNs, relay hop count should be bounded due to a number of reasons. First, underwater environment is very noisy which can increase the chance of packet failure by forwarding a packet over several hops [1]. Second, energy efficiency can be achieved by limiting the number of packet transmissions. Third, there is a limitation on the sensor buffering capacity. Thus, it is not practical to allocate a high number of sensors to a CH for local data aggregation.

The relay hop bound, $d$, is a system parameter which can be set based on the application priorities on the energy saving and delay. For delay-tolerant applications, $d$ is set to a small value to save more energy at sensors. Mobile data-gathering protocols are usually suitable for the applications which are almost delay-insensitive.

During the data-gathering phase, AUV continuously starts its tour from a static sink, which can be placed anywhere on the surface, collects data packets from CHs and then returns to the static sink to forward all gathered data.

B. Updating phase

In updating phase, neighbouring information should be exchanged between sensors to be used during the clustering phase [15], [16]. Each sensor is required to obtain and maintain $d$-hop neighbouring information. To this end, $d$ rounds of neighbouring information exchange should be performed. Each sensor has a neighbouring table to maintain the IDs of neighbouring sensors within its $d$-hop range.

In the first round, each underwater sensor broadcasts a control packet including the packet type and sensor ID. Upon receiving a control packet, each receiving sensor updates its neighbouring table based on the newly discovered sensor. Each receiving sensor also measures its relative distance to the sending sensor via the difference between the initial and received signals strengths and keeps it at the neighbouring table. The initial signal strength is known to each node as all the nodes are homogeneous in terms of the transmission power [11]. If $d = 1$, this procedure is completed, otherwise it should continue until all neighbouring information from $d$-hop is obtained.

In the following rounds, in addition to the packet type and sensor ID, each underwater sensor also broadcasts its neighbouring table and measured distances. Then, each receiving sensor updates its neighbouring table with newly discovered sensor IDs and its distance to each of them. At the end of $d$ rounds of neighbouring information exchange, each sensor is able to calculate its $d$-hop degree which is the number of sensors in its $d$-hop range.

C. Clustering phase

In this section, we propose a new distributed algorithm to determine CHs and their members in order to find a short tour to visit all CHs. The clustering phase only needs to be repeated when the network topology changes [17].

In addition to relay-hop bound, $d$, choosing an appropriate CH has also a direct impact on the length of the tour. A lower number of clusters with higher density, but still with $d$ relay-hop, leads to having a lower number of CHs and consequently a shorter tour for AUV to traverse every time. Therefore, our clustering algorithm is based on this criteria and uses $d$-hop degree to determine a CH. Needless to say, this model does not need to have any localisation information and neither the knowledge about the distance to the surface sink.

Algorithm 1 details the clustering phase. The basic idea is that each sensor competes with other sensors to become a CH based on its $d$-hop degree. A sensor with the higher $d$-hop degree has greater chance to be a CH than others. After determining the CHs, each none-CH sensor may locate a set of CHs in its $d$-hop range. However, it should join a cluster in which its head has the shortest $d$-hop distance, to break the tie.

In our scheme, CHs are selected in a timer-based approach. In this way, each sensor can set a delay timer upon starting the clustering phase. The delay timer has inverse proportion relation with the $d$-hop degree of each node. For the sensors with fewer neighbours within $d$-hop range, they should be delayed for a longer time. The delay timer for each sensor can be computed by

$$t = (\Delta_d - \delta_d) \times (DP/\Delta_d) \pm \lambda$$  \(6\)

where $\delta_d$ indicates the sensor $d$-hop degree and $\Delta_d$ is the maximum $d$-hop degree limit for a network topology. $\Delta_d$ can be estimated based on the sensor deployment strategy, number of sensors, and network area size and it can be known to all sensors during the deployment time. $DP$ is the clustering time interval and $\lambda$ is a short random time duration to differentiate the underwater sensors with the same $\delta_d$. 

Algorithm 1 Distributed Clustering Algorithm

<table>
<thead>
<tr>
<th>Line</th>
<th>Instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><code>procedure Clustering(sensor)</code></td>
</tr>
<tr>
<td>2</td>
<td>Set a delay timer $t = (\Delta_d - \delta_d) \times (DP/\Delta_d) \pm \lambda$</td>
</tr>
<tr>
<td>3</td>
<td><strong>while</strong> the delay timer is not expired <strong>do</strong></td>
</tr>
</tbody>
</table>
| 4    | Listen and maintain the CHs
| 5    | If $CH\_msg$ has been forwarded for less than $d$-hop then |
| 6    | Forward the $CH\_msg$
| 7    | **end while** |
| 8    | If never received any $CH\_msg$ then |
| 9    | `status(sensor) ← CH`
| 10   | Broadcast a $CH\_msg$
| 11   | **else** |
| 12   | `Keep listening and maintaining the CHs`
| 13   | **end if** |
| 14   | If DP is over and `status(sensor) ≠ CH` then |
| 15   | `status(sensor) ← non_CH`
| 16   | $CH(sensor) ←$ the closest CH among accessible CHs
| 17   | Transmit a join_msg
| 18   | **end if** |
| 19   | **end procedure** |

During the listening period, each sensor only listens and receives the CH announcements from the sensors with higher priority and receiving sensors also forward the announcement to their neighbouring sensors if it has been forwarded for less than $d$ hops. After expiring the delay timer, a sensor announces itself as a CH by broadcasting a CH announcement if no other CH announcement received from other $d$-hop neighbouring sensors during the listening period. This announcement should then be relayed by all neighbouring sensors to up to $d$ hops neighbourhood.

At the end of the clustering time interval, a non-CH sensor should join a CH with the closest distance. After finding a suitable CH, sensor transmits a joining message to the selected CH. By using this approach, all sensors are locally clustered and the packet relay is bounded to $d$-hop.

D. Data gathering phase

After the clustering phase, AUV should discover all selected CHs. Thus, AUV needs to travel the entire sensing field to find CHs and to mark their locations. While exploring, AUV can broadcast control packets periodically to discover the CHs. When a CH receives a control packet from AUV without any error, it should respond to AUV using an ACK message. AUV then marks its current location and links it with the ID of CH.

After the exploring of all CHs, AUV returns to the static sink to upload the CHs list. In the next step, the main objective is to find a short path passes through all CHs. Finding such a path is considered as a Travelling Salesman Problem (TSP) which belongs to the class of NP-complete problems [18]. We therefore use a greedy heuristic approach offering comparably fast running time and still yielding near-optimal solutions [18].

Following this approach, if the number of CHs is $N$, a tour is gradually constructed by repeatedly adding the shortest edge to the tour as long as there is no cycle with less than $N$ edges, or no CH with a degree more than 2. The greedy algorithm can be summarised as follow:

1. All edges should be sorted in increasing order of length.
2. The shortest edge is included in the tour if
   - No early cycle is formed, and
   - No vertices has a degree of 3
3. Does tour include $N$ edges? If no, step 2 is repeated.

The complexity of this greedy algorithm to create a path is given as $O(n^2 \log_2(n))$ [18].

Once a tour is generated by the greedy algorithm, it can be optimised using some heuristic techniques. We use 2-opt algorithm which is a basic local search algorithm to take a route that crosses over itself and convert it to a tour without any crossed line [18]. The main idea is to incrementally improve an initial tour by removing two edges from the tour, reconnecting the two paths created, and replacing the current tour with new tour if it decreases the length of the tour. This procedure is continued by swapping all possible pairs of edges in the tour until no 2-opt improvements can be found.

After the tour planning, AUV initiates the data gathering phase. It should visit each CH based on the planned tour to collect the data and return to the static sink to upload the aggregated data. Meanwhile, each sensor monitors the environment and sends data packets to its CH with a fixed data rate. Transmitting data by sensors to their CHs is independent of AUV activity. In this way, data packets are buffered in CHs and are transferred to AUV when it arrives.

When AUV gets close to a CH, it announces its arrival using a control packet. Then, CH starts transmitting the collected data to AUV. After collecting data from a CH, AUV travels to the next CH and repeats the same procedure until it returns to the static sink. It then initiates the next round of data gathering in a similar way.

V. EXPERIMENTAL RESULTS

In this section, the details of our simulation study and also the performance results are presented, to evaluate the proposed algorithm against another existing mobile data gathering scheme, $AUV\_PN$ [3], in an underwater environment.

A. Simulation setup

The underwater acoustic communication channel described in Section III is used in our simulation. We deploy the sensors (ranging from 100 to 500) randomly in a two-dimensional plane $1000m \times 1000m$ at a depth of 300 m of a 3D underwater environment. The transmission power and the power threshold for receiving a packet are set to 105 $dB$ re $\mu$ Pa and 10 $dB$ re $\mu$ Pa, respectively. Each sensor consumes 50 $W$ and 0.158 $W$ energy for sending and receiving a packet, respectively, while the idle power consumption is 0.008 $W$. The signal frequency $f_{data}$ is set to 20 $kHz$ and the acoustic signal propagation speed is 1500 m/s. The transmission range of each sensor is considered as 100 meters.

Each sensor generates a data packet every 100 seconds. The bit rate is set to 10 $kbps$, and the data packet size is fixed at 1024 bits. The static sink is placed at the corner of the network.
B. Results and Analysis

In this section, we assess the performance of our scheme against another existing distributed data-gathering protocol AUV_PN in terms of tour length, number of cluster heads, end-to-end delay, and energy tax.

Tour length: It is defined as the total travelled distance by AUV in each data gathering round.

Number of cluster heads: It indicates the number of CHs which should be visited by AUV during the data gathering phase.

End-to-end delay: It is defined as the average delay time taken from the moment of the creation of packets at sensors until successfully being delivered to the static sink.

Energy tax: The energy tax shows the average energy consumed per message to deliver a packet to the static sink successfully.

1) The impact of sensor density: In this set of simulations, the impact of sensor density on performance metrics are examined. We keep other parameters fixed and change the number of sensors from 100 to 500. The results for the tour length, the number of cluster heads, end-to-end delay, and energy tax are plotted in Fig. 2a, Fig. 2b, Fig. 2c, and Fig. 2d respectively.

From Fig. 2a, we can see that the tour length of CMDG (d=1) increases as the number of sensors increases. It is because that each CH can only cover the sensors with a one-hop relay. Thus, by increasing the number of sensors, higher number of CHs and larger tour length to visit those CHs is required. However, the tour length of CMDG (d=2) and AUV_PN are not correlated with sensor density. As shown in Fig. 2b, the number of CHs in AUV_PN is fixed and less than CMDG (d=2), but it still has higher tour length. This is because, in AUV_PN, AUV starts the tour from the static sink and travels to the closest CH to obtain the list of Path Nodes (PNs) and then visits them to collect data packets; however, it does not guarantee a shorter tour. The constructed tour by AUV_PN is a tour with so many crossed lines which increases the tour length. In CMDG (d=2), any further increase in the number of sensors has little impact on the number of CHs and selection of the preferred CHs. The tour is also constructed by a greedy approach and optimised by a 2-opt algorithm before the AUV travelling.

As shown in Fig. 2c, there is a direct relationship between the tour length and end-to-end delay. In CMDG (d=2), data packets are delivered with the lowest end-to-end delay. It is because AUV travels a shorter path to complete its data-gathering cycle, so data packets are held for a shorter time in CHs.
Fig. 2d shows the average energy consumed per message in each approach. It is observed that the energy consumption of CMDG (d=1) is considerably less than others. It is because the number of transmissions is significantly reduced by bounding the relay hop to lower values.

In addition to having a shorter tour, CMDG (d=2) consumes less energy than AUV_PN. This is because, in AUV_PN, the number of cluster heads is less than CMDG (d=2). Thus, sensors are placed in a longer distance to CHs which require them to send their packets with a higher power. The larger number of CHs results in a shorter distance between sensors and CHs, and therefore saving more energy. Moreover, in AUV_PN, some CHs do not participate in data collecting, and AUV only visits them to obtain the list of PNs. Thus, the number of CHs which participate in local data collecting is less than the actual value.

2) The impact of network size: In this set of simulations, we explore how the network size impacts the performance of protocols. Let SL be the side length of the network area. We fix the number of sensors at 200, and vary SL from 500 m to 2000 m. Fig. 3a, Fig. 3b, Fig. 3c, and Fig. 3d plot the results for the tour length, the number of cluster heads, end-to-end delay, and energy tax respectively.

From Fig. 3a, we can observe that the tour length for all approaches becomes longer as SL increases. This is because sensors are more sparsely deployed in a wider area so AUV should travel a longer route to visit all CHs. In a broader distributed area, the tour length of AUV_PN is less than others because the number of CHs is fixed (as shown in Fig. 3b) and only the geographical distances between CHs are increased. In CMDG, the number of CHs increases to cover more positions to collect data from all the sensors by keeping the relay hop bound limitation. Thus, it is reasonable that the tour length becomes longer as an expense while energy saving is highly improved as shown in Fig. 3d. The direct relationship between the delay in receiving packets and the tour length is evident from Fig. 3c.

Fig. 3d shows the energy consumption of protocols as a function of SL. The energy consumed by AUV_PN is highly increased as SL increases. This is because, by increasing the network size, the number of CHs is still fixed while many sensors are placed further away from the CHs. Thus, a sensor needs to transmit its packet by consuming more power to cope with the path loss in a longer distance. On the other hand, as SL increases, energy consumption decreases in CMDG. This is because the same number of sensors are distributed in a wider area resulted in a sparse density network. Thus, the number of CHs increases while the hop count distances between sensors and CHs is reduced, and more energy saving can be achieved.

VI. CONCLUSION

In this paper, we proposed a mobile data-gathering scheme for UWSNs by exploiting a trade-off between energy and data gathering latency. In the proposed scheme, a group of sensors are selected as CHs in a distributed manner to collect data locally from their members. An optimal tour is then planned by AUV to visit all those CHs to gather data packets and upload them to a static sink on the surface. The simulation results illustrated that CMDG can shorten the tour length while maintaining the relay hop bound, resulting in a better trade-off between energy and data-gathering latency.

REFERENCES