Journal of King Saud University – Computer and Information Sciences 31 (2019) 275–286

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

Data aggregation in underwater wireless sensor network: Recent approaches and issues

Nitin Goyal [a,*], Mayank Dave [a], Anil K. Verma [b]

[a] Dept. of Computer Engineering, National Institute of Technology, Kurukshetra, Haryana, India
[b] Dept. of Computer Science and Engineering, Thapar University, Patiala, Punjab, India

ARTICLE INFO

Article history:
Received 13 February 2017
Accepted 29 April 2017
Available online 2 May 2017

Keywords:
Review
Survey
Future challenges
Data aggregation
UWSN

ABSTRACT

Underwater Wireless Sensor Network (UWSN) technology is widely used in various underwater monitoring and exploration applications and has proven its high stature. Since many years various UWSN protocols have been designed or existing protocols are improvised for effective and qualitative research analysis. The data aggregation is one of the schemes that is widely been used along with UWSN protocols to achieve better results. Thus it is foremost required to present a periodical review on data aggregation. Herein we present a paper based on the survey of UWSN with data aggregation to highlight its benefits and limitations. Ambition behind this paper is to build interest of research fraternity towards future challenges identified on the basis of survey of existing approaches. The existing techniques that aggregate data are divided into cluster based, non-cluster based and other approaches. The existing techniques are analysed along with their advantages and disadvantages. Moreover, the performances of K-means, Distributed underwater clustering scheme and Round-based clustering approaches are compared in terms of delay, packet drop and energy consumption, with and without aggregation. Also the performance of Receiver Oriented Sleep Scheduling, Intra and Inter Cluster Communication and Energy Efficient Distributed Time Synchronization techniques are compared.

Contents

* Corresponding author.
E-mail addresses: er.nitin29@gmail.com (N. Goyal), mdave@nitkkr.ac.in (M. Dave), akverma@thapar.edu (A.K. Verma).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2017.04.007
1319-1578/

1. Introduction

The vast research has already been done on terrestrial sensor networks in many aspects and nowadays the researchers are attracted towards Underwater Wireless Sensor Networks (UWSNs) being a new area for research. UWSN differs from generally used land based sensor networks in terms of communication method of acoustic signal, cost of expensive sensors, greater memory space to save maximum data, maximum power for communication, and for dense deployment of sensors (Gholami et al., 2015; Jadidoleslamy et al., 2016; Rezvani et al., 2015; Rahman et al., 2016; Das and Thampi, 2017; Goyal and Kumar, 2014). UWSNs consist of stationary as well as movable nodes that communicate through acoustic channel (Goyal et al., 2016; Han et al., 2016). UWSN is used in many applications having underwater environment like pollution monitoring especially chemical waste, monitoring of the population of underwater flora, fauna, the examining of the health of rare marine creatures, mine reconnaissance, disaster prevention, assisted navigation, nutrient production, oil leakage detection, distributed tactical surveillance, oceanographic data collection, target detection, tracking and underwater military applications (Izadi et al., 2015; Jia and Meng, 2016; Karimi et al., 2015). The researchers have to face many challenges while working in UWSN like narrow bandwidth, shadow zones, long propagation delay, harsh geographical atmosphere, attenuation, comparatively smaller network scale, high bit error rates, limited energy and temporary losses of connectivity (Vennila and Madhura, 2016; Coutinho et al., 2015).

In UWSN sensor nodes are deployed in a network which collect information or data and transmit it to the sink. In different aquatic mediums such as rivers, lakes, ponds etc. the UWSN protocols are being used for monitoring purpose. However these protocols may not give the effective information in case of oceans, sea etc. as these are larger in area and there are numerous factors like area to be affected, level of water, pressure etc. which are required to be analysed at real time. To overcome this situation the technique of data aggregation is being used as a supplementary technique along with routing protocols where larger information is collected and aggregated at a node before its transmission to sink. The positive and the effective results have revealed that the use of data aggregation in the network has reduced the data redundancy and saved the energy consumption along with data transmission with minimum delay (Curiac, 2016; Xu et al., 2015).

In the recent past number of disasters have occurred like floods in Vietnam, Spain, earthquake in Nepal, Verdah Cyclone in Tamilnadu India and many others that have shown the need or presence of proper disaster management. For effective disaster management it is foremost required for any responder to have complete information about the affected area. In case of disasters occurred due to earthquake, volcanic eruption, tsunami, or other seismic activities the affected area is so large that it becomes cumbersome to collect and store the information data in readily available manner (Senel et al., 2015; Goyal et al., 2017). Various data collection techniques are already devised to gather the information for later use. However, due to inadequate storage capacity or power supply the collected information may not be used timely. Thus the process of data aggregation is adopted that suppresses the size of the collected information so that it uses lesser space while storing. The data aggregation also reduces the number of data packets thus the packet drop ratio in the network gets reduced. In the hierarchical network, data aggregation has to be done at designated nodes for reducing the number of data packets transmitted towards the sink and improving the energy efficiency (Zenia et al., 2016).

An underwater network is generally constituted of various independent sensor nodes that aggregates data and employs forward operations to send the aggregated data to a node designated as sink node. The utmost challenge behind organizing such network are operational energy, cost, memory, communication range and limited lifetime of any individual sensor. Due to power loss in UWSN the sensor nodes stop transmitting the data which further increases with the passage of time. Thus in the same acoustic medium the coverage area of the working nodes diminishes. It remains important for researchers to have a longer lifetime of the network along without degrading performance. Researchers have proposed data aggregation technique to design energy efficient algorithms for UWSN, where aggregator node accumulates the identified data from neighbouring nodes, process and transmit it to sink (Liu et al., 2015). Further researchers have proposed to maintain data accuracy with minimum data redundancy during data aggregation in UWSN.

2. Existing survey on data aggregation in UWSN

In UWSN many supplementary techniques have been proposed like data integration, data collection, data aggregation, data dissemination, and data fusion etc. Although each supplementary techniques has its own value but the process of data aggregation holds the most important position and it is the necessity of time to have periodic reviews on data aggregation techniques. Researchers (Ayaz et al., 2011; Goyal et al., 2014; Shen and Bai, 2016; Kafetzoglou et al., 2008; Kumar and Singh, 2014) have presented a review article to identify the importance of data forwarding, node

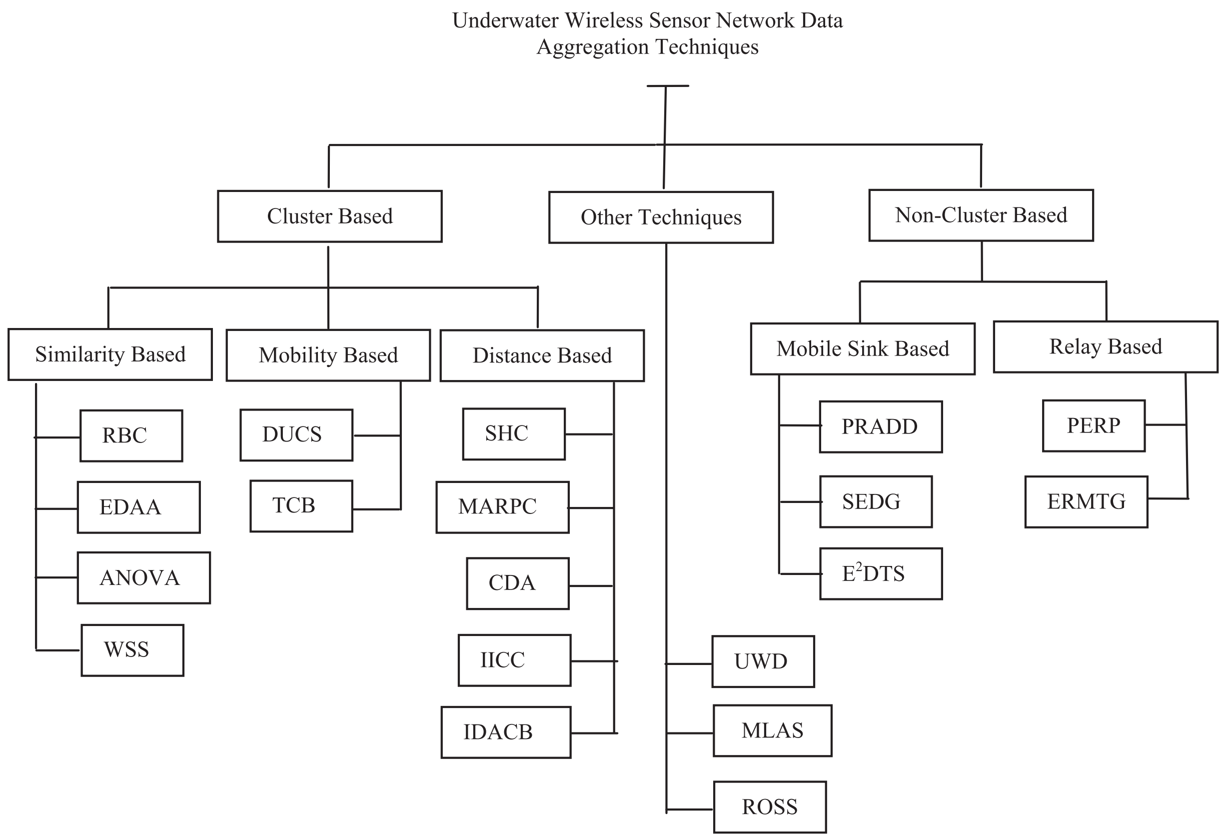

Fig. 1. Classification of UWSN data aggregation techniques.

deployment and node localization in UWSNs under variable conditions which were further segregated in accordance to their characteristics and functions.

Oh and Tran (2013) have determined four similarity functions appropriate to UWSN data aggregation by evaluating and comparing them. The experimental results show that unique characteristics of UWSN, Euclidean or cosine distance is more effective in UWSN, whereas Hamming and Jaccard distances are appropriate for applications that work on exactly the same captured data. But except Kumar and Singh (2014), all other works have hardly focused on data aggregation directly. In Kumar and Singh (2014), existing aggregation technique of UWSN and WSN are discussed. Moreover it does not perform the survey based on any classification. However our manuscript classifies the work on data aggregation in UWSN into various categories and exclusively focuses on UWSN only.

The techniques of data forwarding, deployment, localization and data collection are used to transfer the data from node to sink in UWSN. To achieve better results out of these schemes a method of data aggregation was explored to be used as a supplement. It accumulates data in an energy efficient manner by reducing data redundancy and simultaneously enhancing the accuracy and lifetime of the network.

1. The techniques invented under data aggregation in UWSN are classified into cluster based and non-cluster based techniques that are not done in earlier survey papers. Apart from these two, there exists some other techniques that indirectly apply aggregation by means of diffusion and scheduling is classified as ‘other techniques’.
2. Various UWSN techniques with and without aggregation based on QoS parameters i.e. delay, packet drop, and energy consumption w.r.t. time are compared. The protocol with better performance is also suggested.
3. The performance evaluation of all three data aggregation categories (cluster based, non-cluster based, and other technique) on the basis of parameters that are delay, delivery ratio, and energy consumption w.r.t. varying packet size.
4. The classification of cluster based, non-cluster based and other techniques based on the aggregation method is shown in the Fig. 1 below.

3. Classification and State-of-Art

This section presents the overview of the UWSN data aggregation techniques as shown in the classification above.

3.1. Cluster based techniques

The cluster-based concept splits the network into sets of nodes called as clusters and states a method that connects all clusters to each other. The cluster based arrangement creates a concise and stable network. Moreover, it is a challenging phenomenon to reduce network’s energy consumption, which has recently gained attention in UWSN.

3.1.1. Similarity function based

As the name suggests the similarity based clustering involve the pairing or the group of objects with similar functions. In a way this minimizes the data redundancy however it remains unknown that under which conditions the selected similarity function produces the required form of cluster. Later on it has been seen that similarity function based clustering may increase data latency in a cluster. Following are the schemes that are based on this function.

3.1.1.1. Round based clustering (RBC). Tran et al. (2014) have designed a technique named as round-based clustering (RBC). It performs in rounds which constitute four phases that are initialization, cluster-head selection, clustering and data aggregation phase. In the initialization phase, the sensor nodes and the sink node are deployed in the network. Here sink node initiates a round by setting up its time. In the second phase there occurs selection of cluster-heads. During this phase, information about residual energy, position, and distance to the BS/sink node is gathered. In the clustering phase, clusters are formed for each cluster-head and its members. Then, in the last phase, data is aggregated and transmitted to the sink node by cluster-heads. In re-clustering phase, the clusters are reconstructed and the cluster-heads are reselected whenever any changes occur in network conditions due to energy consumption, network movement etc. This process of re-clustering prolongs the network lifetime. This technique saves energy consumption thereby enhancing throughput of the network.

3.1.1.2. Efficient data aggregation approach (EDAA). Here, at a certain time, the aggregator node accumulates the measured data grouped as a vector. Further the same aggregator node start identifying the pair of grouped data in which there occurs similarities above a notified value of threshold. If the compared data are considered similar to each other, it is not necessary for the aggregator to transmit all sets of data to the sink node. The similarity functions that are generally noticed are Euclidean distance, Edit distance, Cosine similarity, and Jaccard similarity etc. This application of similarity function saves energy consumption and reduces data packet size. Tran et al. (2014) have proven the effectiveness of similarity functions by lowering the size of packet also reducing duplication of data in cluster-based UWSNs.

3.1.1.3. K-Means and ANOVA based. Harb et al. (2015) have presented a clustering technique based on similar node reading function. They assumed that data is sent from node to cluster head (CH) in the form of readings. The scheme constitutes two levels of data aggregation. . Firstly, to reduce data duplication, the reading at each node is cleaned periodically, before sending its data set to its CH. After data transmission, K-means algorithm based on a one-way ANOVA model is applied to notify the nodes that generate similar sets of data thereby aggregating the similar sets and transmit them to sink. Saranya and Arthi (2016) have proposed a new clustering method used to handle similarity between node readings which are sent periodically to the cluster-head. A two tier data aggregation technique is proposed in which at first level node eliminate data redundancies and at second level identify nodes generating identical data sets and aggregated them before sending to sink. In general the readings are sent periodically from sensor nodes to their appropriate CHs. They also used distributed energy-efficient clustering algorithm for efficient communication between cluster head and base station.

3.1.1.4. Well-suited similarity functions (WSSF). Tran et al. (2013) have presented a cluster based UWSN approach. It works on the basis of similarity functions that are Euclidean distance and cosine distance and shown their importance by cutting down the size of data packet and reducing the data duplicacy in the network. Initially the similarity functions are determined and then applied along with data aggregation approach in order to enhance the network lifetime.

3.1.2. Dynamic or mobility based techniques

In UWSN as the nodes move along with water current thus an underlying routing protocol is required for nodes to move. The existing routing protocols on UWSN assume random node mobility. Hence Clustering techniques are designed where nodes change their position along the time. Below are the schemes mentioned that are based on this function.

3.1.2.1. Distributed underwater clustering scheme (DUCS). Domingo et al. (2007) have presented DUCS (Distributed Underwater Clustering Scheme), where random mobility of nodes is assumed and timing was adjusted continually to reduce loss of data. It uses GPS-free routing protocol without using any flooding technique. It minimizes the proactive routing message exchange. Further data aggregation is used to reduce data duplication. The scheme also reduces high propagation delay in the aquatic medium along with better communication, using a continually adjusted timing advance combined with guard time values.

3.1.2.2. Temporary cluster based routing (TCBR). Ayaz et al. (2010) have proposed a protocol suitable for even hybrid networks apart from the stationary and mobile networks. Here preference is given to shortest path while sending data and smaller number of nodes are engaged in the process of end-to-end routing. Moreover, it does not require any location information of sensor nodes. In this way, the same amount of energy is consumed by each node thereby saving energy.

3.1.3. Distance based techniques

Wireless transmission laws define a proportional relationship between power attenuation and the square of distance covered. Clustering approaches should be direction sensitive by considering the unique characteristic of up/down transmission direction in UWSNs. After gathering data from its members, a cluster should forward the aggregated message to its nearest UW-Sink. Hence distance based approaches mainly consider the shortest or nearest distance towards the sink.

3.1.3.1. Self-healing clustering (SHC). Huang et al. (2011a,b) have developed a cluster based algorithm that involves energy-efficient routing and data aggregation, specifically according to the application. It shows improved performance in terms of system lifetime and application-perceived quality. Here sensor nodes are organised into clusters with direction-sensitive, in which a particular node is considered as CH, in order to meet the unique characteristic of up/down transmission direction of UWSN medium. Further, there may occur re-clustering in the network for which the technique of self-healing is used that hampers the growth of excessive re-clustering and increases the robustness of clustered UWSNs.

3.1.3.2. Minimum average routing path clustering Problem (MARPCP). Kim et al. (2010) have proposed a cluster based scheme to select cluster heads so that there exists least hop distance measured from a node to its nearest sink. Also, Minimum Weighted Dominating Set Problem (MWDSP)-Revised is devised to overcome the high complexity of MARPCP. They have developed a fast and constant factor approximation algorithm for MARPCP and also shown that the existing approximation algorithms for MWDSP can be used to find approximated solutions for MARPCP.

3.1.3.3. Cluster based data aggregation (CDA). Manjula and Manvi (2012) proposed a cluster based technique for the UWSN involving formation of clusters and election of CH. Further data is transferred to the sink node with and without applying data aggregation technique (averaging technique).

3.1.3.4. Intra and inter cluster communication (IICC). Goyal et al. (2016) proposed a cluster based scheme that describes optimal election of CH and estimation of cluster size based on fuzzy logic.

In the paper MARPCP algorithm is used for intra-cluster communication and HMR-LEACH algorithm is used for inter-cluster communication. The experimental results show the improvement in packet delivery ratio, energy consumption, and delay in data delivery.

3.1.3.5. Improved data aggregation for cluster based UWSN (IDACB). Goyal et al. (2017) proposed a technique to reduce delay and packet drop at cluster nodes while data aggregation is being carried out in cluster based UWSN. In this technique improvement is done using sleep-wake up algorithm. TDMA based transmission schedule is used to avoid intra and inter cluster collision. Also data fusion is applied to reduce congestion.

3.2. Non-cluster based techniques

In Non-cluster based data aggregation, nodes which can be stationary or dynamic in nature are randomly deployed in the network. These nodes transfer the collected data to sink directly or indirectly. This non-cluster based classification is explained as follows:

3.2.1. Mobile sink based techniques

To improve the data transfer reliability in disruptive environment, a strategy to collect the data on-the-fly is essential. A mobile sink/data collector is required to gather data from the underwater sensor nodes. Different Autonomous Underwater Vehicles (AUVs), ships, and boats being non-stationary in aquatic environment can be deployed for data collection to enhance connectivity and coverage area thus becoming a good source of data collection.

3.2.1.1. Path reliability-aware data delivery (PRADD). Nowsheen et al. (2016) proposed a scheme for improving the reliability of delay tolerant underwater medium. Here a node with higher transmission reliability is selected as next hop forwarder to improve the data delivery in the network. The same node is also expected to have good reachability to gateways and better coverage probability. Mobile message ferries are used to collect urgent data from the gateways. A gateways election technique is also proposed to maximize their lifetime. PRADD requires the technique of active localization only during initial coarse location to detect the approximate location of sensor node. The movement of an anchored node is exploited to estimate its coverage probability. Further Data forwarding solution is also used in the network to improvise the data delivery along with reduction in overheads.

3.2.1.2. Scalable and efficient data gathering (SEDG). Ilyas et al. (2015) have presented a scalable data gathering protocol SEDG that improvises the data delivery ratio and saves energy by assigning member nodes with gateway node optimally. Further, the variable time span of AUV is used that reduces the packet drop ratio and exhibits better network throughput.

3.2.1.3. Energy efficient distributed time synchronization (E^2DTS). Li et al. (2013) proposed an algorithm to determine both clock skew and offset. These were required to attain high level time synchronization accuracy with less energy consumed. But in this scheme synchronization error is larger than mean value.

3.2.2. Relay based techniques

In relay or forwarder based data aggregation techniques, the data packets are forwarded through sensors here referred as relay sensors at the water surface. The data received at sink node in the form of acoustic signals from the sensors. During routing, the next relay node in the route is selected based on various factors like energy and distance.

3.2.2.1. Power efficient routing protocol (PERP). Huang et al. (2011a, b) developed a protocol to tackle the problems in underwater network. Here an idea of a forwarding node selector is used that determines the suitable sensor to forward the data packet further. A forwarding tree trimming method is applied to avoid excessive widespread of forwarded packets.

3.2.2.2. Extended RMTG (ERMTG). Dhurandher et al. (2013) have designed a geo-cast technique for an energy efficient UWSN. This is an extended work on RMTG protocol. The ERMTG algorithm takes into account the present energy state of the node to identify the next relay node. The transmission energy of a node depends on the distance between itself and the next hop node to which it wants to transmit. By preferring shorter paths, the suggested algorithm reduces the energy consumption, hence enhances the lifetime of the node.

3.3. Other techniques

3.3.1. Minimum latency aggregation scheduling (MLAS)

Wu et al. (2011) proposed a realistic aggregation scheduling scheme along with theoretical latency bound based on the hop radius and max degree of the network. Their scheme was based on virtual slot concept to explore multiplexing opportunities of time domain.

3.3.2. Non-intrusive underwater diffusion (UWD)

Lee et al. (2007) proposed a "conservative" communication structure designed in a minimalist's framework by assuming homogeneous GPS-free nodes with random mobility. It does not depend on GPS or power hungry motors to control currents. It is applied through non-intrusive underwater diffusion (UWD), which is an in-network processing with multi-hop ad hoc routing protocol. It has no proactive routing message exchanges and has negligible amount of on-demand floods. Hence the protocol avoids the excessive number of packets transmission so as to reduce underwater collisions in the network.

3.3.3. Receiver oriented sleep scheduling (ROSS)

Hong et al. (2013) proposed a receiver oriented strategy by using TDMA slots with sleep scheduling. The scheme detects and corrects both types of conflicts namely intra-family conflict and inter-family conflict. In this scheme tree structure is used and collision is avoided.

3.3.4. Energy-efficient compressed data aggregation

Hongzhi Lin et al. (2015) have developed a framework for three-dimensional underwater acoustic sensor networks (UASNs). Its goal is to minimize the total energy consumption during data transmission sensed by nodes. It consists of two layers, lower layer and upper layer. Lower layer is the compressed sampling layer, where nodes are divided into clusters. Upper layer is the data

Table 1
Simulation parameters.

Parameter	Value
Network size	75 nodes
Simulation time	10, 20, 30, 40 and 50 s
Packet size	250 bytes
Delta value	0.5
Traffic rate	50 Kbps
Channel capacity	2 Mbps
Range	100 m
Initial energy	1000 J
Transmission power	2.0 W
Receiving power	0.75 W

aggregation layer, where full sampling is adopted. They also incorporated the method to determine the number of clusters and the probability that a node will participate in data sampling or not.

3.3.5. Relaxation of distributed data aggregation

Rather than having each node operates in isolation, Rabbat and Coates (2014) aims to develop scheme under which sensors share information about their local measurements so that the tracking performance of the overall system should be superior to that of any individual sensor. Here, communication plays the role of synchronizing the state estimates across all nodes in the network so that, ideally, all nodes have the same estimate that would be computed by a single tracking algorithm that had direct access to all of the measurements. They used gossip algorithms to diffuse information across the network and drive the state estimates at each node.

4. Simulation analyses

In this paper the simulation results are conducted to analyse the performance of various techniques using AquaSim pack of NS-2 simulator ns-allinone-2.30. In this pack for UWSN, UnderwaterMAC as MAC layer protocol, UnderwaterPropagation as propagation type, UnderwaterChannel for underwater traversing are used to identify the link breakage and notify the same to network layer. Here Constant Bit Rate (CBR) is used as simulated traffic with traffic rate of 50 Kbps and the OmniDirectional Antenna type. These properties differentiate the UWSN simulation from WSN.

4.1. Performance metrics

We evaluate performance of the proposed protocol according to the following parameters:

Average Packet Delivery Ratio: It is the ratio of the number of packets received successfully to the total number of packets transmitted. It reflects the efficiency and reliability of the network.

Average Delay: It is the average time taken by a data packet for moving from source to destination. It involves detection and recovery delays. It is measured in seconds.

Energy Consumption: It is the energy grasped by the nodes during data transmission. It is expressed as the average energy consumption of all the nodes in the network during the simulation.

Packet Drop: It is the number of data packets dropped during the data transmission.

For better understanding we have divided our simulation into two scenarios as below.

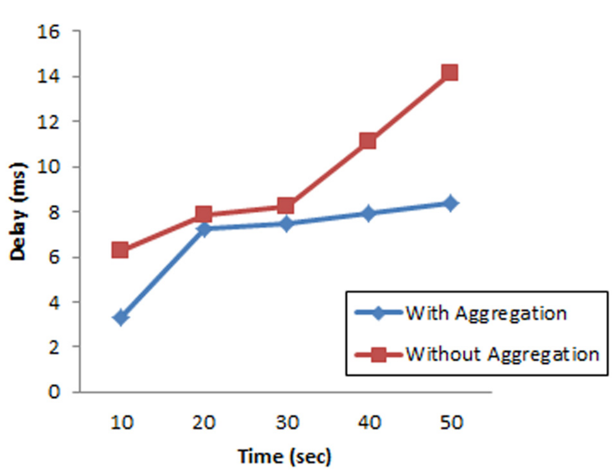

Fig. 2. Average data delivery delay vs time for K-means clustering.

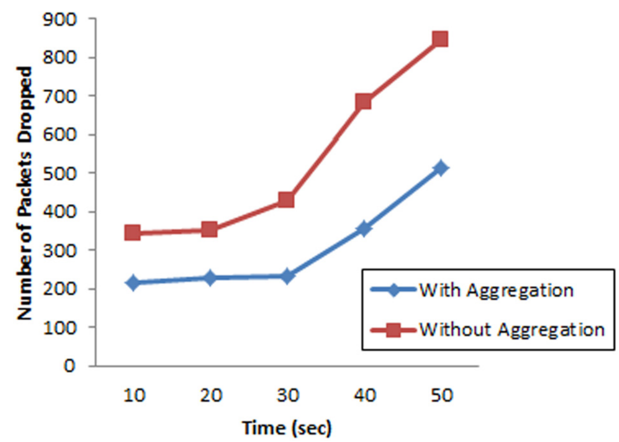

Fig. 3. Average packets dropped vs time for K-means clustering.

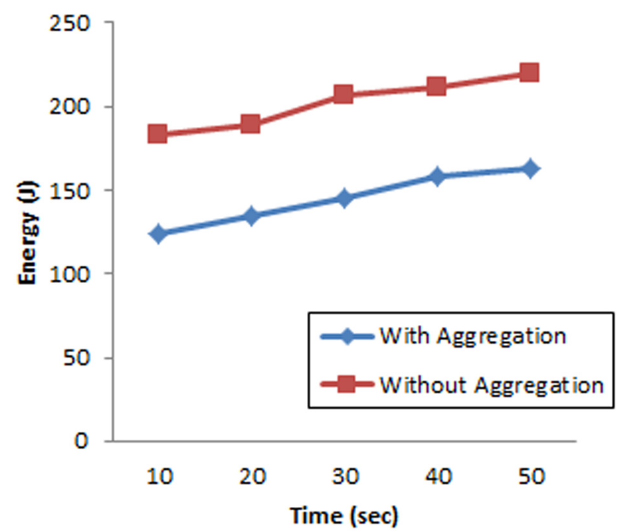

Fig. 4. Average energy consumption vs time for K-means clustering.

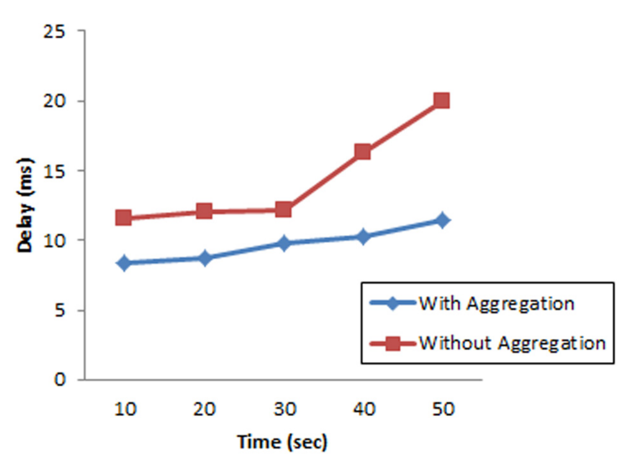

Fig. 5. Average data delivery delay vs time for round based clustering.

4.2. Scenario I

In UWSN each node has to send the data to the sink directly or indirectly. In some cases, node transmits its own data to the sink whereas in other cases, the technique of data aggregation is

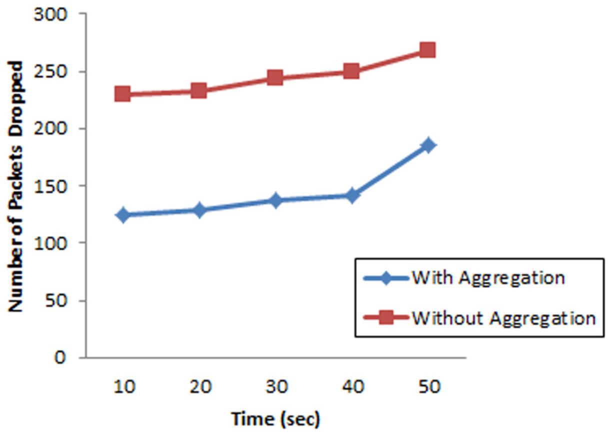

Fig. 6. Average packets dropped vs time for round based clustering.

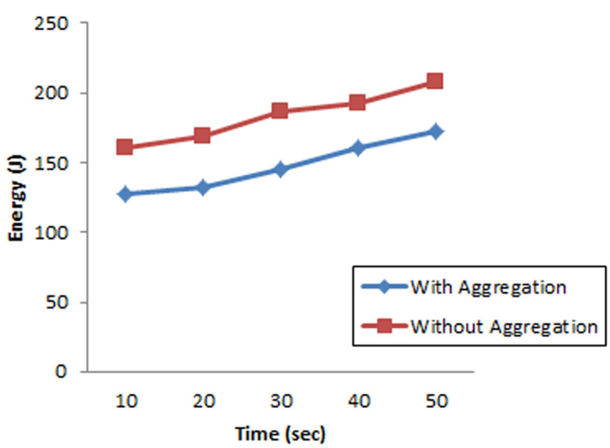

Fig. 7. Average energy consumption vs time for round based clustering.

applied due to which the transmitted data is collected at a particular node and then transmitted to the sink. To analyse the application of data aggregation and its results we have compared the simulation results of three cluster based techniques with and without data aggregation. The three techniques that are considered for evaluation are Round based clustering (Tran et al., 2014), K-means clustering (Harb et al., 2015), and Distributed underwater

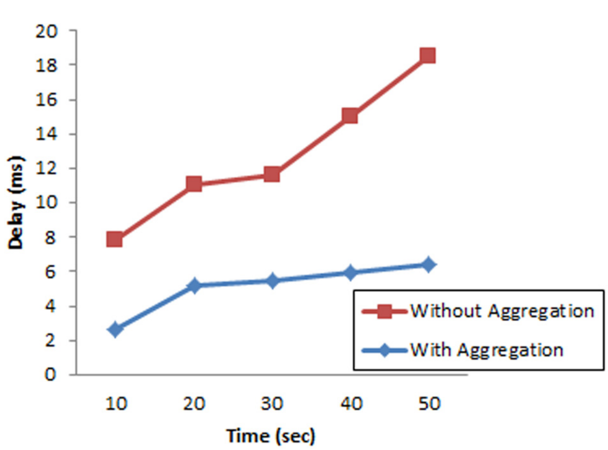

Fig. 8. Average data delivery delay vs time for distributed underwater clustering.

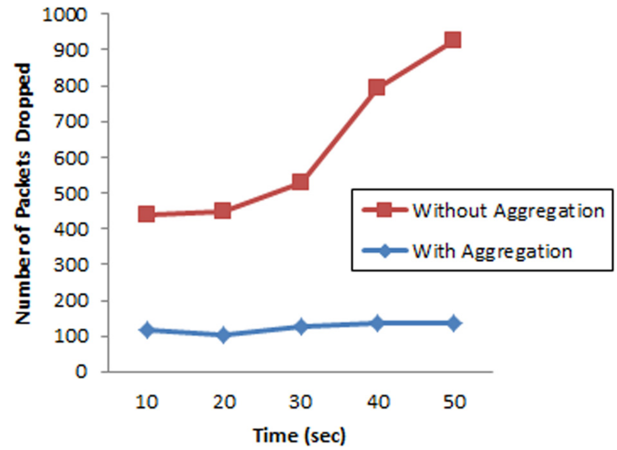

Fig. 9. Average packets dropped vs time for distributed underwater clustering.

clustering scheme (Domingo et al., 2007). In this scenario the performance metrics are taken as average delay, average packet drop and average energy consumption.

4.2.1. Simulation parameters

In the simulation for result evaluation, time variation taken as 10, 20, 30, 40 and 50 s. Because of memory usage and running time of the code, it is restricted to 50 s. But there will not be much deviation in the results, if we choose larger time intervals. The experimental area size considered here is $500 \times 500\,\text{m}^2$. Other simulation parameters are summarized in Table 1 for this scenario.

4.2.2. Simulation results

The simulation results of the cluster based techniques with and without aggregation are presented graphically in this section. In subsection A, we have analysed the performance of K-means clustering technique with and without aggregation based on parameters that are delay, packet drop, and energy consumption w.r.t. time. Further in subsection B and C, for analysis we have considered round based clustering and distributed underwater clustering schemes in both the scenarios of with and without aggregation. The analysis is based on the same parameters as above and the graphical representation of which are as following.

A. Results for K-means clustering

Fig. 2 shows the average delay occurred in K-means clustering technique with and without aggregation. It is inferred from the

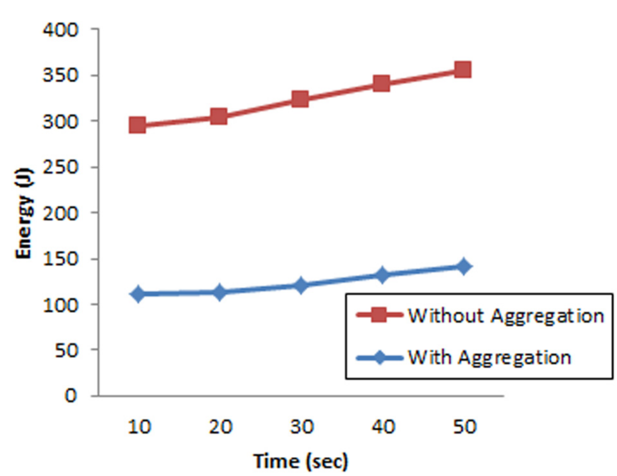

Fig. 10. Average energy consumption vs time for distributed underwater clustering.

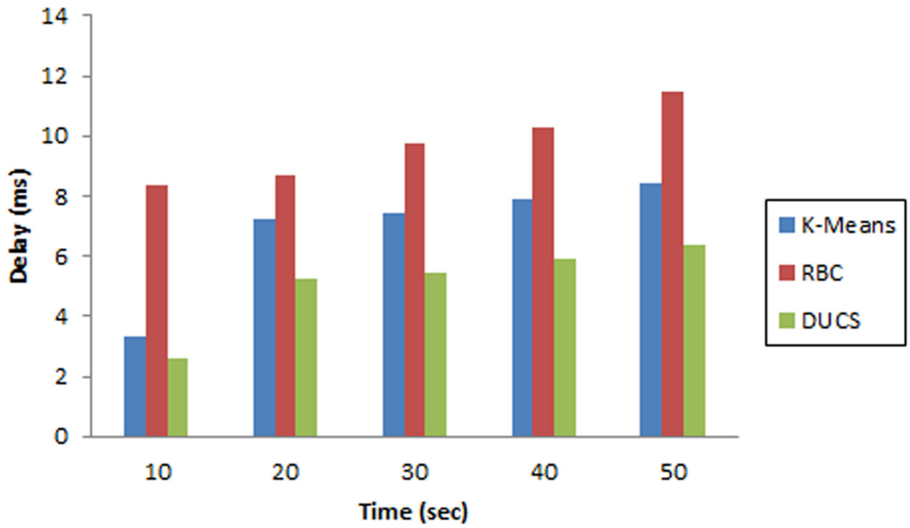

Fig. 11. Delay comparison of K-means, RBC, and DUCS w.r.t. increasing time.

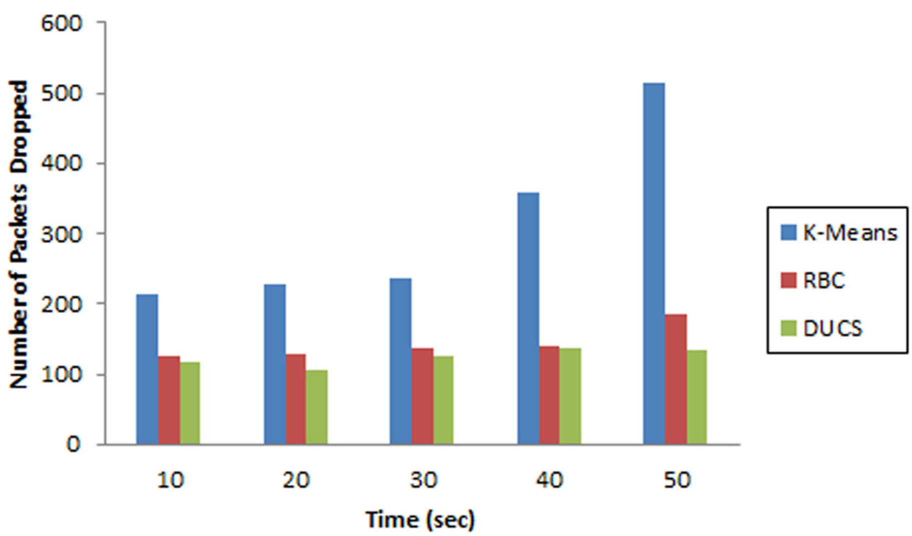

Fig. 12. Packets drop comparison of K-means, RBC, and DUCS w.r.t. increasing time.

results that K-means clustering technique when applied without aggregation in UWSN involves more collision amongst the data packets as compared to the technique applied with aggregation. Thus at particular intervals of time there occurs 27% less delay in case of with aggregation in comparison to the technique without aggregation.

Fig. 3 represents the average packet drop occurred w.r.t. time measured in seconds for K-means clustering technique with and without aggregation. Packet drop after aggregation is 41% less as compared to without aggregation. This is due to the fact that the technique of aggregation when applied leads to lesser number of collisions. However absence of aggregation implies more collision and buffer overflow in the network.

Fig. 4 shows the average energy consumed by K-means clustering technique with and without aggregation. The technique of data aggregation consumes lesser energy thus the K-means clustering technique with aggregation consumes 28% lesser energy in comparison to without data aggregation scenario. The reason behind the same is the occurrence of huge redundancy while sending data without aggregation.

B. Results for round based clustering

The above represented graph in Fig. 5 shows that for round based clustering technique, delay is 31% less in presence of data aggregation as it involves lesser collision amongst data packets in the network.

The graphical representation in the above Fig. 6 reiterates that the presence of aggregation leads to collision free network and lesser packet drop w.r.t. time. The numbers in the graph implies that average number of the packet drop is 42% lesser in the scenario of round based clustering technique with aggregation than without aggregation scenario. Also it has been seen that there is buffer overflow in the network in case of without aggregation.

Fig. 7 infers that the average energy consumption for round based clustering technique with aggregation is 20% less in comparison to the technique without data aggregation. The occurrence of huge redundancy during data transmission in case of without aggregation consumes more energy.

C. Results for distributed underwater clustering scheme

The above represented graph in Fig. 8 shows that delay for DUCS clustering technique is 31% less in presence of data

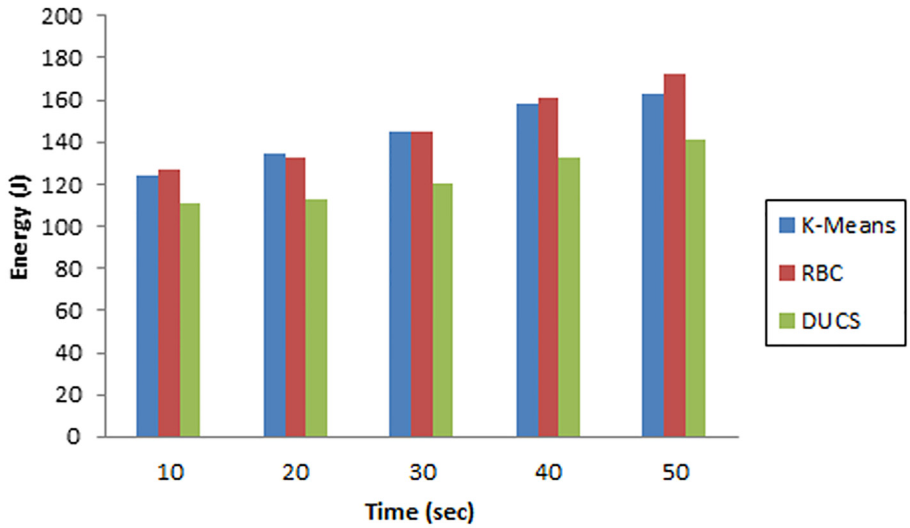

Fig. 13. Energy consumption comparison of K-means, RBC, and DUCS w.r.t. increasing time.

Table 2
Simulation parameters.

Parameter	Value
Network Size	50 nodes
Area size	1000 m × 1000 m
Traffic rate	50 Kbps
Channel capacity	2 Mbps
Range	100 m
Packet size	50, 100, 150, 200 and 250 bytes
Initial energy	1000 J
Transmission power	2.0 W
Receiving power	0.75 W

aggregation as it involves lesser collision amongst data packets in the network.

The graphical representation in the above Fig. 9 reiterates that the presence of aggregation leads to collision free network and lesser packet drop w.r.t. time. The numbers in the graph implies that average number of the packet drop is 73% lesser in the scenario of DUCS clustering technique with aggregation than without aggregation scenario. Also it has been seen that there is buffer overflow in the network in case of without aggregation

Fig. 10 infers that the average energy consumption for DUCS clustering technique with aggregation is 38% less in comparison to the technique without data aggregation. The occurrence of huge redundancy during data transmission in case of without aggregation consumes more energy.

4.2.3. Combined comparison of cluster based routing protocols with data aggregation

When we apply data aggregation on various protocols the performance gets improved. Below shown figures are the comparisons of various protocols with data aggregation. The performance metrics delay, packet drop and energy consumption are compared with varying simulation time from 10 to 50 s.

The above represented graph in Fig. 11 shows the aggregation delay for K-means, Round Based and Distributed underwater clustering techniques. The delay of distributed scheme (DUCS) is 25% less than K-means and 48% less than Round based clustering method (RBC) w.r.t. increasing time by applying data aggregation.

The graphical representation in the Fig. 12 shows the packets drop for K-means, Round Based and Distributed underwater clustering techniques. The graph implies that after data aggregation average number of packet drop of DUCS is 56% less than K-means and 13% less than RBC method w.r.t. increasing time.

The graphical representation in the Fig. 13 shows the average energy consumption for K-means, Round Based and Distributed underwater clustering techniques. The graph implies that using data aggregation average energy consumed of DUCS is 15% less than K-means and 16% less than RBC method w.r.t. increasing time.

It can be concluded from the above graphs that DUCS outperforms the other two techniques in terms of delay, packet drop and energy consumption during aggregation.

4.3. Scenario II

In this scenario to analyse the application of data aggregation, we have compared the simulation results of three categories (cluster based, non-cluster based, and other technique) with aggregation. The three techniques that are considered for evaluation are Receiver Oriented Sleep Scheduling (ROSS) (Hong et al., 2013), Intra and Inter Cluster Communication (IICC) (Goyal et al., 2016) and Energy Efficient Distributed Time Synchronization (E^2DTS) (Li et al., 2013). In this scenario the performance metrics delay, delivery ratio and energy consumption are measured by varying the packet size from 100 to 250 bytes.

4.3.1. Simulation parameters

In the simulation for result evaluation, the area size taken is $1000 \times 1000\ m^2$ for a simulation time of 50 s. However if larger time intervals are chosen, then there will not be much deviation in the results. The other parameters for the simulation are summarized in Table 2.Table 3**

4.3.2. Simulation results

The simulation results of three categories (cluster based, non-cluster based, and other technique) with aggregation are presented graphically in this section below. Here, we have analysed the performance of ROSS, IICC, and E^2DTS techniques based on parameters that are delay, delivery ratio, and energy consumption w.r.t. varying packet size.

4.3.3. Comparison of various data aggregation routing protocols

The above represented graph in Fig. 14 shows the aggregation delay for IICC, ROSS and E^2DTS techniques. The delay of IICC scheme is 66% less than ROSS and 36% less than E^2DTS method w.r.t. increasing packet size by applying data aggregation.

Table 3
Comparison of various data aggregation techniques.

S. No	Authors	Pub. Year	Type	Proposed Method	Metrics	Advantages	Disadvantages
1.	Khoa Thi-Minh Tran et al.	2014	Similarity based	Round-based clustering to reduce redundant data transmission	Throughput, Energy consumption, Data received ratio	High-throughput and low energy consumption	Lack to achieve best performance for node mobility
2.	Khoa Thi-Minh Tran et al.	2013	Similarity based	Similarity functions based Data Aggregation	Data Lost, Data sent ratio and Data deleted ratio	Minimize data redundancy and packet loss	The combination of similarity function and under water protocol are not explained
3.	Hassan Harb et al.	2015	Similarity based	Enhanced K-means and ANOVA based clustering	Data sent ratio and Energy consumption	Decreased data redundancy	Higher Energy consumption
4.	Manjula R. B. et al.	2012	Similarity based	Data aggregation technique for UWSN	Energy consumption	Less Energy consumption and extends the network life time	The proposed method did not explain about the resource utilization.
5.	Mari Carmen Domingo and Rui Prior	2007	Mobility based	A GPS-free clustered routing protocol	Routing Overhead, Packet delivery ratio	Better delivery ratio and Minimized proactive routing exchange, data loss, overhead	
6.	Muhammad Ayaz et al.	2010	Mobility based	Temporary Cluster Based Routing protocol	Packet Delivery Ratios, End to End Delays and Power Consumption	Reduce Energy consumption	Node mobility issue with wide communication coverage
7.	Chenn-Jung Huang et al.	2011	Distance based	Forwarding node selector and tree trimming mechanism	Energy consumption, Packet drop, delay Packet delivery ratio	Outstanding results in terms of data received at base station and number of active nodes.	The Energy consumption ratio is high
8.	Donghyun Kim et al.	2010	Distance based	Minimum Average Routing Path Clustering	Delay, Throughput	Good performance ratio	
9.	Nusrat Nowsheen et al.	2015	Mobile Sink based	Path Reliability-Aware Data Delivery protocol	Packet delivery ratio, Routing Overhead and Energy Consumption, Location Estimation Error	Low overhead and less energy consumption	Didn't explain about the dynamic data transfer
10.	Naveed Ilyasa et al.	2015	Mobile Sink based	Scalable and Efficient Data Gathering protocol	Energy Consumption, Network throughput, Packet delivery ratio, Delay and Network life time	Maximize Network life time, Packet delivery ratio	Ideal route of AUV collect or gather data efficiently remains an issue
11.	Sanjay K. Dhurandher et al.	2013	Relay based	Geocast technique using ERMTG algorithm	Packet delivery ratio, Path Energy, Network Energy, Delay	Good delivery ratio and less energy consumption	
12.sss	Zuodong Wu et al.	2011	Other techniques based	Minimum-Latency aggregation scheduling	Aggregation latency	Reduced latency	
13.	Uichin Lee et al.	2007	Other techniques based	Conservative communications architecture using non-intrusive underwater diffusion	Delay, Delivery Ratio and packet sent	Minimize data packet collision	

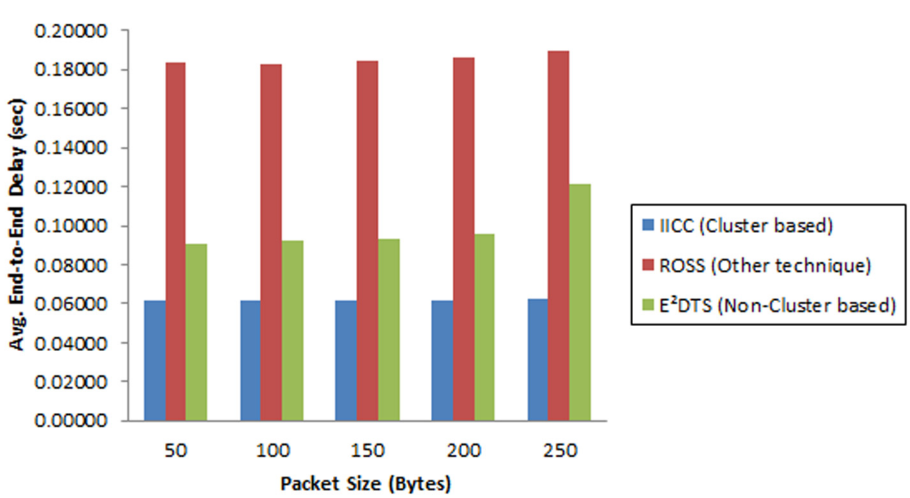

Fig. 14. Average end-to-end delay comparison of IICC, ROSS, E^2DTS w.r.t. increasing packet size.

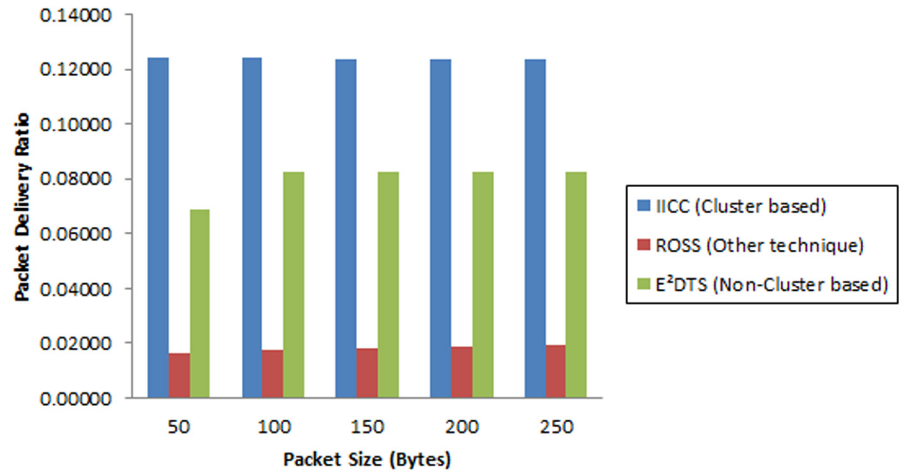

Fig. 15. Average packet delivery ratio comparison of IICC, ROSS, E^2DTS w.r.t. increasing packet size.

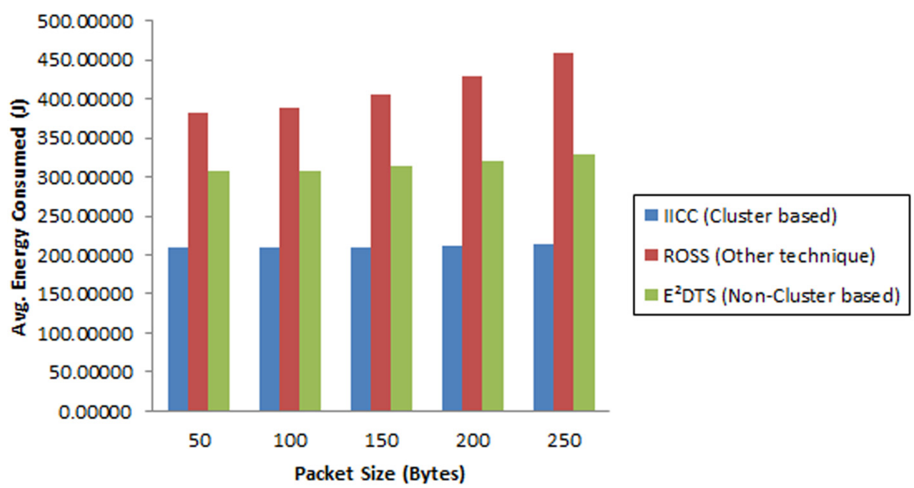

Fig. 16. Average energy consumption comparison of IICC, ROSS, E^2DTS w.r.t. increasing packet size.

The above represented graph in Fig. 15 shows the average packet delivery ratio for IICC, ROSS and E^2DTS techniques. The data delivery ratio of IICC scheme is 85% better than ROSS and 35% better than E^2DTS method w.r.t. increasing packet size by applying data aggregation.

The above represented graph in Fig. 16 shows the average energy consumed for IICC, ROSS and E^2DTS techniques. The graph implies that using data aggregation, IICC consumes 48% less energy than ROSS and 33% less than E^2DTS method w.r.t. increasing packet size.

It can be determined from the above graphs that cluster based techniques are better than the non-cluster based and other techniques using data aggregation.

5. Comparison Table

The following table presents the comparison of various aggregation techniques discussed in this paper for better understanding.

6. Open Issues and future challenges

The future challenges of data aggregation techniques in UWSN are listed below:

(i) The clustering protocol should be energy efficient and reliable.
(ii) The cluster heads involved in data aggregation process should be honest and trust worthy.
(iii) The aggregation should ensure fault tolerance (i e.) it should handle all types of faults involved in the data delivery.
(iv) The aggregation should be free from collision or interference by choosing suitable scheduling technique.
(v) The aggregation should be free from congestion and overloading.
(vi) The aggregation technique should provide suitable loss recovery techniques for recovering burst losses.
(vii) The similarity functions used for data aggregation should eliminate the outliers or inconsistent data.
(viii) Node or cluster head mobility should be handled in an efficient manner.
(ix) The aggregation techniques should ensure minimum latency and overhead.
(x) While constructing aggregation trees or paths, effective void recovery and depth adjustment processes should be applied.

7. Conclusion

This paper presents a survey on data aggregation techniques in UWSN discovered by researchers in the past. Various protocols have been designed for the purpose of surveillance, monitoring, assisted navigation etc. in UWSN. It was further explored that the use of data aggregation will not only increase the lifetime of the network but also saves energy consumption during data transmission. Based on research, the data aggregation techniques are classified into three major categories that are cluster based, non-cluster based, and other techniques. The cluster based techniques are further categorized as similarity function based, distance based and mobility based. Similarly, the non-cluster based techniques are further categorized as mobile data collector based and relay based.

In our paper the simulation results of a technique with data aggregation is compared to the results of same technique without data aggregation to show the impact of data aggregation. For the very purpose we have considered three cluster based techniques that are K-means, RBC, and DUCS clustering, whose results are graphically represented in terms of delay, packet drop and energy consumption using NS-2 simulation tool. The reduction in the three parameters delay, packet drop and energy consumption in case of protocol with data aggregation is mainly due to lesser collision and redundancy. We have also compared IICC, E^2DTS, and ROSS techniques (cluster based, non-cluster based, and other technique respectively) based on parameters that are delay, delivery ratio, and energy consumption w.r.t. varying packet size. A brief comparison table of existing data aggregation techniques in UWSN is presented, from which the future challenges of data aggregation in UWSN are identified. This has truly justified the use of data aggregation technique along with routing protocol in UWSN as an efficient tool and has generated the need of survey on data aggregation that may create further interest towards its scope and to overcome its limitations to meet the identified challenges.

References

Ayaz, M., Abdullah, A., Jung, L.T., 2010. Temporary cluster based routing for underwater wireless sensor networks. IEEE Int. Symp. Inf. Technol. 2, 1009–1014.

Ayaz, M., Baig, I., Abdullah, A., Faye, I., 2011. A survey on routing techniques in underwater wireless sensor networks. J. Netw. Comput. Appl. 34, 1908–1927.

Coutinho, R.W., Boukerche, A., Vieira, L.F., Loureiro, A.A., 2015. A novel void node recovery paradigm for long-term underwater sensor networks. Ad Hoc Netw. 34, 144–156.

Curiac, D.I., 2016. Towards wireless sensor, actuator and robot networks: conceptual framework, challenges and perspectives. J. Netw. Comput. Appl. 63, 14–23.

Das, A.P., Thampi, S.M., 2017. Fault-resilient localization for underwater sensor networks. Ad Hoc Netw. 55, 132–142.

Dhurandher, S.K., Obaidat, M.S., Gupta, M., 2013. Energized geocasting model for underwater wireless sensor networks. Simul. Model. Pract. Theory 37, 125–138.

Domingo, M.C., Prior, R., 2007. A distributed clustering scheme for underwater wireless sensor networks. In: The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1–5.

Gholami, E., Rahmani, A.M., Fooladi, M.D.T., 2015. Adaptive and distributed TDMA scheduling protocol for wireless sensor networks. Wireless Pers. Commun. 80 (3), 947–969.

Goyal, N., Kumar, M., 2014. Reviewing underwater acoustic wireless sensing networks. Int. J. Comput. Sci. Technol. 5 (2), 95–98.

Goyal, N., Dave, M., Verma, A.K., 2014b. Fuzzy based clustering and aggregation technique for under water wireless sensor networks. In: IEEE International Conference on Electronics and Communication System (ICECS-2014), pp. 1–5.

Goyal, N., Dave, M., Verma, A.K., 2016. Energy efficient architecture for intra and inter cluster communication for underwater wireless sensor networks. Wireless Pers. Commun. 89 (2), 687–707.

Goyal, N., Dave, M., Verma, A.K., 2017. Improved data aggregation for cluster based underwater wireless sensor networks. Proc. Natl. Acad. Sci., India, Sect. A Phys. Sci., 1–11

Han, G., Liu, L., Jiang, J., Shu, L., Rodrigues, J.J., 2016. A collaborative secure localization algorithm based on trust model in underwater wireless sensor networks. Sensors 16 (2), 229.

Harb, H., Makhoul, A., Couturier, R., 2015. An Enhanced K-Means and ANOVA-Based clustering approach for similarity aggregation in underwater wireless sensor networks. IEEE Sens. J. 15 (10), 5483–5493.

Hong, L., Hong, F., Yang, B., Guo, Z., 2013. ROSS: receiver oriented sleep scheduling for underwater sensor networks. In: Proceedings of the 8th ACM International Conference on Underwater Networks and Systems, Taiwan, p. 4.

Lin, Hongzhi, Wei, Wei, Zhao, Ping, Ma, Xiaoqiang, Zhang, Rui, Liu, Wenping, Deng, Tianping, Peng, Kai, 2015. Energy-efficient compressed data aggregation in underwater acoustic sensor networks. Wireless Netw. 22 (6), 1985–1997.

Huang, C.J., Wang, Y.W., Lin, C.F., Chen, Y.T., Chen, H.M., Shen, H.Y., Chen, Y.J., Chen, I. F., Hu, K.W., Yang, D.X., 2011a. A self-healing clustering algorithm for underwater sensor networks. Cluster Comput. 14 (1), 91–99.

Huang, C.J., Wang, Y.W., Liao, H.H., Lin, C.F., Hu, K.W., Chang, T.Y., 2011b. A power-efficient routing protocol for underwater wireless sensor networks. Appl. Soft Comput. 11 (2), 2348–2355.

Ilyas, N., Akbar, M., Ullah, R., Khalid, M., Arif, A., Hafeez, A., Qasim, U., Khan, Z.A., Javaid, N., 2015. SEDG: scalable and efficient data gathering routing protocol for underwater wsns. Proc. Comput. Sci. 52, 584–591.

Izadi, D., Abawajy, J., Ghanavati, S., 2015. An alternative clustering scheme in WSN. IEEE Sens. J. 15 (7), 4148–4155.

Jadidoleslamy, H., Aref, M.R., Bahramgiria, H., 2016. A fuzzy fully distributed trust management system in wireless sensor networks. AEU-Int. J. Electron. Commun. 70 (1), 40–49.

Jia, J., Meng, J., 2016. Impulsive noise rejection for ZigBee communication systems using error-balanced wavelet filtering. AEU-Int. J. Electron. Commun. 70 (5), 558–567.

Kafetzoglou, S., Alexandropoulou, M., Papavassiliou, S., 2008. A novel data gathering framework for resource-constrained underwater sensor networks. Ad Hoc & Sensor Wireless Networks 1 (5), 313–329.

Karimi, H., Medhati, O., Zabolzadeh, H., Eftekhari, A., Rezaei, F., Dehno, S.B., 2015. Implementing a reliable, fault tolerance and secure framework in the wireless sensor-actuator networks for events reporting. In: Procedia Computer Science, International Conference on Advanced Wireless Information and Communication Technologies (AWICT 2015) 73, pp. 384–394.

Kim, D., Wang, W., Ding, L., Lim, J., Oh, H., Wu, W., 2010. Minimum average routing path clustering problem in multi-hop 2-d underwater sensor networks. Optim. Lett. 4 (3), 383–392.

Kumar, R., Singh, N., 2014. A survey on data aggregation and clustering schemes in under water sensor networks. Int. J. Grid. Distrib. Comput. 7 (6), 29–52.

Lee, U., Kong, J., Gerla, M., Park, J.S., Magistretti, E., 2007. Time-critical underwater sensor diffusion with no proactive exchanges and negligible reactive floods. Ad Hoc Netw. 5 (6), 943–958.

Li, Z., Guo, Z., Hong, F., Hong, L., 2013. E^2DTS: An energy efficiency distributed time synchronization algorithm for underwater acoustic mobile sensor networks. Ad Hoc Netw. 11 (4), 1372–1380.

Liu, Y., Liu, A., He, S., 2015. A novel joint logging and migrating traceback scheme for achieving low storage requirement and long lifetime in WSNs. AEU-Int. J. Electron. Commun. 69 (10), 1464–1482.

Manjula, R.B., Manvi, S.S., 2012. Cluster Based Data Aggregation in Underwater Acoustic Sensor Networks. In: IEEE India Conference (INDICON), pp. 104–109.

Nowsheen, N., Karmakar, G., Kamruzzaman, J., 2016. PRADD: a path reliability-aware data delivery protocol for under water acoustic sensor networks. J. Netw. Comput. Appl. 75, 385–397.

Oh, S.H., Tran, K.T.M., 2013. A comparative analysis of similarity functions of data aggregation for underwater wireless sensor networks. Int. J. Digital Content Technol. Appl. 7 (2), 830.

Rabbat, Coates, 2014. Relaxation of Distributed Data Aggregation for Underwater Acoustic Sensor Networks. Contract Report. McGill University.

Rahman, A.U., Alharby, A., Hasbullah, H., Almuzaini, K., 2016. Corona based deployment strategies in wireless sensor network: a survey. J. Netw. Comput. Appl. 64, 176–193.

Rezvani, M., Ignjatovic, A., Bertino, E., Jha, S., 2015. Secure data aggregation technique for wireless sensor networks in the presence of collusion attacks. IEEE Trans. Dependable Secure Comput. 12 (1), 98–110.

Saranya, D., Arthi, K., 2016. Underwater wireless sensor networks using enhanced K-mean and clustering approach. Int. J. Emerging Technol. Comput. Sci. Electron. 21 (4).

Senel, F., Akkaya, K., Erol-Kantarci, M., Yilmaz, T., 2015. Self-deployment of mobile underwater acoustic sensor networks for maximized coverage and guaranteed connectivity. Ad Hoc Netw. 34, 170–183.

Shen, H., Bai, G., 2016. Routing in wireless multimedia sensor networks: a survey and challenges ahead. J. Netw. Comput. Appl. 71, 30–49.

Tran, K.T.M., Oh, S.H., Byun, J.Y., 2014. An efficient data aggregation approach for underwater wireless sensor networks 24, 46–48.

Tran, K.T.M., Oh, S.H., Byun, J.Y., 2013. Well-suited similarity functions for data aggregation in cluster-based underwater wireless sensor networks. Int. J. Distrib. Sens. Netw.

Vennila, C., Madhura, M., 2016. An energy-efficient attack resistant trust model for underwater wireless sensor networks. Middle-East J. Sci. Res. 24 (S2), 33–39.

Wu, Z., Tian, C., Jiang, H., Liu, W., 2011. Minimum-latency aggregation scheduling in underwater wireless sensor networks. In: IEEE International Conference on Communications (ICC), pp. 1–5.

Xu, M., Liu, G., Guan, J., 2015. Towards a secure medium access control protocol for cluster-based underwater wireless sensor networks. Int. J. Distrib. Sens. Netw., 1–11 Hindawi Publishing Corporation.

Zenia, N.Z., Aseeri, M., Ahmed, M.R., Chowdhury, Z.I., Kaiser, M.S., 2016. Energy-efficiency and reliability in mac and routing protocols for underwater wireless sensor network: a survey. J. Netw. Comput. Appl. 71, 72–85.