Abstract--- Underwater wireless sensor networks (UWSN) is recently proposed to observe aquatic environments. It uses acoustic technology for underwater communication. In this paper, a new clustering method is used to handle similarity between node readings which are sent periodically to the cluster-head. A two tier data aggregation technique is proposed. First level the node periodically cleans its reading to eliminate data redundancies. At second level an enhanced K-mean algorithm based on one way ANOVA model is applied on the cluster head in order to identify nodes generating identical data sets and aggregated them before sending to sink. Our experimental results show that this technique largely reduces data redundancies thus extending the network lifetime.

Index Terms- Underwater Wireless sensor network, K-mean clustering, ANOVA model, Data aggregation, Network Lifetime.

I. INTRODUCTION

Underwater Wireless Sensor Network consists of a large number of sensors that are deployed in aquatic environment. They are capable of monitoring the surroundings and explore the underwater environments. In Underwater Wireless Sensor Networks (UWSNs) the sensor nodes are deployed in the oceans which are responsible for collecting the sensed information and send it to the sink. The effectiveness of the UWSNs lie in their sensing quality, flexibility, coverage, etc., they can offer. Node clustering and Data aggregation are the two methods to organize data traffic and reduce in network redundancies. Node clustering makes the network smaller and it extends the network lifetime.

Data aggregation is used to minimize the energy consumption by eliminating data redundancies. These two techniques combines to improve the network performances. Many clustering protocols are proposed in UASN. In which few protocols consider sensor readings similarity and the correlation between received data. But, the lack of suitable energy efficient protocols for handling correlations leads us to study the data aggregation and clustering protocol that creates clusters of nodes with identical readings.

The ultimate goal of the UWSNs deployed in the underwater environments is often to deliver the sensing data from sensor nodes to sink node and then conduct further analysis at the sink node. Data collection becomes an important factor in determining the performance of such UWSNs.

APPLICATION OF UWSNs

- Environmental Monitoring,
- Real-time warship Monitoring,
- Locating mooring positions and submerged wrecks,
- Disaster prevention,
- Oceanographic data collection,
- Mine reconnaissance,
- Assisted navigation,
- Undersea explorations,
- Tactical surveillance.

II. RELATED WORK

CLUSTER-BASED FALSE DATA FILTERING SCHEME

In wireless sensor networks, the adversaries can inject false data reports from compromising nodes. Previous approaches for filtering false reports share keys between the source node and its upstream nodes on the path to sink, and rely on intermediate nodes to verify the reports generated by downstream nodes in a probabilistic manner. As a result, false reports have to travel several hops before detected. Worse still, these schemes haven’t balanced the overheads of all nodes in the process of keys distributing. In
response to these, this project proposes a cluster-based filtering scheme, in which nodes are grouped into clusters once deployed by employing some strong nodes act as cluster heads. Then a distributed method of keys assignment is proposed by constructing a sink-rooted tree which comprises of all the cluster heads, guarantees that the keys of a source cluster are stored by several forwarding clusters close to it and thus to filter false reports generated by the source cluster during several hops during forwarding, further, the number of authentication keys held by the forwarding clusters getting smaller with the distance increase from the source cluster and thus to balance the keys stored by each forwarding cluster.

III. CLUSTER-BASED ARCHITECTURE
A 2-D UWSN with cluster based architecture is considered in this paper. Clustering is the efficient topology control method used to increase network lifetime and scalability. The network is divided into a number of clusters in clustering scheme in which each cluster has a CH. Cluster Head manages the clusters. Single-hop communication is used for the data transmission between the sensor nodes and the CH. The periodic data collection system is used in which the sensor sends its data periodically to the CH. The CH sends the received data sets to the sink which is also performed periodically. A two tier data aggregation based transmission efficient technique is proposed. This technique is applied at each cluster. First level the node periodically cleans its reading to eliminate data redundancies. At second level an enhanced K-mean algorithm based on one-way ANOVA model is applied on the cluster head in order to identify nodes generating identical data sets and aggregated them before sending to sink. This technique significantly reduces the amount of data sets which are sent to the sink. The similarities between the collected data sets are defined using two different functions. They are similar function and measure weight function.

K-MEAN BASED CLUSTERING METHOD
Three decades ago, the K-Mean clustering algorithm was proposed by MacQueen. The nodes member sends all datasets to the cluster head, in which the cluster head applies K-Means based clustering algorithm to classify these sets by identical data sets with the aim to eliminate redundancies and reduce the huge amount of data. An enhanced K-mean clustering algorithm using the one way ANOVA model is proposed in order to identify neighboring nodes generating identical data sets.

PERFORMANCE METRICS
The following metrics are used to evaluate the performance of the proposed method.

- **PERCENTAGE OF DATA AGGREGATION:** The ratio of the number of delivered data packet to the destination. This illustrates the level of delivered data to the destination.
- **NETWORK LIFETIME:** The time interval from the start of the network operation until the death of the last alive sensor.
- **ENERGY CONSUMPTION:** The total amount of energy consumed by the node.

IV. PROPOSED METHOD
In this paper, a new clustering method is used to deal with the spatial similarity between the node readings. Cluster Head (CH) receives the readings periodically from sensor nodes. Two common methods to organize data traffic and to reduce in network redundancies which improve the energy consumption and network scalability are node clustering and data aggregation. By reducing the data transmissions between nodes and sink the lifetime of the network will be extended and the size of the network will look smaller by the technique node clustering. Best way to minimize the consumption of energy by eliminating data redundancy is done by data aggregation techniques. It also reduces the number of transmissions to the sink. These two techniques combines to improve the network performances. Then a two tier data aggregation technique is proposed. First level the node periodically cleans its reading to eliminate data redundancies. At second level an enhanced K-mean algorithm based on one-way ANOVA model is applied on the cluster head in order to identify nodes generating identical data sets and aggregated them before sending to sink. Our experimental results show that this technique largely reduces data redundancies thus extending the network lifetime.

V. RESULTS AND DISCUSSION
In this section, the experimental results of this technique, at both sensor node and CH level is performed. This is shown via simulation on real data the efficiency of our approach in saving energy and reducing the huge amount of data thus extending the network lifetime of real UWSNs. This paper consists of 240 sensors over an area of 1000 * 1000m2. They consist of three cluster heads N1 = 3; N2 = 29 and N3 = 63 sensors respectively. Each node periodically reads real.
measures while applying the first aggregation phase. At the end of this step, each node sends its set of measures with frequencies to their corresponding CH which in his turn applies the CH aggregation phase.

SIMULATION RESULTS

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>PERCENTAGE OF DATA SENT TO CH (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>With aggregation T=100</td>
<td>24%</td>
</tr>
<tr>
<td>With aggregation T=500</td>
<td>20%</td>
</tr>
<tr>
<td>With aggregation T=1000</td>
<td>18%</td>
</tr>
<tr>
<td>Without aggregation</td>
<td>100%</td>
</tr>
</tbody>
</table>

T- Total number of slots in a period

<table>
<thead>
<tr>
<th>CLUSTERING TECHNIQUE</th>
<th>ENERGY CONSERVATION (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-Mean algorithm</td>
<td>80%</td>
</tr>
<tr>
<td>Sensor node applies aggregation phase</td>
<td>Upto 96%</td>
</tr>
</tbody>
</table>

DATA AGGREGATION RATIO AT THE SENSOR NODE

Similar function is used to reduce the amount of data collected at each sensor node over a period by the process of eliminating redundancies. The obtained results show that each sensor node will send in the worst case scenario, e.g. threshold = 0.01 and T =200, 24% of its collected data to the CH when the aggregation techniques is applied and when the aggregation is not applied the 100% of the collected data will be sent to sink which consumes more energy.

This figure clearly shows the percentage of aggregated measures at sensor nodes. Depends on the number of slots over a period (T) the percentage of aggregation differs. When the number of data collected increases the amount of data sent to sink will be reduced as the data collected by sensors will be same over a closer region.

ENERGY CONSUMPTION STUDY

This section, deals with the study of energy consumption at the sensor nodes and CH levels. The energy consumption in network depends on amount of data sent and received. Fig. 4 shows the energy consumption comparison with and without applying the aggregation phase by each sensor node and when varying T and threshold. The redundancy among data collected by the sensor node in aggregation phase is reduced significantly, which allows it to save proportionally its energy when transmitting its data to the CH at each period. It is important to notice that our technique can conserve energy of a sensor node up to 96%.

On the other hand, the energy consumption comparison between the ANOVA model with the three tests and the PFF technique at the CHs level is considered. We fixed T and threshold and we varied N to N1, N2 and N3 respectively, while we fixed N and threshold and we varied T to 200, 500 and 1000 respectively. The obtained results show that our technique outperforms PFF for all values of thresholds and it reduces up to 70% of the energy consumption when compared to PFF. Prefix Frequency Filtering techniques does not obtain results as we obtain from the K-mean clustering method. When compared to the PFF technique the ANOVA model eliminates more data sets in CH.

![Figure 3](image3.png) Percentage of aggregated measures at the sensor nodes.

![Figure 4](image4.png) Energy consumption at each sensor node T=200

![Figure 5](image5.png) Energy consumption at each sensor node T=500
When compared to other tests in one way ANOVA model the, Bartlett test decreases the energy consumption of the cluster head (CH). The CH conserves more energy when compared to the Prefix Frequency Filtering Techniques as the value of N is increased. Similarly the CH energy consumption is minimized as there is decrease in the value of threshold.

NETWORK LIFETIME

The experiment results show that this technique has largely reduced the data redundancy and has also extended the network lifetime.

PACKETS SENT TO SINK

After performing the two tier data aggregation techniques which largely eliminate the data redundancies. The amount of data gathered is significantly reduced while they are sent to the sink as they generate same or very similar data sets. Aggregation techniques eliminates data redundancies to conserve energy and to extend network lifetime.

VI CONCLUSION AND FUTURE WORK

In UWSN, researches are done for data aggregation based on clustering to minimize energy consumption and network lifetime extension. This approach is used to avoid multihop communications and develop an aggregated path around network. A two tier data aggregation based transmission efficient technique is proposed. This technique is applied at each cluster. First level the node periodically cleans its reading to eliminate data redundancies. At second level an enhanced K-mean algorithm based on one-way ANOVA model is applied on the cluster head in order to identify nodes generating identical data sets and aggregated them before sending to sink. Our experimental results show that this technique largely reduces data redundancies thus extending the network lifetime.

In future, the sensors are scheduled in a manner that the node generating redundant data will not be active at same time. So that the sensor nodes will conserve more energy and increase network lifetime.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their detailed comments and suggestions.

REFERENCES


D. Saranya received her B.TECH. degree in Electronics and Communication Engineering from SRM University, India in 2014 and now pursuing M.E. degree in Communication Systems from Valliammai Engineering College under Anna University, Chennai, India. Her areas of research interests include, Wireless Sensors Networking, Wireless Networks, Communication Switching Techniques and Network Security.

K. Arthi received her B.E degree in Electrical and Communication Engineering, M.E degree in VLSI Designs. She is serving as Asst. Professor in the Department of Electronics and Communication Engineering, Valliammai Engineering College, Kattankulathur, Tamilnadu, India. Her research interest include Wireless Networking and Wireless Sensor Networks, VLSI Designs.

D. Saranya

K. Arthi