Dynamic Clustering and Data Aggregation for the Internet-of-Underwater-Things Networks

Kenechi G. Omeke*, Michael Mollel*, Syed Tariq Shah*, Kamran Arshad†, Lei Zhang*, Qammer H. Abbasi* and Muhammad Ali Imran*

*James Watt School of Engineering, University of Glasgow, Glasgow, United Kingdom
† Ajman University, United Arab Emirates

***Abstract*—Advances in semiconductor technology have made it possible to have high processing powers in cheap microcontrollers, which is spawning off a revolution in the range of applications of the Internet-of-Things (IoT) and its underwater counterpart, the Internet-of-Underwater-Things (IoUT). As a result, it has now become possible and cost effective to implement powerful data processing algorithms on very cheap microcontrollers and achieve network intelligence on edge devices. In this paper, we evaluate the impact of implementing an unsupervised machine learning technique based on the k-means algorithm, as well as data aggregation, on the performance of a wireless underwater sensor network. A clustering algorithm based on the k-means algorithm is used to divide the network into clusters and to select cluster heads based on network topology and residual energy. Each cluster head collects and aggregates data from nodes within its cluster's coverage and forwards the data to the sink. The network is deployed in a shallow seabed, and it is assumed that the nodes can reach the sink using their full transmission powers. Hence, the performance evaluation compares the sum-throughput, energy efficiency and coverage probability for direct transmissions to the sink against transmissions using the cluster heads. We also propose a special consideration for disaster early warning data, which packets are assigned priority delivery and handled with minimum delay. The evaluation is performed through computer simulations and the results show over a 100% improvement in throughput for cluster-based transmissions compared to direct transmissions.**

I. Introduction

There is an ever-growing reliance on the oceans for life on earth. As the global population continues to grow and the energy requirements of modern life skyrockets, the oceans play a vital role in the supply of oxygen, food, mineral and energy resources to support human activities on earth and to clean the atmosphere of excess carbon and regulate global temperatures. However, climate change continues to threaten the oceans and life on land. To understand the mechanism of marine climate change and proffer suitable solutions, there is a need to monitor ocean climatic conditions (as a predictor of atmospheric conditions on land) and marine life. As more resources are deployed offshore to produce energy, there is also a need to monitor and control such offshore assets. Wireless sensor networks are the cheapest and most robust solution for gathering high-quality and quantity of marine data without interruption.

However, wireless underwater communication is fraught with difficulties, which makes gathering data from underwater networks very challenging. Acoustic communication, which is the primary technology used for this purpose, is slow, prone to errors, have very limited bandwidth and consumes a lot of power. Radio and optical waves are heavily absorbed by water molecules, severely limiting their transmission ranges. Magnetic induction also have very poor range and low bandwidth. Despite these technological limitations, power consumption is the biggest threat facing wireless sensor networks deployed underwater and in terrestrial applications as battery technology has not kept pace with developments in sensing, processing and communications technologies. As a result, energy consumption is the most important concern for Internet-of-Underwater-Things (IoUTs) networks, especially those deployed long-term due to the difficulty of replacing sensor batteries underwater.

Some major schemes for improving energy efficiency in IoUTs include clustering, data aggregation and use of relays to reduce transmission distances for collected data. Clustering divides the network into clusters so that each node sends its collected data to the cluster head, which is usually closer than the sink, thereby minimising the power used for data transmissions [1]. Data aggregation is an intermediate processing stage whereby redundancies in the data collected from each cluster is removed before sending them to the sink. This significantly improves the energy efficiency of the network because the power required for data processing is far less than the power required for packet transmissions. Finally, relaying reduces transmission distances by using intermediate nodes between the data source and sink to forward data. The data source forwards packets containing collected data to its nearest neighbour in the direction of the sink, which in turn forwards it to its nearest neighbour until it gets to the sink. Since the nearest neighbours are usually closer to the source than the sink, lower transmission powers can be used, thereby improving energy efficiency as well as network performance through spatial diversity.

Data aggregation is often performed at intermediate nodes such as cluster heads. It involves combining spatially correlated data generated by different sources within a region of interest to avoid redundancies or duplicates in the collected information so as to reduce the amount of data transmitted to the sink, thereby improving the energy efficiency of the network. Strong correlation exists between the data collected by IoUT nodes deployed within the same region, so sending all the generated packets back to the sink wastes resources. Data aggregation is an important tool in practical wireless sensor networks because of their energy-limited nature. It has been shown that effective data aggregation improves the network energy efficiency and prolongs the network lifetime [2].

An IoUT node comprises a suite of different sensors, such as those used for monitoring temperature, dissolved oxygen, pH, salinity and pressure sensors, etc, depending on the application. The sensors are all connected to and controlled by a processor, such as a microcontroller that serves as the brain of the node. The data generated by each sensor is tagged with the sensor's identity (ID) to differentiate data generated by different types of sensors within the node. In addition, each node in a cluster is also tagged with the node ID. In some applications, it is also possible to geo-tag the data if location information is available. At the cluster head, a data aggregation function is used to filter out duplicates using the IDs from the packets collected by individual nodes. Data aggregation functions can be linear or non-linear and use statistics in the collected data to enhance data quality, remove redundancies and reduce the size of packets sent to the sink [3]. For instance, the mean, minimum, maximum, etc observations can be used to obtain a simple average for each sensed parameter, which is then sent to the sink. Other techniques, such as principal component analysis, analysis of variance (ANOVA), Gaussian Processes (GP), self-organising maps (SOM), adaptive learning vector quantization (ALVQ), etc. can also be used to perform data aggregation to obtain accurate information about the sensed parameter. The reader is referred to the literature in [4]–[6] for a more in-depth coverage of data aggregation techniques, including information security considerations. For underwater sensor networks, a tree-based data aggregation scheme was presented in [7] while a review of data aggregation techniques can be found in [8].

A key question in data aggregation is how to structure it in order to achieve the best results. The approaches that have been proposed include cluster-based aggregation, tree-based, centralised and in-network aggregation [3]. In the cluster-based approach, data aggregation is performed at cluster heads. The tree-based approach is a hierarchical structure that uses spanning trees, with data aggregation performed at parent nodes (any node that other nodes connect to in order to reach the sink) all the way up to the root (sink) of the tree. In the centralised approach, a designated aggregator (a node equipped with more computing and battery resources) is used to collect data from other nodes in the network, aggregate them and forward the result to the sink. In-network data aggregation involves processing at intermediate nodes for multi-hop sensor networks; it can be lossy aggregation, leading to packet size reduction or lossless aggregation without packet size reduction. We have summarised the merits and limitations of each approach based on common data aggregation metrics in Table I. A more in-depth review can be found in [3].

TABLE I
COMPARISON DATA AGGREGATION TECHNIQUES

Approach/ Parameters	Centralized	In Network	Tree	Cluster
Delay	Moderate	Moderate	Low	Low
Redundancy	Moderate	Moderate	Low	Low
Accuracy	Moderate	Low	Moderate	High
Traffic	High	High	Moderate	Low
Energy Consumption	High	Moderate	Low	Low

In addition to routine sensing and monitoring operations, we propose using the IoUTs network as an early warning system for natural disasters in the ocean. The network monitors spikes in temperature and pressure that could suggest an earthquake, similar to the Deep-ocean Assessment and Reporting of Tsunami (DART) system deployed by the National Oceanic and Atmospheric Administration. As a result, packets generated due to spikes from the temperature and pressure sensors are not aggregated but tagged as priority packets and sent immediately to the sink to avoid delays in reporting potential disasters.

This paper introduces a dynamic protocol for clustering wireless sensor nodes in an IoUTs network and performing data aggregation to reduce data redundancies to enhance the network lifetime. The protocol (called DYCADA for dynamic clustering and data aggregation) takes into account the network distribution and the residual energy of nodes before selecting them as potential cluster heads. The selected cluster head is responsible for data collection from nodes within its cluster, performing data aggregation to remove redundancies in the data and forwarding clean copies of the data to the sink. The overarching aim of the paper is to achieve efficient monitoring of underwater assets and offshore energy facilities in an energy-efficient manner. The contributions of this paper include;

- We propose a joint dynamic clustering and data aggregation (DYCADA) scheme to improve energy efficiency in wireless underwater sensor networks.
- An in-depth analysis of throughput and coverage probability is provided as a function of energy efficiency of the network considering direct transmissions to the sink and using cluster heads to perform data aggregation and forwarding of data from individual clusters.
- We propose tagging packets originating from early disaster warning systems as priority packets to minimise delay in forwarding them to the sink. Such packets are not aggregated but sent immediately to the sink to

reduce delays in warning of potential disasters such as earthquakes and tsunamis.

II. System Model

A. Network Model

We consider a three-dimensional static underwater sensor network comprising a set of underwater sensor nodes positioned at a depth of about 100 meters below the sea surface to monitor marine life and other underwater assets. A buoy positioned at the sea surface above the network serves as the sink for data generated within the network. The seabed nodes can serve as a data source, denoted $\mathbb{S}$. Each underwater node is equipped with a single half-duplex acoustic modem for communication, while the sink is a more powerful node fitted with a strong ocean-facing acoustic modem as well as RF links for communicating with an onshore data centre. We assume that it is possible for nodes in the network to transmit directly to the sink instead of using their cluster heads, as shown in the network model in Figure 1. Network instantiation, neighbour discovery and clustering are performed following the method in [1].

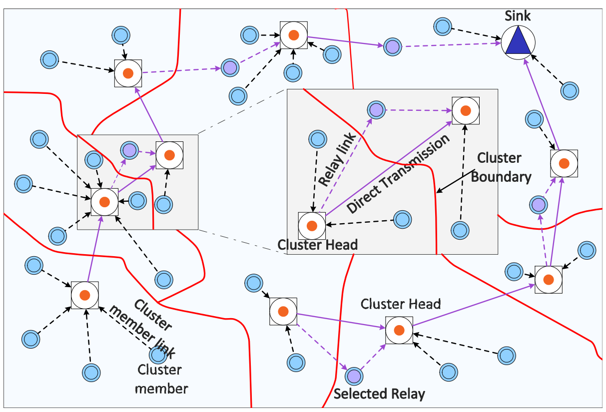

Fig. 1. Network model showing distribution of the nodes and clusters. Data is collected in each cluster by the cluster head, which performs data aggregation and sends only only clean copies of the data to the sink. When there is a spike in temperature and vibration, the data generated are tagged as priority packets and are not aggregated but sent immediately to the sink to warn of potential dangers such as tsunamis.

B. Channel Model

Acoustic communication is considered for forwarding data from sensor nodes to the cluster head and for transmitting aggregated data from the cluster head to the sink buoy at the sea surface. The final connection between the buoy and the data centre uses RF links. The received acoustic power from a source at a given range can be expressed by the passive sonar equation as a function of acoustic losses and modem characteristics. The signal-to-noise ratio (γ) depends on the frequency of the acoustic signal, the transmitting power and transmitter-receiver separation. For a narrowband acoustic signal (represented by a single tone of frequency Δf), this can be expressed as [9]

$$\gamma(d,f) = \frac{P_t/A(d,f)}{N_0(f)\Delta f}, \tag{1}$$

where P_t is the power of the acoustic projector, N_0 is the noise power, which is a function of frequency. A is the acoustic pathloss, which is given by

$$A(d,f) = A_0 d^k \alpha(f)^d, \tag{2}$$

where k is the channel spreading factor (equivalent to the pathloss exponent in terrestrial communications). The values of k ranges from 1–2; $k = 1$ is referred to as cylindrical spreading which dominates when the operating depth is less than the horizontal communication range, $k = 2$ is referred to as spherical spreading while $k = 1.5$ is referred to as practical spreading. The pathloss comprises a spreading loss which depends on the distance of the receiver (hydrophone) away from the acoustic projector (source/transmitter) and an absorption loss that grows with frequency. Absorption losses arise due to the chemical interactions between the wireless signal and water molecules, often leading to some of the propagating acoustic energy being converted to heat. Spreading losses arise from the decreasing intensity of the propagating waves away from the source. As the wavefront increases, the intensity of the signal energy per unit area decreases. The spreading loss can be expressed in dB re 1 μPa at 1m as [10]

$$L = k \times 10log(d), \tag{3}$$

whereas the absorption loss has been obtained empirically by W.H. Thorp (the so-called Thorp formula) for kHz frequencies [11] as

$$\alpha(f) = 0.11\frac{f^2}{1+f^2} + 44\frac{f^2}{4100+f^2} + 2.75\cdot 10^{-4} f^2 + 0.003. \tag{4}$$

Noise in underwater acoustic communication systems is a function of frequency of propagation and include contributions from shipping N_s, thermal noise N_{th}, noise due to water waves, N_w and water turbulence, N_t. The power spectral densities of the different noise sources can be expressed in dB re 1 μPa at 1m per Hz [9] as

$$N_t(f) = 17 - 30\log(f), \tag{5}$$

$$N_s(f) = 40 + 20(s-0.5) + 26\log(f) - 60\log(f+0.03), \tag{6}$$

$$N_w(f) = 50 + 7.5\sqrt{w} + 20\log(f) - 40\log(f+0.4), \tag{7}$$

$$N_{th}(f) = -15 + 20\log(f), \tag{8}$$

where w is the wind speed in m/s, s is the shipping activity factor (0 for low activity and 1 for high activity) and f the frequency. As equations (5 – 8) show, these noise sources are frequency-dependent. For instance, shipping activity noise and noise due to ocean turbulence dominate at very low and low frequencies. Thermal noise is predominant at very high frequencies while noise due to surface waves is strongest between 100 Hz – 100 kHz [10]. N_0 from equation (1) is the cumulative noise power from the different noise sources and can be expressed as

$$N_0 = 10^{\frac{N_t(f)+N_s(f)+N_w(f)+N_{th}(f)}{10}}. \tag{9}$$

III. Dynamic Clustering and Data Aggregation

In addition to extracting similarities in data, it is also important to be able to detect dissimilarities or anomalies/erroneous readings, which could otherwise corrupt the aggregated data. The clustering approach used in this work is based on that proposed in [1] which was shown to lead to significant improvement in network performance. We adopt the data aggregation aggregation approach in [12]. To quantify the results achieved, we analyse the network sum-throughput and coverage probability as a function of the energy consumption of the network.

A. Clustering

The entire network is grouped into clusters for ease of data processing and to save energy by reducing the communication distance between a node and its cluster head. As a baseline, we compare the performance of cluster-based communication with direct transmissions to the sink in Section IV. The clustering process is carried out in three stages, following the method proposed in [1]. The first stage involves clustering using the k-means clustering algorithm. In the second stage, the residual energy of nodes is combined with the k-means algorithm to select cluster heads. This is only relevant after the first round of operation of the network (during the first round, the k-means algorithm is used to select the centroid of each cluster and the nearest node to the centroid is elected as cluster head). The k-means algorithm ensures that the selected cluster head is geometrically close to the centre of the cluster while the residual energy condition ensures that the node has sufficient battery power to perform data aggregation and transmission operations. The final stage is the actual data collection and aggregation stage whereby nodes in a cluster forward their data to the cluster head using a TDMA-Based format.

The k-means algorithm groups an unlabelled multidimensional data set into a set of k distinct clusters, $C_i = \{C_1, C_2, \ldots, C_k\}$, where k is chosen beforehand. Given a sensor network with N sensor nodes $\{n_1, n_2, \ldots, n_i\}$, the aim of the k-means algorithm is to minimise the distance between nodes in the same cluster and maximise the distance between nodes in different clusters [13]. This is achieved using the following minimisation function

$$J_{\min} = \sum_{j=1}^{k} \sum_{c_{\mathrm{n}} \in C_{\mathrm{i}}} \left\| \mathrm{n_i} - \mu_j \right\|^2 , \tag{10}$$

where each cluster, C_i contains N_j nodes, n_i represents the i-th node in the network; μ_j represents the geometric centroid of the sensor nodes in a given cluster [14], which is determined using

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

There are different ways to select the optimal number of clusters based on the network size. One of the most accepted techniques is using the *Elbow method*, which returns the optimal number of clusters by calculating the intra-cluster sum of squares, η for the data set, which is given by

$$\eta = \sum_{i=1}^{n} (n_i - c_i)^2 . \tag{12}$$

If the cluster size is too large, nodes in each cluster spend more energy to reach the cluster, leading to faster disconnection of the network. However, if the cluster size is too small, it leads to high inter-cluster interference, which reduces the throughput and leads to energy wastage. A plot of the Elbow method for a generic sensor network is shown in Figure 2.

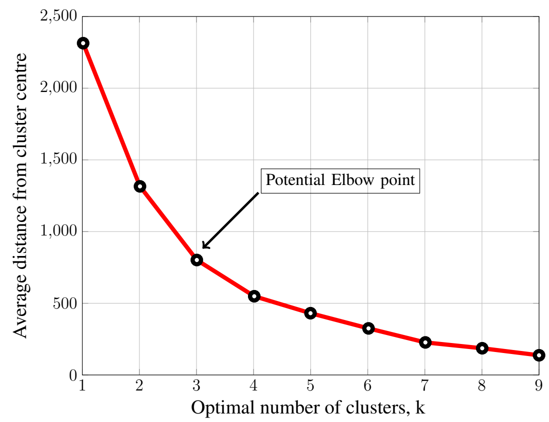

Fig. 2. Initial cluster size selection is is achieved via the Elbow method in Eq. (12). Following this, subsequent clustering is performed dynamically according to changes in the network. That is, as the network size changes due to removal of nodes from the network (due to battery depletion or node failure), re-clustering is done to ensure uniform energy distribution and continued connectivity for all nodes.

B. Cluster-based data aggregation

Data aggregation is performed to average collected data at the cluster head before forwarding it to the sink. This significantly improves the quality of the data collected and reduces energy usage. The technique employed for data aggregation is based on that proposed in [12]. Due to space limitation, we do not reproduce the analysis here but refer the reader to [12] and the references therein.

C. Throughput and Outage Probability

Here, we analyse the throughput and outage probability of the system, assuming that the nodes use the cluster heads to send data to the sink. Data transmission takes place in blocks of T seconds, which is divided into two phases of $T/2$ seconds each. In the first phase, the nodes transmit their data to the cluster head while the cluster heads forward the data to the sink in the second phase. The signal to noise ration (SNR), $\gamma(d,f)_i^{C_k}$ between the ith node, n_i and the kth cluster head, C_k can be expressed as

$$\gamma(d,f)_i^{C_k} = \frac{P_i/A(d,f)_{i,C_k}}{N_{0i,C_k}(f)\Delta f}, \tag{13}$$

where P_i is the transmitting power of the ith node, $A(d,f)_{i,C_k}$ is the transmission loss for the n_i–C_k link and N_{0i,C_k} is the noise power. Similarly, the SNR, $\gamma(d,f)_i^{C_k}$ between the kth cluster head and the sink S can be expressed as

$$\gamma(d,f)_i^{S} = \frac{P_{C_k}/A(d,f)_{C_k,S}}{N_{0C_k,S}(f)\Delta f}, \tag{14}$$

where P_{C_k} is the transmitting power of the kth cluster head, $A(d,f)_{C_k,S}$ is the transmission loss for the C_k–S link and $N_{0C_k,S}(f)$ is the noise power.

Outage occurs when the received SNR falls below a predefined threshold, γ_{th}. For ease of analysis, we assume the same γ_{th} for decoding data at the cluster heads and at the sink. Since the network is clustered, outage occurs either when the SNR received from a sensor node at the cluster head is less than γ_{th} OR the SNR at the cluster head is higher than γ_{th} but that at the sink is less than γ_{th}. That is, outage occurs under the following condition

$$P_{\text{out}} = \Pr(\gamma(d,f)_i^{C_k} < \gamma_{th}) + \Pr(\gamma(d,f)_i^{C_k} > \gamma_{th}, \gamma(d,f)_i^{S} < \gamma_{th}). \tag{15}$$

To evaluate the throughput per transmission block of the network, we consider the transmission rate R (which is limited by the available bandwidth). For ease of analysis, we assume the same R for the node–cluster head links and the cluster head–sink links. The block throughput of the network is considered when the network is not in outage (i.e., $1 - P_{\text{out}}$) and can be expressed as

$$\tau = \frac{(1-P_{\text{out}})R(T/2)}{T}, \tag{16}$$

$$= \frac{(1-P_{\text{out}})R}{2} \tag{17}$$

TABLE II
TABLE OF PARAMETERS I

Parameter	Value
Transmit power (P_t)	170 dB re 1 μPa @ 1m
Frequency, f	24 kHz
Bandwidth	4 kHz (22–26 kHz)
End-to-end reliability (α)	0.95
Link failure rate (p_i) (random)	0.05 – 0.25
Network area, number of nodes	1km^2, 100
Angle of curvature of multipath components, $\theta[0,1,2]$	0, 15, -15
Delay of the main path τ_0	c/pathlength
$\tau[n](n=0,1,2)$	$\frac{3.345\times d(m)}{500} + 0.3345 \times n$
Wind speed, (w)	10 m/s
Shipping activity factor, (s)	0.2
Spreading factor, (k)	1.5

where T is the total block transmission time. γ_{th} can be obtained from the information rate using the following relation [15]

$$R = B\log_2(1+\gamma_{th}), \tag{18}$$

where B is the transmission bandwidth, so that $\gamma_{th} = 2^R - 1$.

IV. Performance Evaluation

This section presents computer simulations used to evaluate the dynamic clustering and data aggregation (DYCADA) scheme proposed for cluster head selection and data aggregation in an IoUTs network. Direct transmissions to the sink from sensor nodes in the network is considered as benchmark model and compared to the DYCADA approach. Due to the shallow depth of deployment of the network, nodes can send their data directly to the sink at the sea surface, hence consideration of direct transmissions as a baseline is justified.

The sensors are deployed in a 2D plane at an operating depth of 100m, from where they can transmit data directly to the sink or to cluster heads that aggregate and forward the data to the sink. The network covers an area of 1km^2, with a node distribution intensity of 100 nodes per area ($\Phi_n = 100$), and the height of the sink is 100m. Other parameters considered in the analysis are summarised in Table II.

We begin by evaluating the achieved average throughput per node as a function of the source power of the nodes. In the first instance, each node transmits collected data with power P_t to the sink via direct transmissions. The simulation considers a range source powers normalised to a range between 0.1 to 1 (corresponding to a full source power of 170.8dB re μPa). Under this setting, the achieved throughput per node ranges between 1 to 4 kbps. For the same source powers but using the DYCADA protocol, the average throughput per node varies between 6 and 9.2 kbps, an improvement of over 100% compared to direct transmissions. The results are shown in Fig 3. The improvement in throughput is clearly due to the proximity of the nodes to the cluster head in the DYCADA scheme, which improves the SNR received at the cluster head as channel effects are minimised due to the shorter distance traversed by the packets compared to direct transmissions. However, data aggregation introduces additional delays on packets due to data processing at intermediate nodes (cluster heads), thereby adding to the overall latency in the received data and lowering their value of information (VoI).

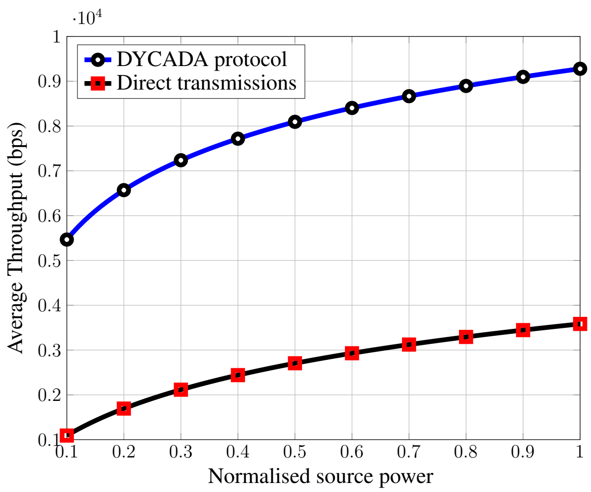

Fig. 3. Throughput per node for direct transmissions and transmissions using cluster heads. By splitting the transmission distance between the node and sink into two, data packets suffer less absoprtive losses when the cluster heads are used to aggregate and relay data. As a result, they arrive with high SNR at the cluster heads which fully decodes the data before forwarding it (with other aggregated packets) to the sink, ensuring that they arrive with high fidelity and significantly high SNR at the sink as well, thereby leading to higher throughputs when cluster heads are used.

Next, the energy efficiency (EE) is evaluated for both direct transmissions and transmissions using the DYCADA protocol as a function of source power. Naturally, the EE curve shows a decreasing trend as the source power increases for both methods, as can be seen in Fig 4. However, it can be seen that for transmissions via the DYCADA protocol, the EE is comparatively high for all ranges of the source power. Similar to the case for throughput, the results indicate that it is more energy efficient to break the transmission distances into shorter distances where lower source powers can be used to reach the receivers than to transmit over long distances using high powers. This is amongst the most important benefits of clustering and other routing schemes such as multi-hop relaying. Since it is more beneficial to partition the network into clusters so as to use lower transmission powers, how can the ideal number of clusters be determined, since it is tempting to keep breaking the clusters into ever smaller clusters? The Elbow method described in Section III-A provides the answer to this, as the algorithm yields the optimal number of clusters that result in the optimal saving of network resources without introducing new problems, as shown in Fig. 2. It can be seen from Fig. 4 that as the value of the source power increases, the two curves begin to converge, indicating that there is an optimal power range in which the transmission of data through the cluster head has an advantage over direct transmission. A point is eventually reached when direct transmissions will result in higher energy efficiency than the DYCADA protocol if the source power keep increasing, which is intuitive because using high powers for short range communications (as is the case within clusters) is wasteful.

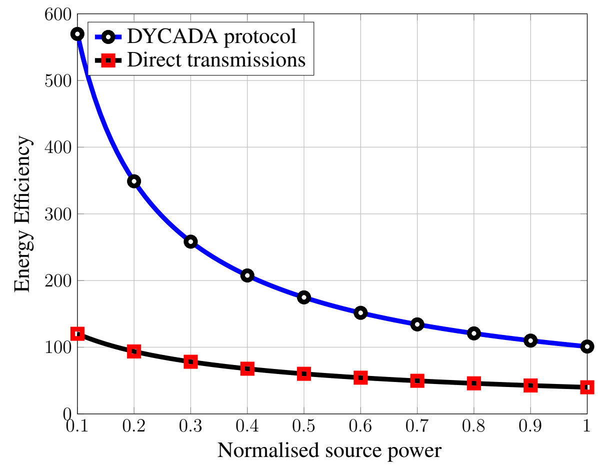

Fig. 4. Energy efficiency for direct transmissions and cluster-based transmissions. Clustering clearly outperforms direct transmissions, especially at low source powers. If there were unlimited power in the network (for e.g. if the sensor nodes are connected to a mains supply), a point is reached where direct transmissions is preferred to clustering in terms of energy efficiency, as shown by the converging trend of the two curves in this figure. One of the major gains of clustering is that lower transmission powers can be used without the nodes falling into outage.

Figs. 5 and 6 illustrate the coverage probability while varying the SNR thresholds (γ_{th}) for different source power levels. Fig. 5 shows the coverage probability for a source power of 163 dB re μPa. It demonstrates that as the SNR threshold grows, the coverage probabilities for both direct transmissions and transmissions via the DYCADA protocol decrease. This is expected because as the minimum SNR threshold required to successfully decode data increases, fewer nodes can achieve γ_{th} unless the source power increases commensurately. In addition, it can be observed from Fig. 5 that the coverage probability of direct transmission decreases significantly compared to the DYCADA protocol. Again, this is expected since the amount of power required to reach the sink is higher for direct transmissions compared to transmissions through the cluster heads which requires much lower transmission powers due to their proximity. Similar to Fig. 5, Fig. 6 shows the coverage probability for a source power of 168 dB re μPa. It also follows a decreasing trend for coverage probability as the SNR threshold increases for both direct transmissions and transmissions via cluster heads using the DYCADA protocol, as fewer nodes are able to attain γ_{th} at the fixed transmission power of 168 dB re μPa.

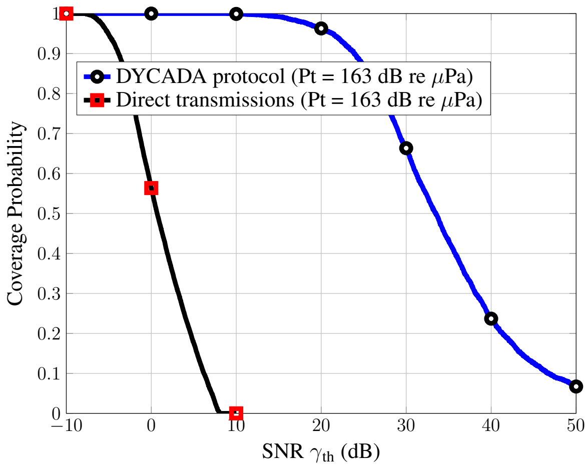

Fig. 5. Coverage probability of direct transmissions and transmissions via cluster head for a source power of 163 dB re μPa. For this fixed source power, more nodes are forced into outage as the SNR threshold is raised, as fewer nodes can achieve the higher demands for SNR demanded at the receiver. This is more severe for direct transmissions, as shown by the steepness of its curve here. Due to the proximity of the cluster heads to the nodes, they show a more gentle degradation in coverage probability, which again shows its super performance over direct transmissions.

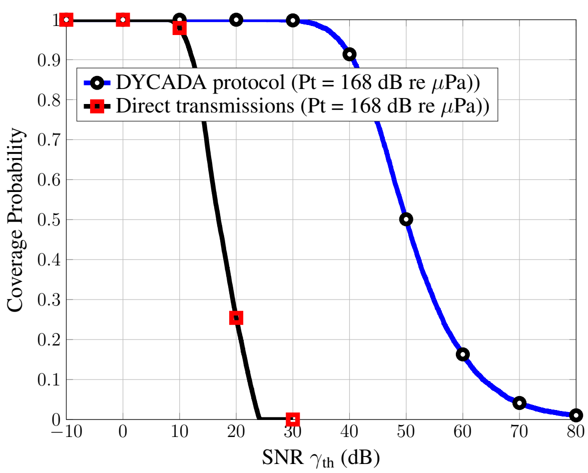

Fig. 6. Coverage probability of direct transmissions and transmissions via cluster head for a source power of 168 dB re μPa. At higher source powers, more sensor nodes are able to achieve the SNR threshold, which reflects in better coverage probability compared to Fig. 5.

V. Conclusion

In this paper, we have studied the performance of a dynamic clustering and data aggregation scheme (DYCADA) in a wireless underwater sensor network, analysed the throughput and derived the outage probability. Through simulations, we evaluated the impact of different system parameters on overall network performance. We showed that clustering and data transmissions via the DYCADA protocol significantly improves the network sum-throughput by over 100% compared to direct transmissions, and enhances coverage probability and energy efficiency, even when the sink is within the transmission range of all the network nodes. Moreover, it was demonstrated that the performance of the proposed scheme is greatly affected by the source transmission power. Our future work will consider the impact of clustering and relaying on the packet delays and how energy efficiency and throughput change when energy harvesting is implemented in the network.

References

[1] K. G. Omeke, M. S. Mollel, M. Ozturk, S. Ansari, L. Zhang, Q. H. Abbasi, and M. A. Imran, "Dekcs: A dynamic clustering protocol to prolong underwater sensor networks," *IEEE Sensors Journal*, vol. 21, no. 7, pp. 9457–9464, 2021.

[2] V.-V. Vo, T.-D. Nguyen, D.-T. Le, M. Kim, and H. Choo, "Link-delay-aware reinforcement scheduling for data aggregation in massive iot," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5353–5367, 2022.

[3] L. K. Ketshabetswe, A. M. Zungeru, B. Mtengi, C. K. Lebekwe, and S. Prabaharan, "Data compression algorithms for wireless sensor networks: A review and comparison," *IEEE Access*, 2021.

[4] X. Xu, S. Wang, X. Mao, S. Tang, P. Xu, and X.-Y. Li, "Efficient data aggregation in multi-hop wsns," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, 2009, pp. 1–6.

[5] X. Liu, J. Yu, F. Li, W. Lv, Y. Wang, and X. Cheng, "Data aggregation in wireless sensor networks: From the perspective of security," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6495–6513, 2020.

[6] M. S. Yousefpoor, E. Yousefpoor, H. Barati, A. Barati, A. Movaghar, and M. Hosseinzadeh, "Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: A comprehensive review," *Journal of Network and Computer Applications*, vol. 190, p. 103118, 2021.

[7] V. Krishnaswamy and S. S. Manvi, "Palm tree structure based data aggregation and routing in underwater wireless acoustic sensor networks: Agent oriented approach," *Journal of King Saud University-Computer and Information Sciences*, 2019.

[8] N. Goyal, M. Dave, and A. K. Verma, "Data aggregation in underwater wireless sensor network: Recent approaches and issues," *Journal of King Saud University-Computer and Information Sciences*, vol. 31, no. 3, pp. 275–286, 2019.

[9] M. Stojanovic, "On the relationship between capacity and distance in an underwater acoustic communication channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007.

[10] N. Morozs, W. Gorma, B. T. Henson, L. Shen, P. D. Mitchell, and Y. V. Zakharov, "Channel modeling for underwater acoustic network simulation," *IEEE Access*, vol. 8, pp. 136 151–136 175, 2020.

[11] W. H. Thorp, "Analytic description of the low-frequency attenuation coefficient," *The Journal of the Acoustical Society of America*, vol. 42, no. 1, pp. 270–270, 1967.

[12] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "Pda: Privacy-preserving data aggregation in wireless sensor networks," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, 2007, pp. 2045–2053.

[13] 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 Transactions on Cybernetics*, vol. 47, no. 3, pp. 772–783, 2017.

[14] 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 *2013 27th International Conference on Advanced Information Networking and Applications Workshops*, 2013, pp. 910–915.

[15] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Wireless-powered relays in cooperative communications: Time-switching relaying protocols and throughput analysis," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1607–1622, 2015.