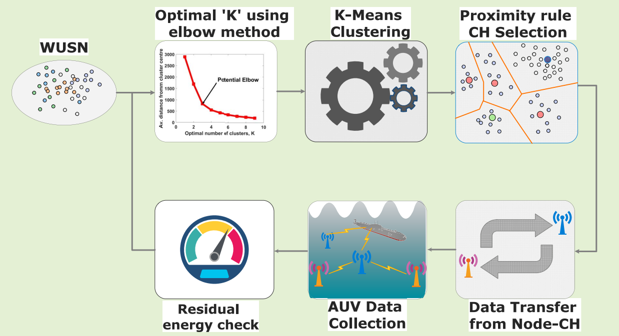
DEKCS: A Dynamic Clustering Protocol to Prolong Underwater Sensor Networks

Kenechi G. Omeke, *Graduate Student Member, IEEE*, Michael S. Mollel, Metin Ozturk, Shuja Ansari, Lei Zhang, *Senior Member, IEEE*, Qammer H. Abbasi, and Muhammad Ali Imran, *Senior Member, IEEE*

***Abstract*—Energy consumption is a critical issue in the design of wireless underwater sensor networks (WUSNs). Data transfer in the harsh underwater channel requires higher transmission powers compared to an equivalent terrestrial-based network to achieve the same range. However, battery-operated underwater sensor nodes are energy-constrained and require that they transmit with low power to conserve power. Clustering is a technique for partitioning wireless networks into groups where a local base station (cluster head) is only one hop away. Due to the proximity to the cluster head, sensor nodes can lower their transmitting power, thereby improving the network energy efficiency. This paper describes the implementation of a new clustering algorithm to prolong the lifetime of WUSNs. We propose a new protocol called distance- and energy-constrained *k*-means clustering scheme (DEKCS) for cluster head selection. A potential cluster head is selected based on its position in the cluster and based on its residual battery level. We dynamically update the residual energy thresholds set for potential cluster heads to ensure that the network fully runs out of energy before it becomes disconnected. Also, we leverage the elbow method to dynamically select the optimal number of clusters according to the network size, thereby making the network scalable. Our evaluations show that the proposed scheme outperforms the conventional low-energy adaptive clustering hierarchy (LEACH) protocol by over 90% and an optimised version of LEACH based on *k*-means clustering by 42%.**



***Index Terms*— Acoustic communication, autonomous underwater vehicles, clustering, k-means, internet of underwater things, sensor networks, underwater networks, wireless sensor networks, wireless underwater sensor networks.**

I. Introduction

ENERGY consumption is a crucial factor in the design of wireless sensor networks since the sensor nodes are usually battery-powered and have a finite energy supply. Energy efficiency is considered one of the most important metrics in evaluating the performance of wireless sensor networks [1], especially in wireless underwater sensor networks (WUSNs) where it is difficult to replace or recharge sensor batteries. Data transmission is the most energy-intensive operation for sensor nodes in wireless sensor networks [2]. Thus, reducing energy consumption requires using lower transmission powers or making fewer transmissions. However, high transmission powers are usually required to overcome the hostile nature of the underwater channel to wireless signals, especially to reach remote base stations. Clustering is a useful technique for addressing this challenge and extending the lifetime of WUSNs [2], [3]. This is because clustering partitions the network into groups; the sensor nodes in each group send their data to a local base station called a cluster head. This allows the sensors to lower their transmission powers due to the proximity of the cluster heads, thereby reducing their energy usage and prolonging the network lifetime. Cluster heads are elected from among the sensor nodes to coordinate data collection from sensors within their clusters. They use multi-access approaches such as time-division multiple access (TDMA) or code-division multiple access (CDMA) to avoid collisions when collecting data from each sensor node. The nodes turn off their radio transmitters when it is not their turn to transmit.

Manuscript received December 16, 2020; revised January 12, 2021; accepted January 12, 2021. Date of publication January 27, 2021; date of current version March 5, 2021. This work was supported by the Petroleum Technology Development Fund of the Federal Republic of Nigeria. The associate editor coordinating the review of this article and approving it for publication was Prof. Yongqiang Zhao. *(Corresponding author: Kenechi G. Omeke.)*

Kenechi G. Omeke, Shuja Ansari, Lei Zhang, Qammer H. Abbasi, and Muhammad Ali Imran are with the James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: k.omeke.1@research.gla.ac.uk; shuja.ansari@glasgow.ac.uk; lei.zhang@glasgow.ac.uk; qammer.abbasi@glasgow.ac.uk; muhammad.imran@glasgow.ac.uk).

Michael S. Mollel is with the Nelson Mandela African Institute of Science and Technology, Arusha 447, Tanzania, and also with the Department of Systems, Power and Energy, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: michaelm@nm-aist.ac.tz).

Metin Ozturk is with the Department of Electrical and Electronics Engineering, Ankara Yıldırım Beyazıt University, 06010 Ankara, Turkey, and also with the Department of Systems, Power and Energy, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: metinozturk@ybu.edu.tr).

Digital Object Identifier 10.1109/JSEN.2021.3054943

Cluster heads also aggregate the collected data by removing duplicates and transmit them to the base station. The cluster heads are routinely rotated to ensure uniform dissipation of energy.

Sensor nodes in a network can be selected to serve as cluster heads in a centralised or distributed manner. The former uses a base station to coordinate cluster head selection whereas the latter is self-organised. Some distributed protocols include the low-energy adaptive clustering hierarchy (LEACH) [2], hybrid energy-efficient distributed clustering (HEED) [3], distributed underwater clustering scheme (DUCS) [4], etc. Increasingly, machine learning is employed to partition the network into clusters from which cluster heads are selected according predefined criteria. This can be done using methods such as k-means [5]–[7] and fuzzy c-means [7], etc., which find increasing use in wireless sensor networks, internet of things and crowd sensing applications. For such clustered networks, AUVs can be used to collect data from the network to avoid the high-power transmissions required to reach remote base stations [6], [8].

The manual cluster head selection schemes above have unavoidable limitations; for instance, in the LEACH protocol, multiple cluster heads can be located next to one another during a transmission round, which leaves other regions of the network with poor coverage. Also, LEACH has no way of ensuring that a dying node is not selected as cluster head. Similarly, there is excessive overhead data exchange during cluster head selection in the DUCS protocol proposed in [4]. This high message overhead lowers the energy and spectral efficiencies of the network. Besides, all the aforementioned algorithms are centered around sending collected data to a distant base station.

In this work, we propose a new protocol called distance- and energy-constrained k-means clustering scheme (DEKCS) for clustering, cluster head selection and data retrieval to prolong the lifetime of wireless underwater sensor networks. DEKCS uses the k-means algorithm for clustering but unlike previous works based on k-means [5], [9] where the node nearest to the k-means centroid is selected as cluster head, DEKCS selects the node that is closer to most nodes in the network as cluster head. This rule for selecting nodes closer to most other nodes as cluster heads has been termed the *proximity rule* because it ensures that nodes in each cluster are always in close proximity to their cluster head, thereby allowing them to transmit with much lower powers. This change in how the cluster head is selected significantly affects the energy consumption of the network, as we show later in Section III. In addition, to ensure scalability in the network design, we modify the elbow method [10], [11] to make it dynamic in choosing the optimal number of clusters as the network size changes (due to the addition or removal of sensor nodes). AUV data collection significantly improves WUSNs lifetime [12]. It finds common application in the offshore energy industry [13]. AUVs eliminate the hot spot problem which arises when multiple sensor nodes employ the same intermediate nodes (usually closer to the base station) to relay their data, causing such nodes to die more rapidly.

To the best of our knowledge, this is the first work that deals with AUV-based data collection from underwater clusters in which the cluster heads are selected based on their distance from all nodes within the cluster and their residual energies. This work is also the first to set adaptive energy thresholds for selecting cluster heads to ensure higher utilisation of the network energy. Thus, the contributions of this work are as follows:

- We propose a new clustering scheme called DEKCS that prolongs network lifetime. This is based on the *proximity rule* and residual energy of sensor nodes.
- We combined a new clustering scheme (DEKCS) with AUV-based data collection to prolong underwater network lifetime.
- We implemented underwater channel inversion for estimating the required transmit power and prevent energy wastage through power control.

The remainder of this paper is structured as follows. Section II examines the underwater channel model based on RF signaling, the energy consumption model and AUV communication model. In Section III, we introduce the DEKCS protocol, which uses the dynamic elbow method to determine the optimal number of clusters required and selects cluster heads based on the position of the nodes in a cluster and their residual energy levels. We evaluate the performance of the proposed algorithm in Section IV, in terms of overall network energy consumption and compare the results with two benchmark algorithms based on LEACH and LEACH k-means clustering [5]. Finally, conclusions are drawn in Section V.

II. System Model

We consider a two-dimensional static wireless underwater sensor network deployed to monitor a subsea oil production field or a marine protected zone, such as the East of Garner and Montrose fields operated by BP exploration operating company, off the coast of United Kingdom, shown in Fig. 1. We model the two-dimensional distribution of sensor nodes used for monitoring as a Poisson point process represented by the parameter, λ. These sensors are deployed to monitor flow rate, pipeline pressure, water salinity, light penetration, chemical pH, etc. They are grouped into clusters to facilitate data reporting through cluster heads selected from amongst them. An AUV is deployed from a floating production storage and offloading (FPSO) platform on the water surface to visit the network location and retrieve data collected by the underwater sensors. The frequency with which the AUV visits the network location depends on the reporting rate of the network, a parameter which varies from application to application. The AUV specifications can be found in [14]. The AUV is loaded with a map of the undersea network in terms of the depth coordinates of the clusters [8]. The AUV battery can be recharged or replaced at the FPSO after a few cycles. During each visit, the AUV hovers in a position that allows it to provide simultaneous coverage to multiple clusters.

A. Underwater Channel Model

The channel experienced by the network during data transmission can be obtained directly from first principles

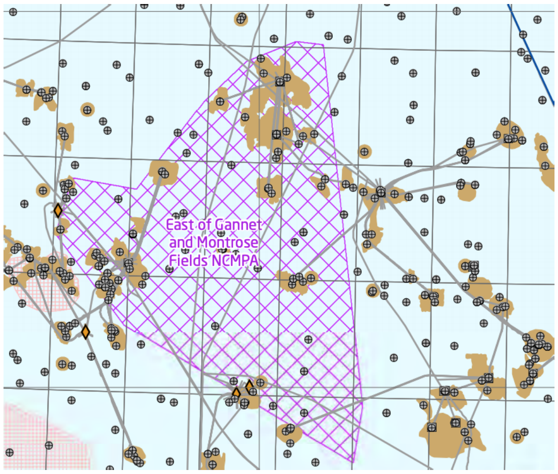


Fig. 1. Typical distribution of subsea facilities in real-life applications showing that applications that wireless sensors are deployed to monitor are seldom uniformly distributed, justifying the use of a Poisson point process to distribute sensor nodes in this paper[1].

using Maxwell's equations. For an RF-based underwater network, the propagation channel between the transmitter and receiver can be modelled by the exponentially attenuating wave model [15], whereby the transmitted power falls off exponentially with distance away from the transmitter.

From Maxwell's equations, the electric field component of a linearly polarised plane electromagnetic wave travelling in the z-direction can be expressed [15] as $E_x = E_0 \exp^{(j\omega t-\gamma z)}$, where γ is the propagation constant which comprises the attenuation and phase constants

$$\gamma = \text{beta-j}\alpha = \omega\sqrt{\mu\varepsilon\left(1-j\frac{\sigma}{\omega\varepsilon}\right)}. \tag{1}$$

The field can be resolved into radial and tangential components, whereby the tangential component can be expressed in terms of the magnitude and phase as [16]

$$E_\theta = E(\omega, r)e^{-\alpha r}e^{-j(\beta r+\vartheta(\omega,r))}. \tag{2}$$

The received signal, $Y(d, f)$ is related to the transmitted signal [16], $X(d, f)$ by

$$Y(d, f) = H(d, f)\,X(f), \tag{3}$$

where d is the transmission range and f is the operating frequency. These are the two major factors that affect RF signal propagation underwater. The magnitude of the channel response decays exponentially with distance for a fixed frequency. For a given transmitter-receiver distance, the channel transfer function $H(d, f)$ can be simplified [16] as

$$H(d, f) = H_0 \mathrm{e}^{-\alpha(f)d}\mathrm{e}^{-\mathrm{j}\theta(f)}, \tag{4}$$

where H_0 and $\theta(f)$ represent the channel gain at DC and the channel phase, respectively. This model does not account for spreading losses, which is usually far less than absorption losses.

The channel attenuation constant, α is given [17] by

$$\alpha = \omega\sqrt{\mu\varepsilon}\left\{\frac{1}{2}[\sqrt{1+\left(\frac{\sigma}{\omega\varepsilon}\right)^2}-1]\right\}^{1/2} \text{Np/m}, \tag{5}$$

where $\alpha(\text{Np/m})| = \frac{1}{8.68}|\alpha(\text{dB/m})|$; ω, μ, ε and σ represent the angular frequency $(2\pi f)$, magnetic permeability, dielectric permittivity and electrical conductivity, respectively. For frequencies where $\sigma \gg \omega\varepsilon$ (this holds true for all megahertz frequencies), the attenuation and phase constants can be simplified [16] to $\alpha = \beta = \sqrt{\pi\sigma\mu_0}\sqrt{f}$.

The received power can be obtained by modifying the Friis equation to account for the nature of the underwater channel [18], [19]

$$P_r = P_t G_t G_r\left(\frac{\kappa\lambda}{4\pi d}\right)^2 \exp(-2\alpha d), \tag{6}$$

where λ is the wavelength.

Equation (6) can be expressed in dB as $P_r = P_t + G_r + G_t - L_p$, where L_p is underwater pathloss given by $L_p = L_\beta + L_\alpha$. L_α is the attenuation loss, given [17] by $L_\alpha = 10\log(e^{2\alpha d}) = 8.69\alpha d$, L_β is called the attenuation loss due to the difference in the wavelength of the signal in air and underwater [19] and is given [18] by

$$L_\beta = 20\log(4\pi d/\lambda). \tag{7}$$

For RF-based systems, attenuation losses dominate other losses in underwater environments.

B. Energy Consumption Model

Underwater sensor nodes expend energy for sensing, packet transmission, packet reception, data processing, network maintenance, staying awake, etc. The power consumed to transmit a packet is related to the packet size and transmission distance [20]. It includes power consumed by the radio electronics and the power consumed by the transmitter power amplifier. For the receiver, only power consumed in the radio electronics is considered relevant in receiving a packet [2]; the value of this power is constant for a given packet size.

To transmit an m-bit message at a distance d, the transmitter requires an amount of power, $E_{TX}(m, d)$ [2] given by

$$\begin{aligned} E_{TX}(m, d) &= E_{TXelec}(m) + E_{TXamp}(m, d) \\ &= mE_{\text{elec}} + m\epsilon_{\text{uw}}d^2, \end{aligned} \tag{8}$$

where E_{TXelec} is the power dissipated in the transmitter electronics, E_{TXamp} is the power dissipated in the transmitter power amplifier and ϵ_{uw} is the power dissipated by the transmit amplifier to maintain an acceptable E_b/N_0 underwater. Note that only the direct pathloss is considered here since the nodes are close to the cluster head or the base station, and there are no objects to cause significant obstructions; hence, multipath losses can be ignored. The energy required to receive a packet is constant. It is given [2] by

$$E_{Rx}(m) = E_{Rx-\text{elec}}(m) = mE_{\text{elec}}, \tag{9}$$

[1]Retrieved from https://www.bp.com/content/dam/bp/country-sites/en_gb/united-kingdom/home/pdf/BP-Vorlich-Environmental-Statement-April2018.pdf. ©Crown Copyright. Contains public sector information licensed under Open Government Licence v3.0.

where E_{RXelec} is the power dissipated in the receiver electronics.

For the transmit amplifier, the energy expended is a function of how far the transmitter is away from the receiver, and how much power is required to achieve an acceptable E_b/N_0 [2]. In each data frame, cluster heads consume energy to receive transmissions, aggregate the data and transmit them to the AUV base station. Non-cluster head nodes only consume energy to transmit to the cluster head once per frame.

The major source of energy dissipation for sensor nodes is in uploading data to the base station. As mentioned earlier, channel inversion is exploited to implement power control for all nodes; hence, each transmitting sensor node can use the lowest transmit power that guarantees an acceptable E_b/N_0 at the receiver. Thus, our scheme considers both the nature of the channel and the distance between the transmitter and receiver. Energy consumption models for acoustic communication can be found in [21] for different water depths.

The nodes are stationary, since they are fixed to the facility they are deployed to monitor. The communication link between a transmitting node and a receiving node is assumed to be symmetric. We assume that nodes always have data to transmit; the cluster head uses TDMA to schedule transmissions for each node.

C. AUV Data Collection Model

An AUV is deployed from an FPSO anchored at the ocean surface to collect data from the underwater sensor network. Since AUVs are battery-operated, it is necessary to reduce their energy expenditure by minimising their travel time while maximising the data collected [22]. The AUV is supplied with the network map as input to enable it to locate the first cluster head [14]. The AUV has a communication radius within which it can reliably gather data from the sensor nodes [23]. We define h as the altitude that maximises the communication link between the AUV and an underwater cluster head, which can be expressed as $h = r\tan(\phi)$, where r is the maximum coverage radius the AUV can provide and ϕ is the elevation angle.

III. The DEKCS Protocol

The implementation of the DEKCS algorithm is divided into three stages. The first stage involves clustering via the k-means clustering algorithm. The second deals with selection of cluster heads. This stage takes into account the residual energy of the nodes and their locations (with respect to other nodes) within their cluster. DEKCS implements a new metric by choosing the sensor node closest to all other nodes as cluster head, instead of the node closer to the centroid. The final stage involves data transmissions between nodes and the cluster head within the clusters. Finally, the available data is collected from the cluster heads using an AUV deployed from an FPSO on the surface of the ocean. We adopted k-means clustering in this work because it is an optimised algorithm suitable for non-uniform network distributions. If the sensor nodes were uniformly distributed, a machine learning algorithm is not necessarily required to cluster them, since it is simpler to manually partition the network. We adopted a Poisson point process for our network distribution in this work; however, k-means is a generalised optimisation algorithm and should work irrespective of the distribution.

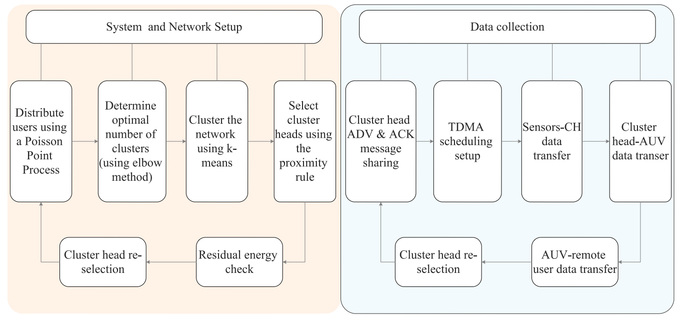


Fig. 2. Block diagram of the DEKCS protocol showing the two broad divisions of the proposed system model implementation: the setup blocks showing how the model is organised and the data collection blocks showing the flow of data from the sensors to the sink.

A. k-Means Clustering

The k-means algorithm partitions an unlabelled multidimensional data set into a set of k clusters, $C_i = \{C_1, C_2, \ldots, C_{k_i}\}$, where the desired number of clusters, k also corresponds to the number of cluster heads and C_i represents the i-th cluster. The number of clusters is determined *a priori*. Given a sensor network with N sensor nodes, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, the goal of the k-means algorithm is to partition the network into a set of non-overlapping clusters so as to minimise the intra-cluster distance between the nodes and the centroid of the cluster while maximising the inter-cluster distance [7]

$$J_{\min} = \sum_{j=1}^{k} \sum_{c_n \in C_i} \left\| \mathrm{x_i} - \mu_j \right\|^2, \tag{10}$$

where each cluster, C_i contains N_j nodes, x_i represents the i-th node in the network; μ_j represents the geometric centroid of the sensor nodes in a given cluster [24] and is given by

$$\mu_j = \left(\frac{1}{N} \sum_{i=1}^{N} x_i, \frac{1}{N} \sum_{i=1}^{N} y_i \right). \tag{11}$$

In this work, the cluster sizes are designed to ensure that each node requires only one-hop communication to reach its cluster head. We also take into account the severe restriction on the transmission range imposed by the underwater channel, especially for RF-based communication.

1) Optimal Number of Clusters: The efficiency of the k-means algorithm relies on choosing the optimum number of clusters for the network. This can be done using the elbow method. The LEACH protocol also has a legacy scheme for deciding the optimal number of clusters. For the elbow method, the optimal number of clusters is found by calculating the intra-cluster sum of squares (ISS) for the data set, where

$$\mathrm{ISS} = \sum_{i=1}^{n} (x_i - c_i)^2. \tag{12}$$

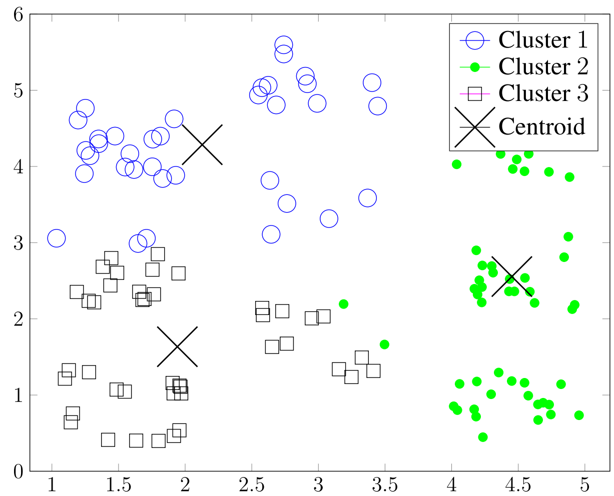


Fig. 3. Network distribution showing cluster centroids. This figure shows that there are regions with dense deployments and regions with sparse deployments, which has implications on the cluster head selection protocol used.

The optimal number of clusters k_{opt} is found in LEACH [2] using

$$k_{\text{opt}} = \frac{\sqrt{N}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{\text{fs}}}{\varepsilon_{\text{mp}}}} \frac{M}{d^2_{\text{AUV}_{\text{BS}}}}, \tag{13}$$

where N is the number of nodes in the network, M is the area of the network, ε_{mp} and ε_{fs} are the transmit power amplifier energy consumption for direct and multipath transmissions, respectively. The distance between the cluster head and the AUV base station is given by d_{AUV_BS}. The elbow method was used in this work to determine the optimal number of clusters. We introduce a dynamic approach to the elbow method that adaptively chooses the optimal number of cluster heads as the network size changes.

B. Cluster Head Selection

In this work, clustering is performed before cluster head selection to reduce the energy spent in the cluster formation process. The cluster head selection policy must satisfy two conditions:

1) In-Cluster Position: The node that is *nearest to most other nodes within the cluster* is selected as cluster head, instead of the node nearest to the centre of the cluster [5], [7], [11]. This condition, which we term *the proximity rule*, is more useful than proximity of the potential cluster head to the centre of the cluster since our goal is to minimise energy consumed by sensors for transmitting to the cluster head, not to select the node at the centre of cluster. The proximity rule is important since the sensors are not uniformly distributed and will ensure that the selected cluster head comes from a region with more dense deployment as it will be closer to more sensor nodes. However, it comes at the cost of proportional fairness, as the selected cluster head will likely be far from regions with sparse sensor deployment.

To find the sensor that is nearest to most other nodes and requires the minimum energy to transmit to within its cluster,

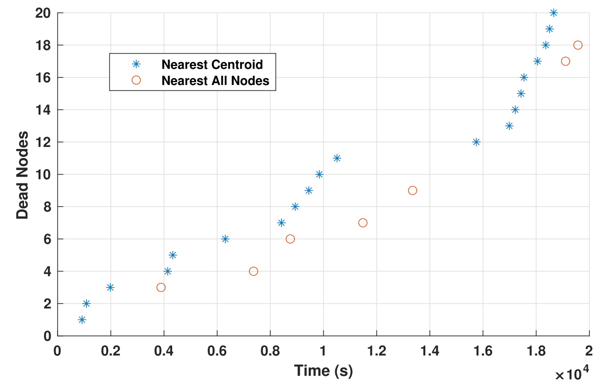


Fig. 4. Scatter plot showing the impact of the cluster head selection policy on the energy efficiency of the network. The benchmark 'Nearest Centroid' scheme selects the node that is closest to the *k*-means centroid as the cluster head whereas the proposed scheme selects the node that is closest to all other nodes in the network as the cluster head, leading to fewer dead nodes per unit of simulation time than the benchmark.

we introduce a cost function, ζ that measures the Euclidean distance between the selected node and all other in-cluster nodes

$$\zeta = \sum_{j=1}^{k} \sum_{s_i \in C_j} \mathrm{d}\left(x_i, \mathrm{x}_j\right), \tag{14}$$

where the Euclidean distance $d\left(x_i, x_j\right)$ is given by

$$d\left(x_i, x_j\right) = \left\|x_i - x_j\right\|^2. \tag{15}$$

To show how the proximity consideration affects the energy performance of the network, we compared the energy consumption of the network for our approach and the conventional k-means approach. The DEKCS algorithm selects a qualifying node (one that meets the residual energy criterion) that is closest to all other nodes in the cluster as head, whereas the conventional approach selects the node nearest to the centroid identified by the k-means algorithm. A comparison of both approaches is shown in the scatter plot in Fig. 5, which indicates that under the same energy threshold conditions, using the proximity rule to select the cluster head results in fewer dead nodes per unit time, which will enable the network to last longer.

2) Residual Energy Level: For a node to qualify to be selected as cluster head, its residual energy must be above the set threshold. This condition is necessary to ensure that the cluster head does not die prematurely, which will cause the network to become disconnected. In this work, we propose setting a dynamic threshold for the residual energy, where nodes will qualify as cluster heads until the network dies entirely. When there exists a cluster for which no node meets the residual energy level condition to become a cluster head, the threshold is updated to ensure that the network does not disconnect. This criterion ensures that cluster heads can continue to be selected after the energy level of each node has dropped below the initial threshold but there is still enough energy left in the network to continue monitoring and reporting. This consideration enables the network to last much longer than conventional means, as the results show in Section IV. When the energy of a cluster head drops below the set threshold, it is demoted to an ordinary node to continue sensing activities. Dead nodes are also removed from the network and a new cluster head is chosen based on the position and residual energy conditions.

C. AUV-Based Data Collection

An AUV is deployed from an FPSO on the ocean surface to collect data from the underwater sensor network. Due to the proximity of the AUV to the network location, it can use high-speed RF link to collect data from the cluster heads. The AUV is positioned in such a way as to provide coverage for multiple clusters simultaneously to reduce the number of locations visited. We set a pathloss threshold that guarantees reliable data transfer, which implements a trade-off between coverage and reliability (more clusters can be covered if the threshold is set lower). The threshold is based on the receiver sensitivity and link channel conditions. In this work, we did not consider losses due to multipath fading because attenuation losses dominate fading losses in underwater environments.

IV. Performance Evaluation

We simulated the performance of the DEKCS algorithm in an underwater wireless sensor network based on RF signaling. However, since DEKCS is based on optimised computer algorithms, it should also work for acoustic signaling or other transmission techniques. The network parameters are chosen according to the channel and energy models set out in Section II. Each sensor node has an initial energy of 5 joules and is equipped with a radio and a flash memory [2]. The radio range is set to 2 m while the minimum required received power is set to −90 dBm; power control is used to assign just enough transmit power to guarantee this value at the receiver (channel inversion). The network size comprises an average of 200 sensor nodes deployed in an area of 400 m^2 following a Poisson Point Process. These parameters were carefully chosen considering the nature of the underwater channel and the resource limitations of the nodes. We used k-means to cluster the network and dynamically determine the number of clusters through the elbow method. Due to the small size of our model, the algorithm converged in less than a second both for identifying the cluster centroids and for selecting the optimal number of clusters (elbow method). We also set a dynamic energy threshold that varies with the overall residual energy in the network for cluster heads selection. The simulation was performed in rounds, where each round involves TDMA-scheduled data collection from all nodes within a cluster, aggregation of the data and collection of the aggregated data from cluster heads by the AUV base station. Our analysis was conducted in MATLAB (R2020a) and Python. The *proximity rule* was adopted for DEKCS after comparing with other cluster head selection criteria.

We assess the energy efficiency by evaluating the number of nodes that have died or remain alive in the network over time. Figure 5 shows the number of nodes that remain alive per simulation round for different algorithms. The termination criteria for the simulation is reached when there is a cluster

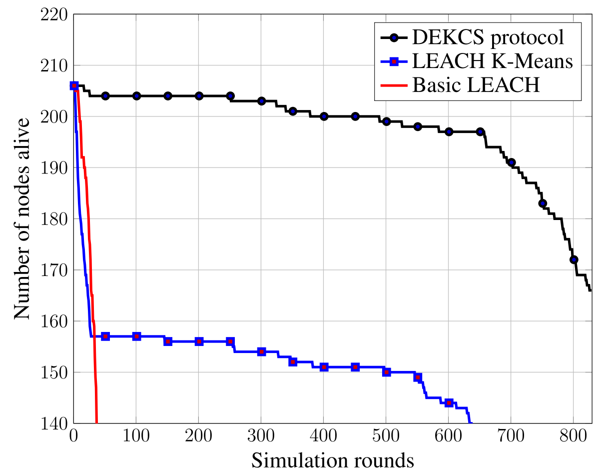


Fig. 5. Plot of the number of nodes alive in the network after it has become disconnected or non-functional. This shows the impact of the clustering policy on the energy efficiency of the network. More efficient the clustering algorithms prolong the network lifetime by reducing the transmit power of the nodes. In this figure, DEKCS keeps more nodes alive per unit time even though it allows more transmissions. The basic LEACH [2] algorithm terminates after only 80 simulation rounds, while LEACH k-Means [5] and DEKCS run for 640 and 830 rounds, respectively.

where there is no node that qualifies to be elected as a cluster head. It can be seen from the figure that under the same residual energy threshold conditions, DEKCS outlasts LEACH and LEACH-K [5] (based on k-means and LEACH) protocols before the network becomes disconnected. Under the same network conditions, the termination criteria is reached after only 80 rounds for LEACH, compared to over 800 rounds for the DEKCS algorithm. For instance, when we set the minimum energy threshold to become a cluster head as 3 Joules, the network becomes disconnected after only 80 rounds for LEACH, with about 49% of the nodes still dead. To ensure consistent comparison, we set the same conditions for the DEKCS algorithm and it ran for 830 rounds before the energy threshold condition is reached, with only 20% of the nodes dead. This shows that the proposed DEKCS algorithm is very energy efficient and requires less energy per round due to the proximity of the cluster heads to the nodes. By considering the area under the curve in Figure 5, we show that the DEKCS algorithm outperforms the LEACH algorithm by 90.5% and the optimised k-means algorithm based on LEACH by 41.2%. The performance disparity is because the LEACH algorithm selects cluster heads randomly without taking into account their position in the network or their residual energy level. LEACH k-means selects nodes that are closest to the centroid of their cluster as cluster heads, but does not consider the residual energy of the nodes in most implementations.

Even in LEACH and k-means implementations that consider the residual energy of nodes, selecting centrally placed nodes may result in poor performance in the network as outlier nodes will cause nodes closer to them to be selected, making all other nodes to waste energy in long transmissions. In our scheme, we sacrifice proportional fairness for overall network energy efficiency by selecting nodes that have the minimum distance to all other nodes. This condition may cause outlier nodes to

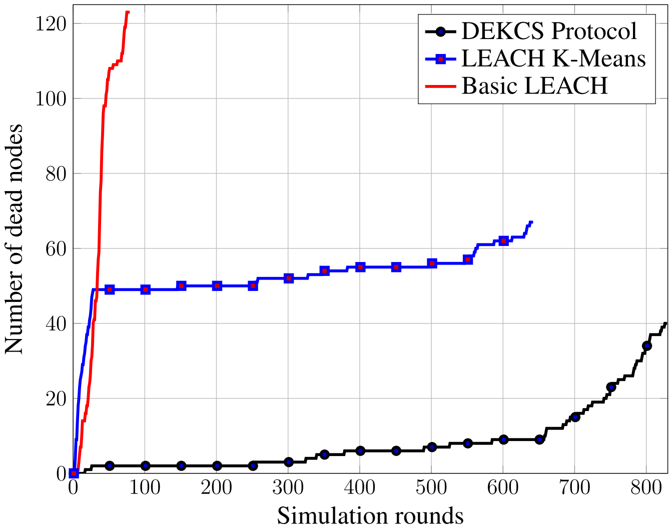


Fig. 6. Plot of network lifetime considering number of dead nodes over time. The proposed algorithms lead to fewer dead nodes per simulation time compared to the other algorithms. This enables the network to run for longer before becoming disconnected.

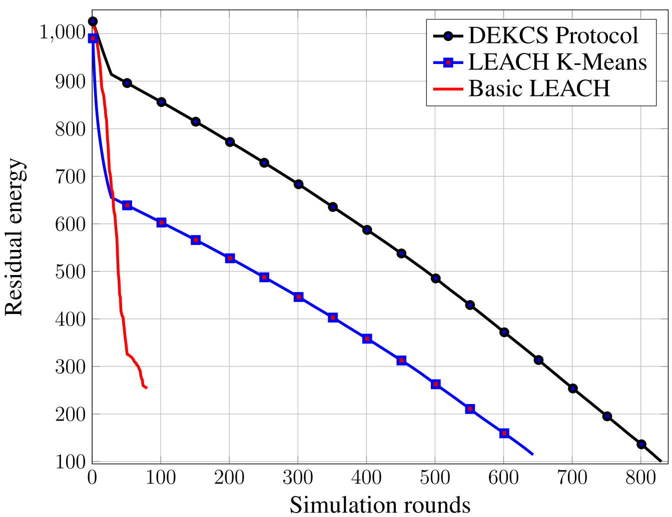


Fig. 7. Plot of the residual energy in the network after it has become disconnected or non-functional showing the impact of the clustering policy on the energy efficiency of the network. The proposed DEKCS algorithm keeps the network running longer than the benchmark algorithms and ensures that the network energy is nearly all used up before the it is disconnected.

run out of battery quickly but will improve the longevity of the network, as clearly shown in Fig. 5.

It is often necessary to show how the sensors in the network die over time. In Fig. 6, we compare the number of dead nodes for different algorithms. It is clearly seen that the DEKCS algorithm outperforms other algorithms under similar network conditions, which allows the network to run for far longer with fewer nodes dying.

Residual energy is another important criterion used in judging the energy efficiency of a wireless sensor network [25]. The residual energy shows the total amount of energy left unused in the network after it has become non-functional. We show in Fig. 7 how the DEKCS algorithm compares with other algorithms in terms of network residual energy.

The lower the residual energy in the network at the point it becomes disconnected, the better and more efficient the clustering algorithm. Figure 7 shows that the DEKCS algorithm prolongs the operational life of the network by ensuring that all or most of the energy in the network is used up before the network disconnects. This is achieved through three techniques: the cluster head selection process based on the *proximity rule* allows more nodes to transmit with lower powers and conserve energy. Secondly, setting dynamic energy thresholds ensures continued operation until some clusters become disconnected, which is the only time the minimum energy threshold is reached. Finally, iteratively updating the elbow method ensures that the optimal number of clusters is selected as the network size changes, which is important in underwater sensor networks based on RF signaling due to the limited transmission range. Since edge nodes are likely to die first, the network size keeps shrinking over time, requiring fewer cluster heads. Also, by accommodating changes in the network size, the DEKCS algorithm is well suited for large networks.

Static minimum thresholds cause the network to terminate once the battery level of all nodes falls below the set threshold, even if there is still enough energy remaining in the network to continue operation. The DEKCS algorithm addresses this shortcoming through dynamic threshold setting that changes with the network size. We showed that relaxing the rigid energy threshold condition significantly improves the network energy utilisation and prolongs its operational lifetime. Our results indicate that the energy in the network is used up to 90% every time.

This paper has assumed that the sensor nodes are fixed in position at the seabed. However, it is challenging to anchor sensor nodes to the seabed in practical deployments due to the impact of the water current that might cause swaying for the nodes. Water current can also introduce errors in estimating the position of the AUV.

V. Conclusion

In this paper, we proposed a new clustering protocol called DEKCS for clustering wireless sensor networks. DEKCS combines k-means, node position and leftover power in choosing a cluster head during clustering while an AUV is used for data collection. These conditions guarantee that the nodes transmit with the minimum power possible by ensuring that the cluster head is in close proximity. More nodes thereby conserve their power, which improves the network energy efficiency. In addition, the network was set up to ensure that nearly all of its energy is used up before it becomes disconnected, as shown by the residual energy efficiency evaluation. We conducted performance evaluations to test the proposed protocol and our results showed that it leads to fewer dead nodes per unit time and uses up more of the network energy than previous algorithms. Our future work will consider the trade-offs between coverage, delay and reliability for AUV-aided data collection in WUSNs.

References

[1] I. Dietrich and F. Dressler, "On the lifetime of wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 5, no. 1, pp. 1–39, 2009.

[2] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.

[3] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 366–379, Oct. 2004.

[4] M. C. Domingo and R. Prior, "A distributed clustering scheme for underwater wireless sensor networks," in *Proc. IEEE 18th Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2007, pp. 1–5.

[5] R. Gantassi, B. B. Gouissem, and J. B. Othmen, "Routing protocol leach-K using K-means algorithm in wireless sensor network," in *Web, Artificial Intelligence and Network Applications*, L. Barolli, F. Amato, F. Moscato, T. Enokido, and M. Takizawa, Eds. Cham, Switzerland: Springer, 2020, pp. 299–309.

[6] S. Cai, Y. Zhu, T. Wang, G. Xu, A. Liu, and X. Liu, "Data collection in underwater sensor networks based on mobile edge computing," *IEEE Access*, vol. 7, pp. 65357–65367, 2019.

[7] J. Qin, W. Fu, H. Gao, and W. X. Zheng, "Distributed *k*-means algorithm and fuzzy *c*-means algorithm for sensor networks based on multiagent consensus theory," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 772–783, Mar. 2017.

[8] M. T. R. Khan, S. H. Ahmed, and D. Kim, "AUV-assisted energy-efficient clustering in underwater wireless sensor networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[9] M. Rida, A. Makhoul, H. Harb, D. Laiymani, and M. Barhamgi, "EK-means: A new clustering approach for datasets classification in sensor networks," *Ad Hoc Netw.*, vol. 84, pp. 158–169, Mar. 2019.

[10] M. Huang, K. Zhang, Z. Zeng, T. Wang, and Y. Liu, "An AUV-assisted data gathering scheme based on clustering and matrix completion for smart ocean," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9904–9918, Oct. 2020.

[11] P. Bholowalia and A. Kumar, "EBK-means: A clustering technique based on elbow method and K-means in WSN," *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 1–8, 2014.

[12] J. Yan, X. Yang, X. Luo, and C. Chen, "Energy-efficient data collection over AUV-assisted underwater acoustic sensor network," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3519–3530, Dec. 2018.

[13] I. Jawhar, N. Mohamed, J. Al-Jaroodi, and S. Zhang, "An architecture for using autonomous underwater vehicles in wireless sensor networks for underwater pipeline monitoring," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1329–1340, Mar. 2019.

[14] I. Vasilescu, K. Kotay, D. Rus, M. Dunbabin, and P. Corke, "Data collection, storage, and retrieval with an underwater sensor network," in *Proc. 3rd Int. Conf. Embedded Netw. Sensor Syst.*, San Diego, CA, USA, 2005, pp. 154–165.

[15] A. I. Al-Shamma'a, A. Shaw, and S. Saman, "Propagation of electromagnetic waves at MHz frequencies through seawater," *IEEE Trans. Antennas Propag.*, vol. 52, no. 11, pp. 2843–2849, Nov. 2004.

[16] A. Zoksimovski, D. Sexton, M. Stojanovic, and C. Rappaport, "Underwater electromagnetic communications using conduction—Channel characterization," *Ad Hoc Netw.*, vol. 34, pp. 42–51, Nov. 2015.

[17] S. Jiang and S. Georgakopoulos, "Electromagnetic wave propagation into fresh water," *J. Electromagn. Anal. Appl.*, vol. 3, no. 7, p. 261, 2011.

[18] K. P. Hunt, J. J. Niemeier, and A. Kruger, "RF communications in underwater wireless sensor networks," in *Proc. IEEE Int. Conf. Electro/Inf. Technol.*, May 2010, pp. 1–6.

[19] M. C. Domingo, "Magnetic induction for underwater wireless communication networks," *IEEE Trans. Antennas Propag.*, vol. 60, no. 6, pp. 2929–2939, Jun. 2012.

[20] G. Han, S. Shen, H. Song, T. Yang, and W. Zhang, "A stratification-based data collection scheme in underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10671–10682, Nov. 2018.

[21] A. Sehgal, C. David, and J. Schonwalder, "Energy consumption analysis of underwater acoustic sensor networks," in *Proc. OCEANS MTS/IEEE KONA*, Sep. 2011, pp. 1–6.

[22] G. A. Hollinger *et al.*, "Underwater data collection using robotic sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 5, pp. 899–911, Jun. 2012.

[23] B. Zhang, Y. Wang, H. Wang, X. Guan, and Z. Zhuang, "Tracking a duty-cycled autonomous underwater vehicle by underwater wireless sensor networks," *IEEE Access*, vol. 5, pp. 18016–18032, 2017.

[24] G. Y. Park, H. Kim, H. W. Jeong, and H. Y. Youn, "A novel cluster head selection method based on K-means algorithm for energy efficient wireless sensor network," in *Proc. 27th Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Mar. 2013, pp. 910–915.

[25] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 976–978, Nov. 2005.