Fitness function X-means for prolonging wireless sensor networks lifetime

Abdelrahman Radwan, Nasser Abdellatif, Eyad Radwan, Maryam Akhozahieh
Department Electrical Engineering, Faculty of Engineering and Technology, Applied Science Private University, Amman, Jordan

ABSTRACT

X-means and k-means are clustering algorithms proposed as a solution for prolonging wireless sensor networks (WSN) lifetime. In general, X-means overcomes k-means limitations such as predetermined number of clusters. The main concept of X-means is to create a network with basic clusters called parents and then generate \((j)\) number of children clusters by parents splitting. X-means did not provide any criteria for splitting parent’s clusters, nor does it provide a method to determine the acceptable number of children. This article proposes fitness function X-means (FFX-means) as an enhancement of X-means; FFX-means has a new method that determines if the parent clusters are worth splitting or not based on predefined network criteria, and later on it determines the number of children. Furthermore, FFX-means proposes a new cluster-heads selection method, where the cluster-head is selected based on the remaining energy of the node and the intra-cluster distance. The simulation results show that FFX-means extend network lifetime by 11.5\% over X-means and 75.34\% over k-means. Furthermore, the results show that FFX-means balance the node’s energy consumption, and nearly all nodes depleted their energy within an acceptable range of simulation rounds.

Keywords: Clustering, Energy efficient, k-means, Routing, Sensor networks, Wireless, X-means

1. INTRODUCTION

Low power communication networks such as wireless sensor networks provide a low-cost infrastructure to implement machine to machine communication (M2M) and the internet of things (IoT) [1]. Wireless sensor networks (WSNs) are designed as application-specific and perform dedicated tasks. The network consists of wireless nodes equipped with sensors and actuators. The wireless node’s main source of energy are batteries and transmitting data is the main energy consumer [2]; therefore, there is a need for an inexpensive routing algorithm to extend the network lifetime [3]–[7]. WSN literature such as [8]–[10], identify clustering as efficient routing algorithms that have shown promising enhancements in the WSN lifetime. Classical clustering algorithms such as low-energy adaptive clustering hierarchy (LEACH) [11], hybrid energy-efficient distributed (HEED) [12], and energy efficient clustering scheme (EECS) [13] divide their networks into random clusters, and each cluster comprises a cluster-head (CH) and wireless nodes. Wireless nodes collect data from the surrounding and relay them to CHs; then, CHs have to aggregate and forward the data to a central unit (the sink). CHs use long-range communications to the sink, called backhauling, and nodes within a cluster take turns into becoming the CH; thus, re-clustering is initiated every round of operation. Rotating the CH role among nodes distributes the backhaul communication energy
consumption among nodes. The concept of random clustering in classical algorithms suffers from an unbalanced intra-cluster distance, unbalanced cluster size, and unbalanced energy consumptions; thus, the concept of fixed clustering has been introduced in k-means algorithms [14]–[19]. k-means algorithm has many applications such as developing vehicle driving cycle [20], fingerprint recognition system [21], and data mining [22]. The k-means algorithms estimate the number of clusters, their initial centroids and then assign nodes to clusters based on the minimum distance to centroids. After that, the average Euclidean distance to these centroids was determined and centroids repositioned; later, wireless nodes were re-clustered and assigned to the new centroids. Determining the mean distance and re-clustering the nodes are repeated until centroid positions are fixed, and cluster members are static.

Unlike classical algorithms, k-means algorithms balance clusters’ intra-cluster distance, but these algorithms still construct bad clusters. A bad choice of initial centroid leads to unbalanced cluster sizes and unbalanced CHs consumption that causes premature CHs death at the early stage of the network lifetime. Therefore, Radwan et al. [23]–[25] proposed the X-means algorithm in three articles. Their concept is to split k-means centroid (parents) into multiple new positions (children), thus expanding the search space for better positioning and the best number of clusters. Their work provided a leap in the network lifetime and reduced k-means limitation, but the problem of determining the number of clusters persists at the children’s level. Furthermore, X-means did not provide splitting criteria to determine which cluster should split or not but assumed that all clusters were worth splitting to a number of children determined by the user. This article proposes fitness function X-means (FFX-means) as a clustering algorithm based on X-means. Simply, each parent centroid will have a fitness based on its cluster size and the Euclidean distance to the sink. The number of children determined if the cluster is worth splitting; otherwise, the parent cluster remains; the position of the children centroids selected randomly within the average intra-cluster distance of the parent centroid.

The rest of the article is as follows; in section 2, FFX-means are proposed and described with mathematical equations. In section 3, the simulation results, and a discussion on how the proposed algorithm prolongs WSN lifetime. Finally, in section 4, we conclude our work with future recommendations to further enhance the FFX-means algorithm.

2. PROPOSED METHOD

The proposed method in this article includes energy model and introduces new methods for clusters formation, cluster splitting criteria, and cluster-heads selection. The energy model is obtained from [2]. The Energy-model consists of transmitter and receiver parts, as in (1) and (2). The transmitter accounts for the energy required to aggregate and process data ($E_{elec}$), data size ($D$), the number of nodes ($n$), and the amplification energy required to transmit ($D$) data over ($d$) distance. The amplification energy in (1) is described by two types of signal attenuation, where $e_{fs}$ describes the free-space attenuation model, and $e_{mp}$ describes the multipath fading attenuation model. ($d_{o}$) is the crossover distance between the free-space and multipath models, described in [15]. The total transmission energy ($E_T$) is proportional to the distance between a source and destination nodes. It corresponds to $d^2$ attenuation when $d$ is less than $d_{o}$ and corresponds to $d^4$ attenuation when $d$ is greater than or equal to $d_{o}$. The receiving consumption ($E_R$) includes the energy required to process data packet ($D$) that is received from ($n$) nodes in bits as in (2).

\[
E_T = \begin{cases} 
D \times n \times E_{elec} + D \times e_{fs} \times d^2, & d < d_{o} \\
D \times n \times E_{elec} + D \times e_{mp} \times d^4, & d \geq d_{o}
\end{cases}
\]  

(1)

\[
E_R = D \times n \times E_{elec}
\]  

(2)

2.1. Clusters formation and splitting criteria

The concept of the X-means clustering is summarized in previous work [25]. The first phase is constructing random clusters with random centroids, then, using k-means [15], optimizing these positions as in (3). The process is recursively repeated until a final copy of centroids is determined, and the final copy of centroids is called parent centroids. The second process is searching for updates of the parent centroids; here, a new method proposed. The method is driven from metaheuristic algorithms; a parent centroid’s fitness is determined as in (4). A parent is fit to split if its fitness ($F$) is greater than average fitness ($AvgF$). $AvgF$ found by the crossover distance ($d_{o}$) divided by the average distance of the parents’ centroids to the sink, added to them the expected cluster-size divided by the average cluster-size of the parents as in (5). The expected cluster-size found by (OCS * N), OCS is user defined value (1%, 10%, 20% ...).

\[
P_n = \alpha_{n-1}(x, y) + \frac{1}{S_n} \sum_{i=1}^{S_n} \beta_i(x, y) \begin{cases} 
n \in 1, 2, 3, ..., k \\
i \in 1, 2, 3, ..., S_n
\end{cases}
\]  

(3)
Fitness function X-means for prolonging wireless sensor networks lifetime

\[ F_n = \frac{d_n}{d_{n,Sink}} + \frac{C_n + k}{\sum C_n}; \quad n \in 1,2,3 \ldots k \]  
\[ AvgF = \frac{k \ast d_n}{\sum d_{n,Sink}} + \frac{OCS \ast N + k}{\sum C_n}; \quad n \in 1,2,3 \ldots k \]  

\( P_n \): parent (n) centroid,
\( a_{n-1}(x,y) \): x and y position of previous (n-1) parent centroid,
\( S_n \): set of nodes assigned in cluster to \( P_n \),
\( \beta_i(x,y) \): the x and y position of node (i),
\( F_n \): the fitness of the parent (n) centroid to split,
\( K \): predefined initial number of centroids,
\( C_n \): cluster-size of parent (n),
\( d_{n,Sink} \): the Euclidean distance between parent (n) centroid and the sink,
\( OCS \): optimal cluster size (percentage of total nodes of the network),
\( N \): total number of nodes in the network.

After determining if parents are fit to split, the number of children per parent has to be determined as in (6). Then every child location is determined as in (7), where a random value is selected from the range 1 to \( d_m \) and multiplied by \( \cos \theta \) to get x position and \( \sin \theta \) to get y position. Finally, \( (d_m) \) is the mean Euclidean distance within the parent cluster, which is determined as in (8).

\[ P_{n,j} = 2 \frac{AvgF}{F} \]  
\[ C_j = \begin{cases} 
    X_j = \text{rand} [1, d_{m,n}] \ast \cos \theta \\
    Y_j = \text{rand} [1, d_{m,n}] \ast \sin \theta 
\end{cases}; \quad j \in 1,2,3 \ldots P_{n,j} \]  
\[ d_{m,n} = \frac{1}{\sum_{i=1}^{S_n}} \sqrt{(x_{P_n} - x_i)^2 + (y_{P_n} - y_i)^2} \]  

\( P_{n,j} \): number of children (j) generated from the parent (n).
\( C_j \): the location of the child (j) of parent (n).
\( d_{m,n} \): the average intra-cluster distance between parent centroid and member nodes in the cluster.

After that, new children \( C_j \) are set as cluster centroids, then nodes are re-clustered and assigned using a minimization technique, where nodes are grouped by the minimum average sum of their distance from the sink and children (\( C_j \)) as in (9) and (10). Finally, the X-means algorithm recursively executes (11) until children’s centroid positions converge. The recursive run has three outcomes; children form their clusters and parents collapse, second some children and parents diminish, and third, children collapse, and parents repositioned to best locations. Figure 1 shows an example for parent centroids, Figure 2 shows parent splitting, and Figure 3 shows the remaining centroids and their final position.

![Figure 1. Parents initial position selected randomly and adjusted to their final position using the k-means algorithm](Image)
Figure 2. According to their fitness, parents split to a number of children marked with (x); each mark has the parent’s name above it and the child’s name below it.

Figure 3. After applying X-means, parents have diminished, and children remained, each child adjusted to the final location and forms its own cluster.

\[
AD_{i,j} = ((d_{i,j} + d_{i,sink})/2), i \in 1,2,3 \ldots S_n \quad (9)
\]

\[
AD_i = \min\{AD_{i,1}, AD_{i,2}, AD_{i,3}, \ldots AD_{i,j}\}, i \in S_n, j \in P_n \quad (10)
\]

\[
CC_{n,j} = \emptyset(x,y) + \frac{1}{S_n} \sum_{i=1}^{S_n} \beta_i(x,y) \left\{ \begin{array}{ll} n \in 1,2,3 \ldots k \\
\end{array} \right. \quad (11)
\]

\[
AD_{i,j} \text{: average distance of node (i) from the child (j) and from the sink.}
\]

\[
d_{i,j} \text{: node (i) Euclidean distance to child (j).}
\]

\[
d_{i,sink} \text{: node (i) Euclidean distance to the sink.}
\]

\[
\emptyset(x,y) \text{: position of the sink.}
\]

\[
CC_{n,j} \text{: final position of child (j) of parent (n).}
\]

2.2. Cluster-heads selection

In previous work [25], cluster-heads were selected and rotated based on their remaining energy; if cluster-head energy dropped below a certain threshold, the cluster-head step-down and a new cluster-head with energy that exceeds the threshold is selected. The threshold itself is not a fixed value, and a node updates its threshold by a small step when its energy drops below it; thus, the node will have a chance of being selected at a later stage. However, this technique benefits the cluster-head only, as it ignores the importance of minimizing the intra-cluster distance to extend the network lifetime. Here, this article proposes a new technique to select cluster-heads as in (12) to (14). At the beginning of the simulation, all nodes with the same initial energy and nodes with smaller distances to their cluster centroid are selected to be cluster heads. Then, when these nodes’ energy decay, new cluster-heads selected have the minimum distance to centroids and the highest remaining energy among their peers.

\[
\omega = \frac{E_{\text{initial}}}{E_{\text{remaining}}} \quad (12)
\]

\[
F_{C,i} = e^{\omega d_{i,C}}, i \in 1,2,3 \ldots S_c \quad (13)
\]

\[
CH_{Sc} = \min\{F_{C,1}, F_{C,2}, F_{C,3}, \ldots F_{C,S_c}\} \quad (14)
\]

\[
\Omega \quad : \text{Represent the ratio of initial energy of a node to its remaining energy,}
\]

\[
d_{i,C} \quad : \text{The distance of node (i) to the child/cluster centroid,}
\]

\[
F_{C,I} \quad : \text{Fitness of node (i) to become cluster-head, nodes with small distance to centroid and high remaining energy will have the smallest fitness,}
\]

\[
CH_{Sc} \quad : \text{The selected cluster-head from set of nodes (Sc) that forms a cluster.}
\]
RESULTS AND DISCUSSION

An extensive simulation has been carried out in MATLAB R2020b to test FFX-means. Table 1 shows the simulation parameters and the result benchmarked among traditional k-means and X-means as in Figure 4. The initial number of clusters is five, and their centroids position is selected randomly; the simulation area is 220×220 unit², and the density of the nodes is uniform.

Figure 4 shows that FFX-means has prolonged the network lifetime by 11.5% over X-means and 75.34% over k-means. However, a sharp decline in the number of alive nodes of FFX-means appeared after round 2750 of the simulation; this decline indicates a balanced consumption among nodes and a balanced cluster size in the network; thus, nodes depleted their energy at the same round of simulation. Heinzelman et al. [11], [26] estimated that the optimal cluster size (OCS) will vary between 9-11% of total network nodes; thus, FFX-means simulated for OCS values and Figure 5 shows that the first node death (FND) appeared between round 2,600 and 2,750 for ten different runs. To further test the stability of FFX-means, the simulation repeated for fifty runs with OCS set to 10% and results in Figure 6 capture FND and last node death (LND); similar to Figure 5, there is no abnormal behavior in FND, and it is noticeable that whenever FND increases the LND decreases and vice versa.

Table 1. X-means simulation parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_0 )</td>
<td>Crossover distance</td>
<td>89 m</td>
</tr>
<tr>
<td>( \epsilon_{fs} )</td>
<td>Free space model amplification energy</td>
<td>6.3 pJ/bit/m²</td>
</tr>
<tr>
<td>( \epsilon_{mp} )</td>
<td>Multipath amplification energy</td>
<td>0.0008 pJ/bit/m⁴</td>
</tr>
<tr>
<td>( E_{elec} )</td>
<td>Single bit processing energy</td>
<td>( Tx ) or ( Rx = 50 \text{ nJ/bit} ) (Aggregation = 5 nJ/bit)</td>
</tr>
<tr>
<td>( E_{init} )</td>
<td>Initial energy</td>
<td>0.5 J</td>
</tr>
<tr>
<td>OCS</td>
<td>Optimal cluster size</td>
<td>10%</td>
</tr>
<tr>
<td>Total number of nodes</td>
<td>-------</td>
<td>220</td>
</tr>
<tr>
<td>Packet size</td>
<td>Data size</td>
<td>4.2 Kb/packet</td>
</tr>
<tr>
<td>Control packet</td>
<td>Packet headers</td>
<td>0.2 Kb/packet</td>
</tr>
<tr>
<td>Sink position</td>
<td>-------</td>
<td>Center</td>
</tr>
<tr>
<td>Simulation area</td>
<td>-------</td>
<td>220 × 220</td>
</tr>
</tbody>
</table>

Figure 4. FFX-means simulation results benchmarked against X-means and k-means
4. CONCLUSION

Splitting and merging clusters in wireless sensor networks aims to extend the network lifetime. X-means algorithm is an example of splitting the clusters into multiple children searching for new centroids. Still, the algorithm neither provides any criteria or measurements to determine if the clusters were worth splitting nor the number of new clusters. In this article, X-means updated with fitness function (FFX-means) to resolve these issues, the fitness function determines if the cluster is worth splitting based on its centroid distance from the sink and how its cluster-size compares to average network cluster-size. Furthermore, FFX-means equipped with a new cluster-heads selection algorithm that balances the remaining energy of the node and its intra-cluster distance. For future works, new splitting criteria such as the intra-cluster distance and average network energy should be explored, and further simulation with a large-scale network should be conducted.

REFERENCES


**BIographies of AUTHORS**

**Abdelrahman Radwan** obtained a BEE (2007), M.Sc. (2009) from Northumbria University, UK, and Ph.D. (2020) from UCSI. From 2010 to 2016, he worked as a senior wireless development engineer, and from 2016-2017 he was a WIMAX consultant for FRC corporation, USA. Currently he is an assistant professor at Applied Science University (ASU), Amman, Jordan. His research interests are trapped ion quantum computer, wireless systems, IoT, embedded systems, and artificial intelligence. He can be contacted at email: a_radwan@asu.edu.jo.

**Nasser Abdellatif** is a professor and dean of the Faculty of Engineering and Technology. He earned a BEEE and MEEE in 1985 and 1986, respectively, from the City College of the City University of New York and a Ph.D. in electrical engineering in 1991 from the City University of New York. He has been employed as a professor at Applied Science University in Amman, Jordan, since 2005. During this time, he served as Vice President, Vice President for Academic Affairs, Dean of Students, and currently, as Dean of the Faculty of Engineering and Technology. He was employed by CSC Mobile Comm. and CSC Telecomm from 1989 to 1991. He served at Bronx Community College from 1991 to 2005. Prof. Abdellatif served as the Department of Physics and Technology chairperson from 1993 to 2005 and the Science, Mathematics, and Technology Division coordinator from 1997 to 2005. Over his career, Prof. Abdellatif has published extensively and has made numerous presentations at local, regional, and national conferences. He was awarded several grants totaling about 2.5 million dollars. Over the last ten years, one of his primary interests has been teaching methodology and student-centered learning. He has conducted many faculty training workshops on integrating skills into the curriculum. He has been an advocate for faculty training and has trained over 300 faculty members worldwide. His research interests include error-correcting coding, communication systems, and optical communications. He can be contacted at email: nasser_abdellatif@asu.edu.jo.
Eyad Radwan completed his B.Sc. from JUST in 1996 (Jordan), M.Sc. and Ph.D. from UPM (Malaysia) in 1999 and 2004, respectively. From 1999 till 2011 he was with UCSI University, Malaysia. During his term with UCSI University, Prof. Radwan served as the Dean of the Faculty of Engineering and Built Environment from 2007 to 2009, and as Chief Operating Officer from 2009 to 2011. Prof. Radwan also served as member of the Malaysia Engineering Accreditation Council from 2009-2011. He joined the Department of Electrical Engineering at the Applied Science Private University (ASU), Amman, Jordan in 2012. In addition to his role as a faculty member, Prof. Radwan has been appointed as the president’s assistant for International Relations at ASU. His research interests are in the areas of renewable energy, electric drives, and control of electric converters. He can be contacted at email: e_redwan@asu.edu.jo.

Maryam Akhozahieh was born in Kuwait. She received her B.S.E. and M.S.E. degrees in electrical engineering from Kuwait University in 1984 and 1989, respectively, and her Ph.D. degree from The University of Akron, Akron, OH, in 2006. In 1984 to 1989, she was a teaching assistant at Kuwait University. From 1992 to present, she is a faculty member with the Department of Electrical Engineering, Applied Science University, Amman, Jordan. Currently she is a professor and the head of the department of electrical engineering department. Her areas of interest include, but not limited to, signal processing, multicarrier modulation, and multiple-access communication techniques. Prof. Akho-Zahieh was the recipient of the Graduate Teaching Assistant of the Year Award from the Department of Electrical and Computer Engineering, The University of Akron, in 2005. Prof. Akho-Zahieh is a member of several IEEE societies. She can be contacted at maryamm@asu.edu.jo.