

CHAPTER 5

ENHANCEMENT OF CLUSTERING ALGORITHM

5.1 Introduction

The Wireless Sensor Network (WSN) is a collection of implanted systems that are able to transmit or process information through various sensor nodes to their neighbouring sensor nodes. The WSN is composed of sensor nodes that are densely deployed in high density manner and in large quantity. A sensor node refers to an object that performs some sensing tasks, apart from processing signals. Sensing is a method that gathers information about a physical object or a process. Sensor nodes do not only have sensing skills, but also various functionalities, communication, and storage capabilities. Due to these factors, the nodes can be connected to the central processing CH for communication and assembly of data.

The objective of this chapter is to design a clustering algorithm based on the behaviour of the correlated nodes. The algorithm focuses only on energy consumption to enhance network connectivity. Correlation clustering is a method for clustering a set of objects into the optimum number of clusters without specifying that number in advance. In order to retain the connectivity of the sensor network, the parameters depicted in Chapter 4 were applied to propose the Enhancement Clustering Algorithm (ECA) to improve network connectivity under correlated node behaviour.

5.1.1 Original Clustering Algorithm

The previously clustering algorithm which has used frequently thus far resolves energy consumption between nodes and increases the lifetime. However, because the clustering algorithm method selects cluster heads probabilistically and re-establishes clusters regularly, unnecessary energy is consumed. Therefore, this research is based on previously Clustering Algorithm for Efficient Energy Management (Yang et al., 2017). In order to design the enhancement of the clustering algorithm, that set an energy threshold value to resolve the network connectivity of unnecessary energy consumption in the previously clustering algorithm. This research enhances existing original clustering algorithm for WSN in Figure 5.1 (page 81).

5.2 Enhancement Clustering Algorithm (ECA)

The main nodes of the clustering algorithms are meant to find and implement an appropriate distance measure between the nodes. In particular, this study proposes Enhancement Clustering Algorithm (ECA) based on 3D Euclidean distance, where the nodes must deal with the distance and later, focus on the challenge of clustering. Figure 5.2 (Page 82) shows a three-stage flowchart to formulate the Enhancement Clustering Algorithm.

The clustering algorithm is composed of three stages; initial setup, enhancement clustering, and maintenance clustering. In the first stage of the initial setup, all variables are declared to start the clustering process. In the second stage, the clustering selects the neighbouring nodes to form a cluster. The formation is based on 3D Euclidean distance and correlated degree, as elaborated in Chapter 4. The final stage is maintenance of the clustering algorithm to assess the nodes to determine the correlated degree value. If the correlated degree value falls below the energy threshold, the node is discarded and new clustering is formed. This process continues until all nodes are out of energy.

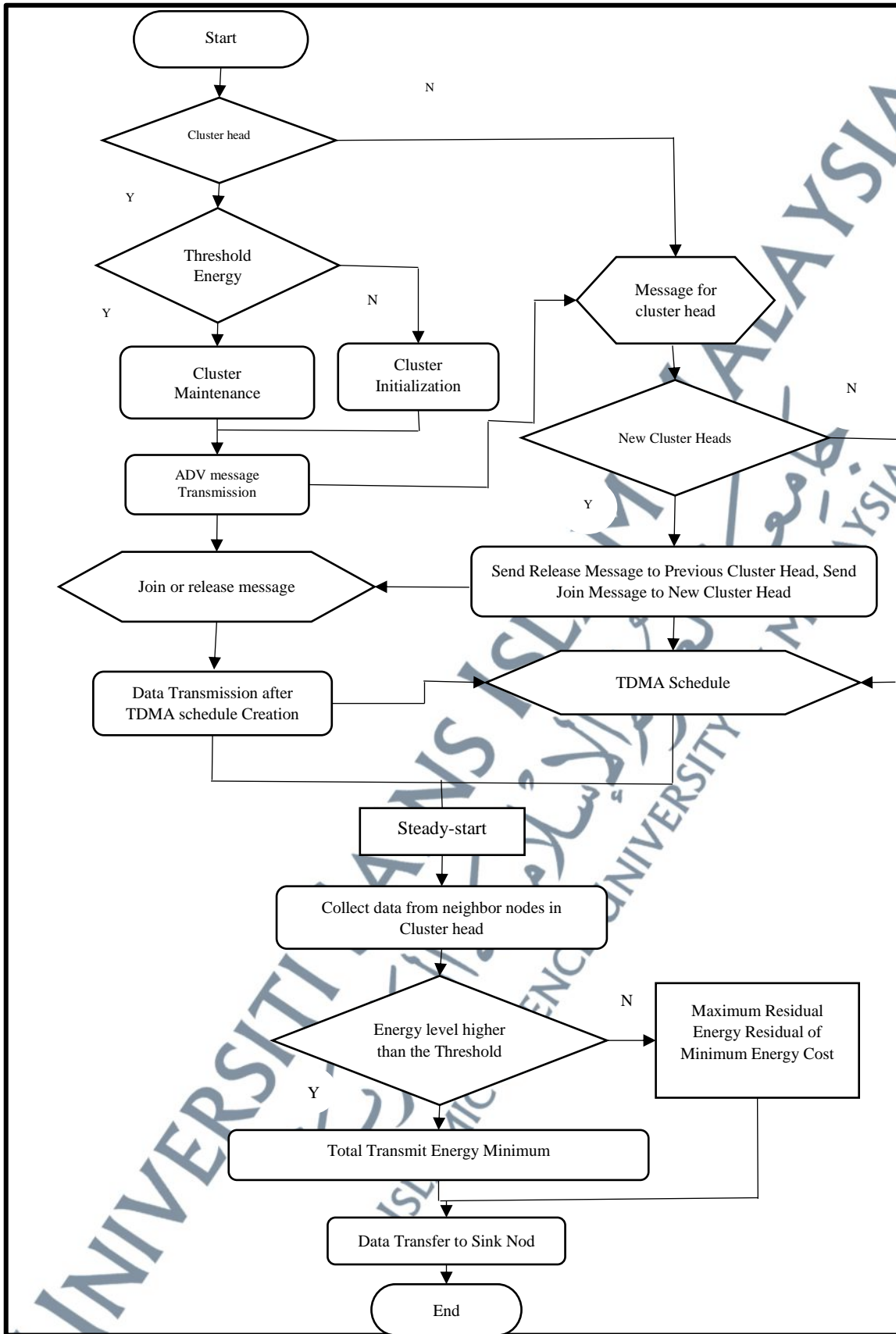


Figure 5.1: Original Clustering Algorithm

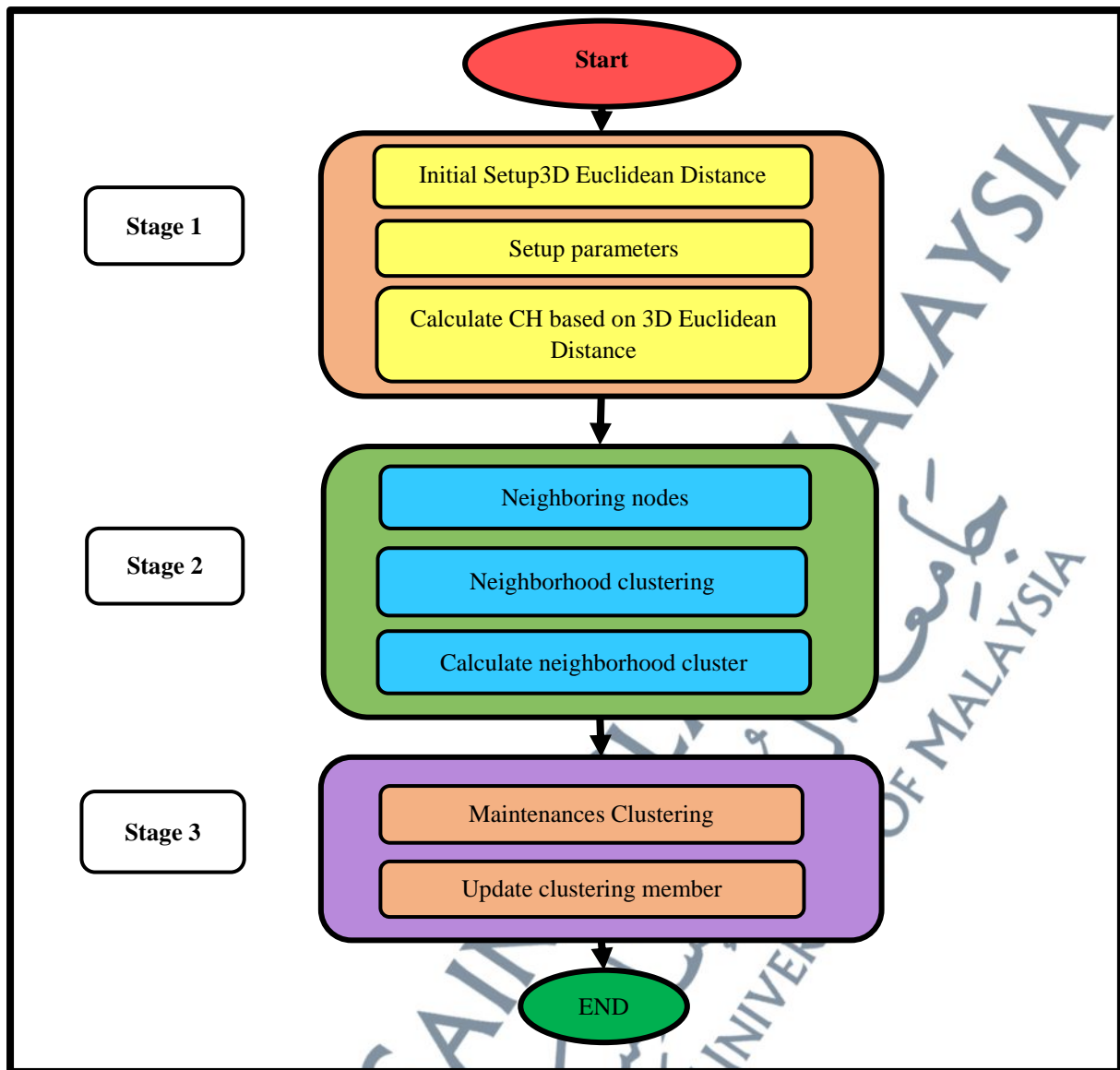


Figure 5.2: Enhancement clustering algorithm flowchart

5.2.1 Stage 1: Initial setup for Enhancement Clustering Algorithm using 3D Euclidean Distance

In this step, the design enhancement clustering algorithm using 3D Euclidean Distance based on the correlated degree that connects all coordinators of nodes to the sink node, while maximising data and reducing distances. The node selects a single neighbour node at each cell of correlated degree for each notification from the sink nodes. Due to differing times in the event region, neighbouring nodes selected from the set of neighbour nodes are only those with

fewer notifications and belong to the same CH. If there is a tie, the neighbouring node with higher energy level transmits the data packets containing energy level field. Next, the initial 3D Euclidean Distance is calculated by incorporating the parameters of the nodes, such as number of nodes a , Initial Energy ℓ , Packet β , Radius/ Distance D , Energy Threshold T , and correlated degree ζ , in respective nodes, to gain the initial CH. The following lists the steps:

Step 1: Calculate the parameters of the node. The parameters are evaluated periodically in order to predict the future state of the network.

Step 2: Compute the correlated degree in 3D Euclidean Distance, where CH is selected if the correlated degree exceeds the energy threshold.

Step 3: The energy threshold $T(a)$ is selected randomly from correlated degree (Neamatollahi & Naghibzadeh, 2018). Justify, each node selects a random number between 0 and 1 and if the random number is greater than the threshold value then it is elected as a CH for that specific round. If the sensor node is greater than threshold $T(a)$, then it is selected as the CH. The threshold value varies in each round based on 3D Euclidean Distance.

Step 4: Calculate the energy level of the node when the energy is evaluated periodically to allow adaptation to the future state of the network so that the total energy left can be expressed.

Figure 5.3 shows the formation of cluster and the stage of its members. Each CH defines its neighbours at two hops and maximum. At this stage, the neighbour nodes form the cluster members. Next, each CH stores all information about its members, and all nodes record the CH identifier. This exchange of information enables the routing protocol to function in the cluster and between the clusters.

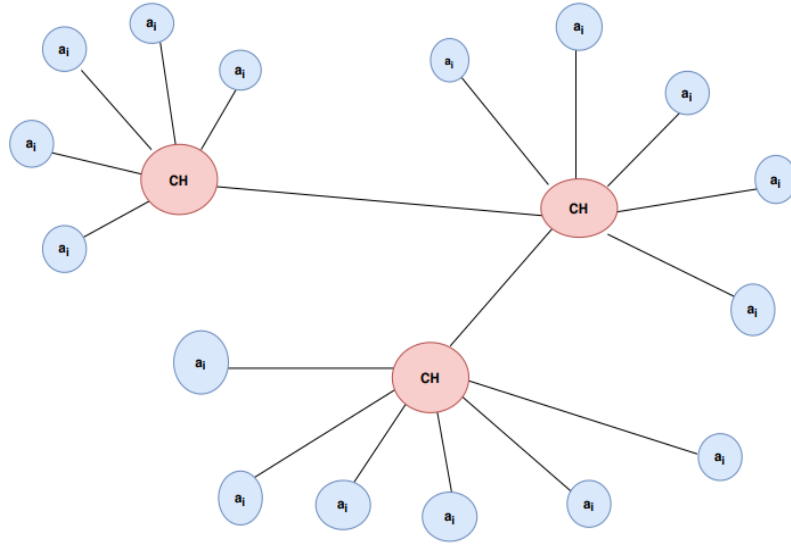


Figure 5.3: Formation of Cluster

In order to determine the initial clusters, the distance between each node and all other nodes is computed, as represented by the neighbours in the domain using 3D Euclidean Distance in Equation 4.5. This representative sensor node has selected in each merged CH, measured distance between node f and anchor nodes a_1 to a_4 based on Correlation Coefficient the correlated degree value in Equation 4.8. Here, the correlation between the parameters (distance, packet, & energy) is positively correlated if $\zeta < 1.0$ based on 3D Euclidean distance, but negatively correlated between the parameters if $\zeta > 1.0$. Finally, the initial CH are obtained by averaging all the nodes in each point set. The 3D Euclidean Distance is used to determine the proximity of each node to the CHs. The clustering algorithm should be independent of the number of iterations to prevent computing distance of each node to CHs repeatedly, especially to save running time and to minimise computational complexity. In the literature, the dynamic means algorithm is proposed to improve the cluster performance and to optimise the number of clusters. Some studies by Na, Shi, Xumin, Liu. (2010) Bholowalia, Purnima and Arvind (2014) and Bhatia, M P S Ph and Khurana. (2013) have

assessed the clustering algorithm as well. In the correlated node, clustering the algorithm can be applied to compute new CHs by increasing the CHs by one in every iteration until they satisfy the validity of the cluster performance.

5.2.2 Setup Parameters

Table 5.1 presents the parameters required for the initial clustering process:

Table 5.1: Setup Parameters

Parameters	Symbol	Value
Number of nodes	$a_1, a_2, a_3, \dots, a_{100}$	100
Correlated degree	ζ	0.0 to 1.0
Initial energy	ℓ	0.5 Joules
Packet size	β	100 bytes/packets
Radius/distance	D	200 meters
Energy threshold	$T(a)$	≥ 1.0
Cooperative nodes	κ	0.1 to 0.69 connectivity range
Malicious nodes	δ	> 1.0 connectivity range
Failures nodes	θ	< 0.0 connectivity range
Selfish nodes	μ	$0.7 \geq \mu < 1.0$ connectivity range
Broadcasting range	BR	15m
Energy per message	ℓ_{pm}	10 aJ/bit/m ²
Energy per step	ℓ_{ps}	50 aJ/bit
Energy produced	ℓ_P	5 aJ/bit

Cooperative node is defined as nodes connected where the connectivity value is between 0.1 and 0.69. Failure node is defined as inactive node where the connectivity value is < 0 . Selfish node is defined as the nodes that is unable to establish any communication where connectivity value is between 0.7 and 1.0. Malicious value is defined as a node that is involved in any suspicious activity where the connectivity value is > 1.0 .

5.2.3 Cluster-Heads (CHs) of Selection

In the clustering algorithm, competition for CHs no longer depends on a random number as in cluster as displayed in Figure 5.4. The CH for correlated clustering is formed based on 3D Euclidean (see Chapter 4 for calculation). Each node broadcasts its value parameter to all its neighbours in the same transmission range. Each neighbour node receives the broadcasted message and estimates its distance from the strength of the signal received.

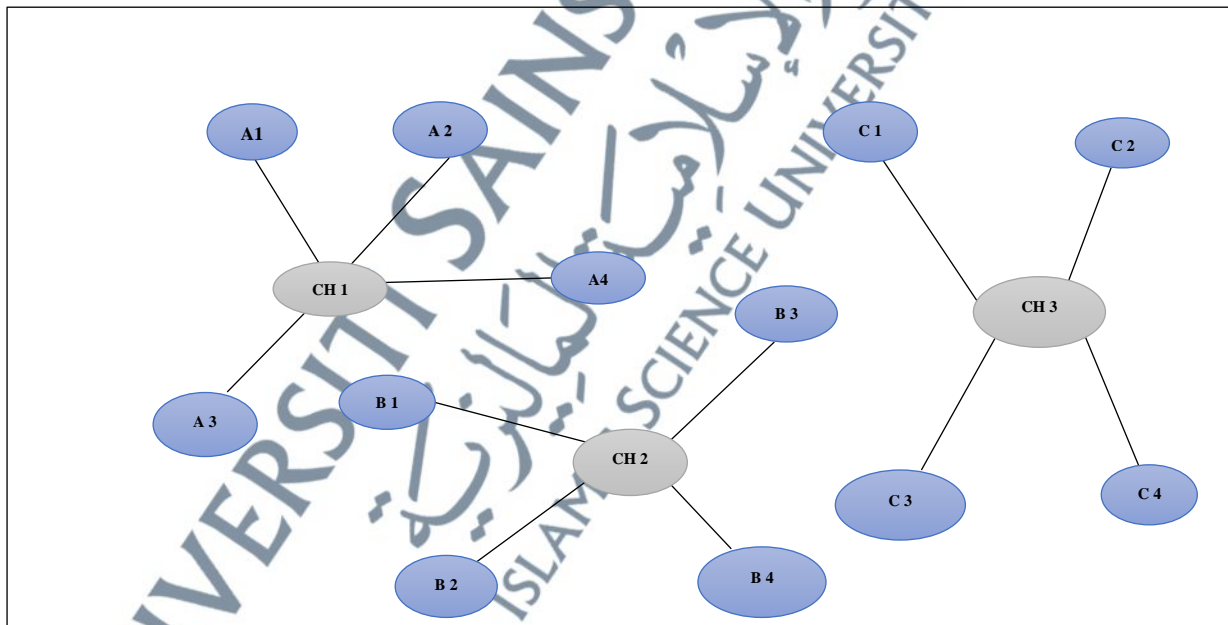


Figure 5.4: Cluster Head

Figure 5.5 illustrates a number of clusters. A key motivation is that almost every clustering algorithm finds the clusters in a dataset, even if that dataset has no natural cluster structure of nodes. The CHs are red nodes, while the cluster members are green ones. The nodes in green are called cluster members because they are about to leave the current cluster and hop onto a new one.

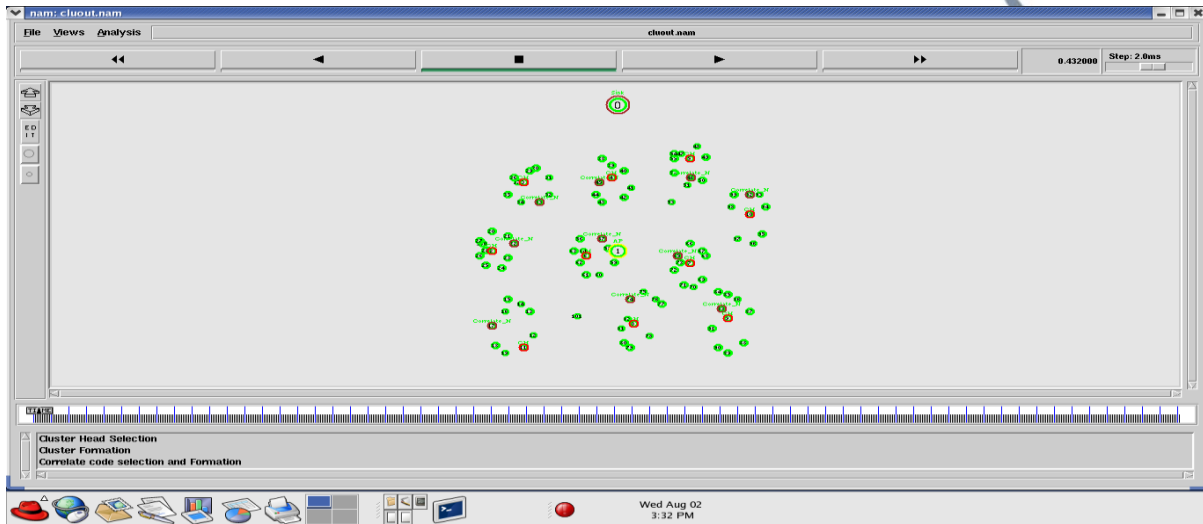


Figure 5.5: Cluster Head Selection

A cluster consists of a CH and several cluster members. The CH is responsible for handling the intra-cluster communication and relays the inter-cluster communication among clusters (Sharma & Nayyar, 2014).

Once a node joins in a cluster and becomes a cluster member, called related to its predicted stay in the cluster starts, where is predicted to become of the member node in the network clustering. Meanwhile, the cluster formation procedure that includes packet transmission value and parameters comparison value, which signifies packet transmission delay as the time taken by the physical layer at the source to transmit the packets over the link.

This delay relies on a number of factors, including the following: The number of active sessions. The physical layer processes the packets in the data order. Hence, if there are node active sessions, this delay becomes quite significant, especially if it does not support the real-

time enhancement clustering algorithms to support network traffic, or the capacity of the transmission of the link. Increasing the transmission capacity decreases the transmission delay. For example, an upgrade from 10 Mbps data to 100 Mbps fast data is bound to reduce the transmission delay by a factor of 10 (Zeb, A., Islam, A. M., Zareei, M., Al Mamoon, I., Mansoor, N., Baharun, S., ... & Komaki, 2016) (Khanvilkar, Bashir, Schonfeld, & Khokhar, 2005).

As direct communication using WSN, the sensor nodes directly transmit their sensed data to the sink without any coordination between two. However, in cluster-based WSNs, the network is divided into clusters when a unique node is the one that is not connected to any other CHs. The cluster-based routing protocols are the CH selection procedure that ensures uniform distribution of energy among the sensors, thus increasing the lifetime of the sensor network (Behera et al., 2019).

5.2.4 Cluster Head Selection Using Energy Threshold (T)

The selection of CH is based on the threshold $T(a)$ value, as determined by Equation 5.1. The CHs can be selected based on 3D Euclidean Distance and it largely affects WSN lifetime. The ideal CH is the one with the highest residual energy, the maximum number of neighbour nodes, and the shortest distance. In WSNs, network clustering for correlated degree is generally preferred, as each round of the collection process will result in sensor nodes $a_1, a_2, a_3, \dots, a_{100}$ as the CHs. Nodes select themselves as CHs based on the energy threshold value of $T(a) < 0.7$.

At the beginning of each round of the process, each node selected is decided by the energy threshold value, ($T(a) > 1$). Before the start of the first round, each node generates a random number between (0-1) If the node is to be greater than a threshold value $T(a)$ given by Equation 5.1, then that node will become CH for that round. With each round, the CH changes

based on the election parameters which behaviour node that all the nodes in the cluster have the same chance to be elected as CH of its energy threshold, a based on the correlated degree. According to Jan, Nanda, & He., (2013), every sensor of the energy threshold value ($T(a) > 1.0$) is compared with the correlated degree. If the correlated degree is greater than energy threshold value, $T(a) > 1.0$, the sensor node will periodically broadcast message to its neighbouring nodes to inform that it will be the CH. CH election process gives rise to the possibility of electing a CH with minimal residual energy which will die out as compared to the one another node has a higher energy level. Therefore, the parameters of each node included in the equation of election probability of CH such that the nodes with higher energy level have a greater chance to be elected as CH. The formula for the threshold value is given as follows:

$$T_a = \begin{cases} k & a_i \geq 1, \\ a_{D,\beta,\ell} - k(R \bmod(a_{D,\beta,\ell})) & a_i < 1. \end{cases} \quad (5.1)$$

where $T(a)$ refers to the energy threshold of node a_i , that serves as the CH. Let a_i denote the number of sensor nodes in a WSN, k represent the number of CHs at each round, and R is the number of nodes round for selection. As a_i would indicate if node a_i has the right to become a CH, a_i must fulfil the condition of $a_i \geq 1.0$, then node a_i is entitled to become a CH. However, when $a_i < 1.0$, node a_i is not entitled to become a CH.

5.2.5 Initial Setup Algorithm 3D Euclidean Distance enhancement in Clustering Algorithm

In the initial setup, all variables are declared to start the network clustering. Next, setup of the parameters is to implement to the algorithm than calculate CH based on 3D Euclidean Distance.

Algorithm 1: Initial Setup 3D Euclidean Distance for the Enhancement in Clustering Algorithm

Initialise network

INPUT: Neighbouring node $(a_1, a_2, \dots, a_{100})$

OUTPUT: Cluster head selection.

1. **Initialise the network**
2. Setup Parameters $(a, \ell, \beta, D, T(a), \ell_{pm}, \ell_{ps}, \ell_p)$
3. Node a broadcast (ℓ, β, D)
4. Calculate 3D Euclidean Distance based on ζ to select CH_i .
5. While $a_i > T(a)$ of ζ select CH_i
6. Otherwise go to line 3
7. Node a_i is selected as CH_i
8. If sensor node $a_i < 0.0$, no CH_i is selected
9. Returns true for a_i then, a_i is selected as CH_i
10. End of Algorithm 1

5.2.6 Stage 2: Neighbouring node

The cluster nodes are selected on the basis of correlated degree in Equation 4.3. As 100 nodes are taken in the simulation of a new protocol, the CH forms a neighbouring node by calculating the correlated degree as follows:

Algorithm 2: Neighbouring node

Algorithm 2: Neighbouring node

INPUT: Neighbouring node $(a_1, a_2, \dots, a_{100}), CH_i$, malicious δ , Selfish μ , cooperative k , failure θ

OUTPUT: Clustering Neighbourhood

1. Find neighbour nodes a
2. Select node with shorter distance D from source node as CH_i
3. Calculate ζ for nodes $(a_1, a_2, \dots, a_{100})$ on CH_i ,
4. If $a_i > 1.0$ then
5. Declare as Malicious node when $a_i > 1.0$ and prohibit from taking part in CH selection
6. If $a_i > 0.7$ and < 1.0 then
7. Declare as Selfish node, it is unable to initiate or have response route discoveries, thus prohibit from taking part in CH selection
8. If $a_i < 0$ then
9. Declare as node failure and prohibit from taking part in CH selection
10. If $a_i < 0.7$ then
11. Declare as node cooperative and allow taking part in CH selection
12. Add into the neighbouring cluster set
13. If number of cluster member > 10 , discard the node
14. End of Algorithm 2
15. Go To Cluster Maintenance

The above depicts cluster formation in algorithm 2. The CHs are selected among the nodes to form a clustering network and in cluster maintenance; the network of cluster nodes is maintained by preventing nodes from malicious attack. Clustering algorithm allows the distance between cluster members and the cluster head, which can effectively expand the cluster coverage. The large cluster coverage can not only effectively reduce the number of cluster head, but also enhance the clustering stability to some certain extent. Unfortunately, as the number of clusters becomes larger, it is increasingly likely that at least one pair of clusters will have only one initial centroid. At this stage, cluster member was set to < 10 nodes because to maintain cluster coverage area. However, for example, the process of forming a stable cluster that is, the maintenance time of the cluster head and cluster members in a cluster is stable, and

the energy consumed by the frequent exchange of information will consume more energy than without cluster (Zhang, D., Ge, H., Zhang, T., Cui, Y. Y., Liu, X., & Mao, 2018) and (Toor, A. S., & Jain, 2019). Unfair distribution on large number nodes leads to one CH that manages big cluster members; some clusters have a small number of cluster members, and the rest of the nodes will not be covered by any CH control. CH that serves a large number of sensor nodes will be loaded and die rapidly in comparison with CH that has limited nodes (Al, Ayman, & Sayed, 2018).

Draining nodes affects the life of the network, thus affecting the network performance. The calculation of sensor nodes a_1, a_2, \dots, a_{100} is based on 3D Euclidean Distance. Each sensor node maintains a list of its neighbours, and their correlated degree values from the network clustering are updated each time a local decision on an event is made to reflect the correctness of their reports. Two nodes become neighbours if their distance is below the correlated degree value. When correlated degree > 1.0 , it represents the malicious node.

The effect of malicious nodes δ value ranges from 1.0 to 1.5 or > 1.0 . Similarly, Rehman et al. (2017) incorporated correlated degree with energy threshold values. A CH of 10 nodes was deployed to determine the value of each node in the whole network, in order to identify the avoidance rate of malicious nodes selection. The effect of malicious node value $\delta > 1.0$ is based on the Energy Threshold value ≥ 1.0 , regardless of the actual sensor reading. The values of malicious nodes reported against their readings refer to Energy Threshold value $T(a)$. The effect of failures node θ values is $a_i < 0.0$. If the Threshold value is more than 1.0, the reported malicious nodes δ against their values refers to correlated degree that is greater than 1.0. This prevents the selection of a malicious node as CH more efficiently due to node behaviour measurement via avoidance of unauthorised node to join any cluster.

After calculating the correlated degree calculation, the node is classified based on the correlated degree value. Equation 4.4 classifies the threshold level of the node as a malicious

node. When the threshold value less than 1.0, it becomes a member cluster. If the malicious node is more than 1.0, the node is declared as malicious and is exempted from the CH selection process. The effects of Selfish node, μ values range at ≥ 0.7 and < 1.0 and cooperative node, k values range at 0.0 to 0.69, respectively. Next, the node is declared as malicious and is exempted from the CH selection process (Rehman et al., 2017).

5.2.7 Stage 3: Cluster Maintenance and Update

Based on the clustering algorithm, a new CH must be selected in every round, thus the need to form new cluster. Clustering maintains the transformed data after clustering formation is performed. The clustering algorithm is a grid correlated degree based 3D Euclidean distance that can identify clusters of arbitrary shape. The clustering maintenance is conducted every 10 seconds of recycling (Azni, Ahmad, & Seman, 2016). In WSN component clusters, the incoming data streams apply Clustering Algorithm 3.

First, if several nodes in the cluster have the highest correlated degree, the node must be added to the start CH for all its neighbours to see if any of them contains the required data, or else, the search is forwarded to the neighbour with the highest degree. For example, consider the cluster topology.

Each node broadcasts CH, cluster member, and packets periodically. Assume that CHs are selected using three parameters (distance (D), packet (β), and energy (ℓ)), all sensor nodes in 3D Euclidean Distance made up of $a_1, a_2, a_3, \dots, a_{100}$, and CH as the anchor nodes. Consider distance is represented as (D), packet transfer (β), and energy (ℓ). The T_{a_i} represents the threshold value to become a CH or otherwise. The recalculation of CHs based on the 3D Euclidean Distance function uses three criteria; distance (D), packet (β), and energy (ℓ), periodically from CHs, while correlated degree is needed to recalculate the cluster member. If the correlated degree is lower

than the energy threshold value $T(a)$, the sensor node will periodically broadcast message to its neighbouring nodes to inform that it will be the CH.

5.2.8 Algorithm 3: Cluster Maintenance

The details of every step are described in the following to propose cluster maintenance.

Algorithm 3: Clustering Maintenance

INPUT: The CH_i , Neighbouring node $(a_1, a_2, \dots, a_{100})$

OUTPUT: New Clustering Neighbourhood

1. Each node broadcast cluster head CH , cluster member a , and packets periodically.
2. Recalculate $(a, \ell, \beta, D, T(a), \ell_{pm}, \ell_{ps}, \ell_p)$ periodically from CH_i , and recalculate ζ for cluster member a
3. if $a \in a_i$ and $a_i < 1.0$, no ζ , then
4. a_i replaces CH_i to become a new CH.
5. If node $a < 0.0$, no CH_i
6. Go to algorithm 2
7. End of cluster Maintenance Algorithm

In Algorithm 3, the Enhancement Clustering Algorithm is run as long as data transmission is present during the maintenance in order to protect the cluster communication. After receiving the broadcast packets, the capability metric is recalculated to identify which node should be changed into CH, so as to maintain cluster stability.

5.3 Experimental Study

This section presents the experimental study that looked into the effects of parameters in WSN. Consider the network of a sensor nodes and a CH distributed over the area. The wireless nature of the sensor networks model causes all sensor nodes to be deployed randomly

within the coverage area. In this study, 100 static nodes in a wireless network were randomly and uniformly distributed over a 3D square within area $A=1000\text{ m} \times 1000\text{ m}$. The node transmission radius, denoted by $r=200\text{ m}$, was assumed to be identical for all nodes. The underlying communication graph of the WSN is modelled by employing $G(a, D)$, where a denotes the vertex set with $|a|=a$, and an edge that exists between two vertices only if their distance is below D .

This experimental study assessed the neighbouring node of energy consumption for WSNs with network parameters using 3D Euclidean Distance. The WSNs often use individual node power efficiency as the key performance parameter, mainly because the battery power of individual nodes can lead to multiple failure types within the network. The study applied experimental-based simulation of Enhancement Clustering Algorithm using NS-2 and C++ programming. The correlated degree in the clustering algorithm contributes to the connectivity of the neighbouring nodes.

5.3.1 Design and Setup

In this simulation, some commonly used WSN simulators are described in the following and the characteristics of these simulators are compared in terms of different aspects, apart from energy evaluation ability. The NS-2, which is the most popular and widely used general-purpose network simulator, supports simulations for various kinds of wired and wireless networks, such as IP-based network protocol and 802.11 standard-based wireless network. The NS-2 is a discrete event and object-oriented simulator, written in C++ and OTcl (an object-oriented version of Tcl). The OTcl is used as the script language to create the simulation environment and to control the simulation processes (Delaney, 2015). The trace file used for data storage is generated during the simulation and the results in the Trace Output trace file can be graphically observed and analysed via Network Animator (NAM). With exceptional

extensibility, the NS-2 has been maintained by a large number of researchers and many third party add-ons can be adopted to address emerging problems. However, NS-2 still lacks appropriate 3D Euclidean Distance, packet formats, sensing models, protocols, and a configurable GUI, apart from the requirement of expert skills that seems unsuitable for beginners. Additionally, both exponential time and execution overhead affect the performance of NS-2 in terms of scalability for large sensor networks. The energy consumption model in NS-2 is very simple, as an initial amount of energy is given to each node, and after packet transmission and reception, the corresponding energy will be subtracted from the total energy until the later value hits zero.

5.3.2 Experimental Scenario for Correlated Node

In order to correlated the clustering algorithm, four scenarios of correlated node behaviour move selection for experiment. Such as cooperative node, selfish node, malicious node, and failures node. The CHs are selected stochastically at first and then adjusted based on the sensor node to maintain their distance. Here, the scenario is restricted to highly correlated region environments and the construction of sensor node takes several rounds in the cluster. Hence, in order to achieve network connectivity of WSN, the scenario demands degree-dense deployment of sensor.

5.3.2.1 Scenario 1: Cooperative Node

Scenario 1 describes the cooperative node, whereby all sensor nodes $a_1, a_2, a_3, \dots, a_{100}$ were calculated, and the CH based on correlated degree decided if the node itself should be a CH or otherwise. All nodes need to hold up before broadcasting CH declaration messages to evade extreme crash and conflict among the nodes. In a scenario where selfish node drops all traffic forwarded through them, the correlated degree value ≥ 1.0 does not affect the results obtained

for isolation. Network disruption is a side effect of the behaviour of a selfish node, while disrupting the network is the intent of malicious nodes. One way to recognise and isolate such disruptive node behaviour is through trust management mechanisms. This work focused on the detection and isolation of selfish nodes in WSNs clustering.

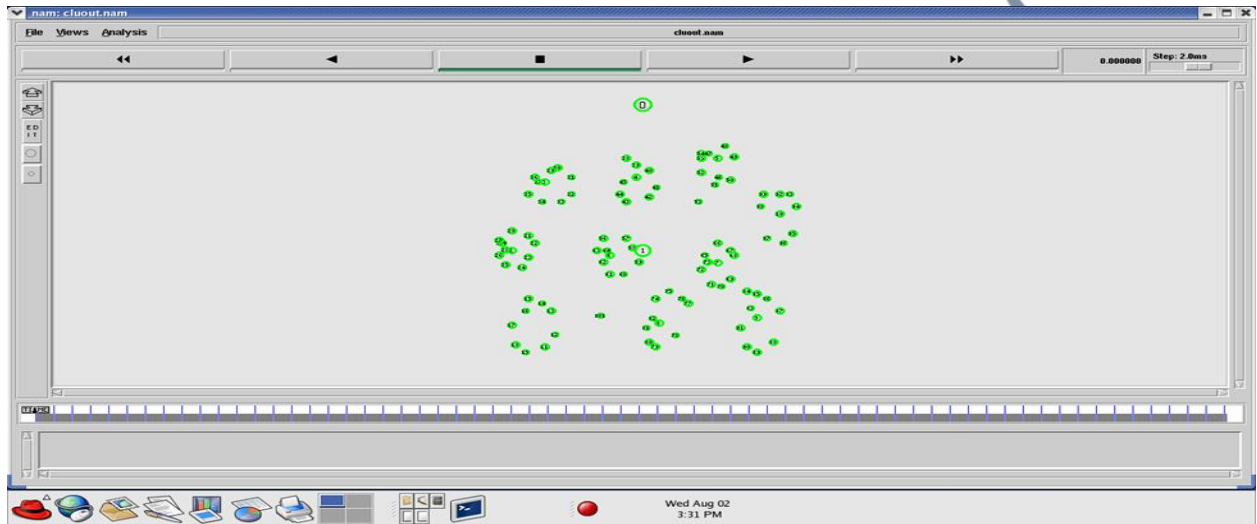


Figure 5.6: Cooperative nodes

In Figure 5.6, assume that each node has one-hop connectivity, which means that each node can communicate with its direct neighbours found in the one-hop communication distance in terms of radio range distance D . In order to select the CHs, Algorithm 1 in Section 5.2.1.4 was employed. For example, a scheme determined the CHs based on Euclidean distance algorithm calculations. Let's denote the number of established CH connections for node a_n with the total number of a nodes in the network. Each node calculates its own correlated degree value and sends it to all its neighbours. After receiving the correlated degree values from its neighbours, a , where $a \neq i$, for all $i = 1 \dots a$.

A specific situation is when the nodes are physically distributed to form a line, as presented in the Enhancement Clustering Algorithm. This arrangement occurs in the case when

all the nodes have a correlation with the nodes of connectivity network in WSN, thus becoming the determiner of nodes level. In such a case, it can be considered that the correlated degree of all the neighbouring, except the failure node in the cluster, is less than a predefined threshold. Hence, all will be classified as next level nodes that are capable of controlling their neighbours. In this case, if two nodes share similar value, one of them can be chosen as the CH, while the topology will change to a correlated degree value, which reflects the enhancement clustering algorithm.

The measure that tells how good WSN is connected is based on 3D Euclidean Distance between any nodes in the network, whether this communication is single-hop or multiple-hop from the source to the destination. Measuring and deciding on the connectivity in a sensor node in WSN is complex as the distance nodes. This complexity is first due to the unpredictable distance node, D , in a given instant of time due to the nodes $(a_1, a_2, \dots, a_{100})$ movement pattern, and second, due to the correlated degree by transmission range, r , which may be affected by interferences, thus the theoretical range may not coincide with the effective one.

Nodes of the cluster have varied roles based on their responsibilities in the cluster. The topology may change often due to the fact that nodes can join and leave the network because that changes their transmission range. This makes it possible to assume that the network is represented by geometric random graph $G(a, k)$, where a represents the set of all the nodes in the network, and k is the set of the entire correlated node. For example, it was assumed that the connection of sensor node a_j and sensor node a_i to be equal. The connectivity in graph theory considers the probability a_{ji} of communication between two nodes a_j and a_i to be equal to 1, which means that they can reach each other (Eletreby & Ya, 2019). Here, to measure network connectivity based on correlated degree if every pair of nodes in the network is connected by either a single-hop or multiple-hop, which is known as 3D Euclidean Distance. This is measured using, where a is the entire number of nodes in the network, CH is the

number of clusters with connected pairs of nodes, and is 3D Euclidean Distance, if the 3D Euclidean Distance is at most the communication range is correlated (Chen & Zhong, 2015) (Sajadian et al., 2011) (Perur & Iyer, 2006).

If the a nodes form CH component with a set of sensor node parameters (distance(D), packet(β), and energy(ℓ)), it denotes the corresponding value of correlated degree based on Equation 4.4. This method of connectivity measurement measures only the availability of correlation between nodes in the network. Thus, for the whole network, the connectivity value ranges from 0.0 to 1.0, where 0 implies no connected nodes, whereas 1 implies connected nodes in the network. However, the goal is not only to know if two nodes can reach each other, but also how well the network is connected and how well the CH is correlated to each node in the network. For this reason, the weights for the three factors that contain parameters with the same values were used in the correlated degree to form the clusters. Based on correlated degree, the calculation of 3D Euclidean Distance measure is added.

Taking into account the above observations, this work proposes the concept of Enhancement Clustering Algorithm based on the calculation of 3D Euclidean Distance based on correlated degree, and all on the circle with radius (Carrabs, Cerulli, Ambrosio, & Raiconi, 2015) (Baidya & Bhattacharyya, 2012), can be applied to measure network connectivity. The correlated degree is used to attribute a weight to each node considered in the calculation of the 3D Euclidean Distance, providing a measured parameters, where node f is the set of all the nodes in the network, regardless with connection or otherwise, and k is a subset of a that has all the misbehaving nodes connected. $\ell_{a_i}, \beta_{a_i}, D_{a_i}$ are the parameters of single-hop communication between nodes for the correlated node a . (Rasheedl et al., 2013) asserted that based on the distance between a and CH , the nodes send their data to CH for each process in different rounds near the sink. Table 5.2 shows correlated degree value for cooperative node refer in the network. And selected in the scenario to experiment.

Table 5.2: Correlated degree of cooperative nodes

No node	Distance (Meters)	Packet (bytes)	Energy (Joule)	Correlated degree
1	200	96	0.552	0.2
2	197	94	0.521	0.2
3	185	90	0.521	0.8
4	190	93	0.512	0.5
5	195	100	0.613	0.3
6	189	86	0.523	0.4
7	197	94	0.621	0.3
8	195	99	0.602	0.2
9	190	96	0.514	0.5
10	187	93	0.624	0.9
11	183	96	0.512	0.5
12	197	94	0.524	0.8
13	189	86	0.524	0.9
14	191	93	0.514	0.8
15	183	92	0.512	0.3
16	190	91	0.541	0.7
17	183	92	0.523	0.5
18	198	84	0.521	0.3
19	184	90	0.502	0.4
20	191	91	0.521	0.5

As shows a cooperative node to search for CH according to the scopes, the correlated value of the selected cooperative node will be almost a correlated cluster the maximum number of hops increases, the nodes in the network achieve more communications with the other member nodes, and as a result, the network requires less number of CHs. It needs to reach the network to become CH. However, if a node selects as CH, it should have an energy threshold

value between 0.0 and 1.0. If the correlated degree is above the threshold value, the node may not be a CH. To support this hypothesis, a ran the CH selection algorithm on a 20 nodes network clustering for three different maximum number of hops one, two, and three. The resulting total number of neighbours, connectivity indices, voting results, and the selected CHs are shown in Figures 5.6, respectively. Accordingly, Table 5.2 shows the correlated degree plot of maximum number of hops versus the total number of node selected CHs, for network clustering. The correlated degree is above the threshold value, the node may not be a CH. Once a node is selected as CH, it announces its decision at the nodes of its immediate region. Based on the strength of the CH advertising message, a cooperative node chooses its CH.

5.3.2.2 Scenario 2: Selfish Node

The second scenario simulated the selfish node behaviour in an experimental setup with 100 nodes. It defines nodes present far from CH as misbehaviour nodes and it offers them an extra amount of energy. Now, in each cluster, the nodes present at the corners have more energy levels compared to sensor nodes. The Enhancement Clustering Algorithm proposed in Chapter 5 was implemented to form the network cluster. In order to form the cluster, CH is needed to determine the cluster region. Based on the algorithm, the CHs were selected using three parameters (distance D , packet β , and energy ℓ). In the selfish scenario, 20 nodes were selected randomly to become selfish, as their energy level was set below 0.5 *joules*. The CH was selected based on the data sense from the nodes within the radius distance of 200 *meters*. The CH was selected based on the packet delivery ratio from the nodes within the packet size of 100 *bytes*. The CH later formed the clusters based on the calculated correlated degree. Figure 5.7 illustrates the formation of ten clusters. The blue node is the selfish node. The selfish node did not take part in the formation because the selfish neighbours were reluctant to broadcast the route request for CH.

Therefore, the number of CHs of cooperative nodes (κ) must be equal to that of misbehaviour nodes. According to Shankar & Shanmugavel (2014), a cooperative node starts to change its status to selfish node when its energy level (0.5 Joules) falls below the energy threshold value $T(a)$. CH selection process gives rise to the possibility of selecting a CH with minimal residual energy which will die out as compared to the one cooperative node has an initial energy level. Therefore, the parameters of each node included selection probability of CH such that the nodes with initial energy level have a greater chance to be selected as CH. Thus, the node is considered as a misbehaving node in selfish nodes, as illustrated in blue circle, while the cooperative nodes (κ) are in green circle. When the node initiates a route discovery to another node, the selfish neighbour node may be reluctant to broadcast the route request. In this case, node behaves like a failure node. It is also possible for the node to forward control packet. However, the situation could become worse as the node may be selected as the next hop and data sent to it. Consequently, the node may discard all data to be forwarded, which hinders communication between node and node. When all node neighbours are selfish, the node is unable to establish any communication with other nodes at a distance of more than one-hop away. In this case, a node is isolated by its selfish neighbours and the network becomes partitioned.

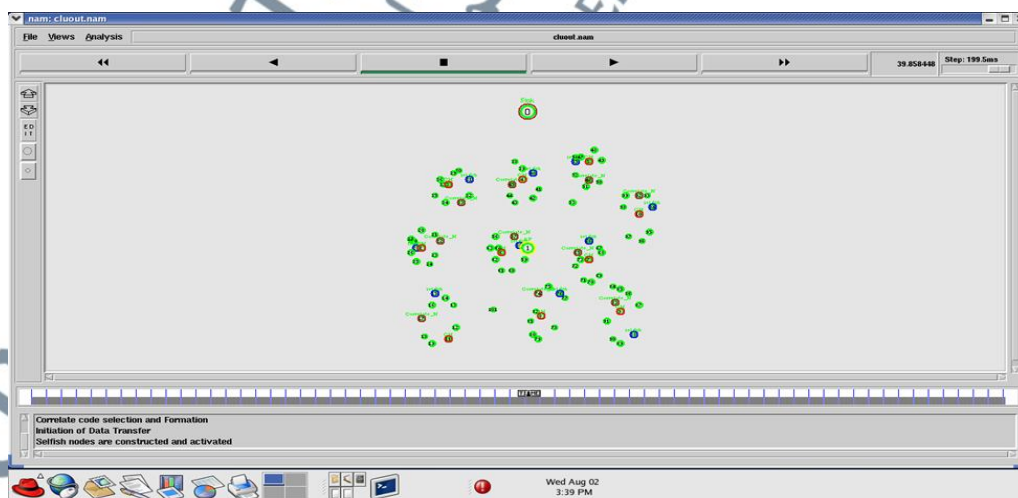


Figure 5.7: Selfish nodes

Table 5.3 presents the number of selfish nodes being detected, in which the message path is recognised and sent to the destination. Then 20 nodes are different node from cooperative node.

Table 5.3: Correlated degree of selfish nodes

No node	Distance (Meters)	Packet (Bytes)	Energy (Joule)	Correlated degree
1	200	16	0.252	0.2
2	170	24	0.821	0.1
3	163	40	0.321	0.1
4	190	53	0.812	0.2
5	110	113	0.213	0.1
6	100	16	0.323	0.2
7	170	24	0.021	0.1
8	195	49	0.602	0.1
9	90	36	0.514	0.1
10	181	13	0.124	0.2
11	100	16	0.812	0.1
12	170	24	0.224	0.2
13	189	46	0.324	0.2
14	191	33	0.514	0.1
15	130	12	0.312	0.1
16	100	61	0.441	0.2
17	173	42	0.323	0.1
18	98	54	0.521	0.2
19	184	30	0.902	0.1
20	100	11	0.521	0.1

If a selfish node is used to search CH based on parameters, the correlated value of the selected selfish node will almost be a correlated cluster. The positions of the CH and the WSN

nodes are as shown in Figure 5.7. CH is located in the far field of the network clustering for WSN, meaning that the distance between nodes for selected CHs and the neighbor nodes is greater. The simulation parameters are the same as shown. β is chosen as *100 bytes packets*, ℓ as *0.5 joules* and the D as *200 meters*. By using these parameters, we ran the simulation for correlated degree selfish node of the maximum hops zero, one, two, and three. Figure 5.7 shows the selfish node behaviour of the network clustering with respect to maximum hops. Because the node itself chooses to be a CH, its energy threshold value should fall between 0.0 to 1.0. If the value of the correlated degree is greater than of the energy threshold, the sensor node will periodically broadcast message to its neighboring nodes to inform that it will be the CH. Once selected as CH, the node announces its decision to the other nodes in the vicinity. Accordingly, a conclusion that, as the maximum hops decreases, the lifetime of the network clustering decreases. From the correlated degree of the cluster, that can deduce that this decrease connected node at a certain number of hops of the cluster heads. This is the point where each node cannot reach any node in the network hops in this case.

5.3.2.3 Scenario 3: Malicious Node

The scenario simulated the malicious node behaviour. The Enhancement Clustering Algorithm proposed in this chapter was implemented to form the network cluster. To form the cluster, CH was needed to determine the cluster region. Based on the algorithm, the CHs are selected using three parameters; distance D , packet β , and energy ℓ . In the malicious scenario, 20 nodes were selected randomly with energy level set below *0.5 joules*. The CH was selected based on the data sense from the nodes within the radius distance of *200 meters*. The CH was selected based on the packet delivery ratio from the nodes within the packet size of *100 bytes*. The CH later formed the clusters based on the calculated correlated degree. Figure 5.8 shows the formation of ten clusters. The brown node is the malicious node, which did not take part in the formation as the neighbours were reluctant to broadcast the route request for CH.

For the scenario of malicious node, the experimental setup incorporated 20 nodes, and a malicious node that broadcasts a routing beacon with an extra high power could lead a large number of nodes to attempt to use it as their next hop in their route to the sink. However, those sufficiently far away would be simply sending their messages into the oblivion. A similar scenario resulted from a wormhole attack. A malicious node could convince nodes that are normally multiple hops from the sink node that is just one-hop away. These nodes would try sending their packets directly to the sink node, which would not be able to hear them.

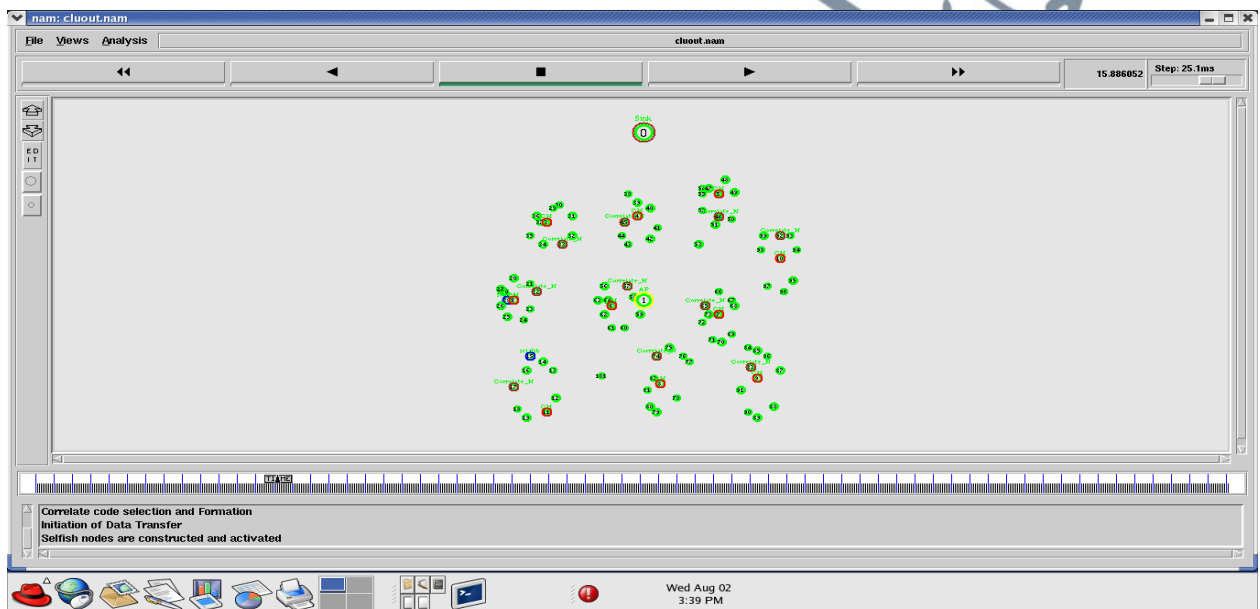


Figure 5.8: Malicious Node

The next effect exerted by malicious node can be illustrated through a scenario with one or more malicious nodes in the neighbourhood of nodes. These malicious nodes are active both in route discovery and launching attacks. If malicious nodes launch DoS attack on the network layer with cooperative node in the routing stage, they become reluctant in forwarding data packets, or disrupting legitimate path selections by broadcasting fake route replies. Thus, if a node has a node as the next hop, communication would be lost with nodes at least two-hop

away. When all node neighbours are selfish nodes, the node gets isolated by malicious neighbours, which is similar to having selfish neighbours (see Table 5.4).

Table 5.4: Correlated degree of Malicious Nodes

No node	Distance (Meters)	Packet (bytes)	Energy (Joule)	Correlated degree
1	134	67	0.3252	0.2
2	35	56	0.2528	0.4
3	51	86	0.2129	0.4
4	47	58	0.1278	0.3
5	87	76	0.1413	0.3
6	56	41	0.2179	0.2
7	24	42	0.2137	0.3
8	45	34	0.2067	0.4
9	46	36	0.1354	0.4
10	84	34	0.3024	0.2
11	65	54	0.2608	0.4
12	40	61	0.2872	0.4
13	95	33	0.2019	0.3
14	133	45	0.1104	0.4
15	127	41	0.2031	0.3
16	87	67	0.1365	0.3
17	75	57	0.1278	0.3
18	90	34	0.1386	0.2
19	120	47	0.1488	0.4
20	59	55	0.2119	0.3

If a malicious node to seek CH based on perimeters, the correlated value of the selected malicious node will almost be a correlated cluster. A malicious node may fail to be a CH for

several reasons, such as energy exhaustion and misconfiguration. It is also prone to be configured from being selfish for the sake of power saving, or to become a malicious node. To investigate this, considered the same network shown in Figure 5.8 along with the simulation parameters the same as. The only parameter that is different here is the maximum hops. There is a total number of CHs and the total lifetime of the network. There is an optimum number of CHs which lead the network to survive the most total lifetime possible without having any partitioning. An energy threshold fall between 0 and 1 of the network clustering. As the energy increases, more data need to be processed and more packets need to be transmitted causing more power to be spent; therefore, the total lifetime of the network decreases.

A node that wishes to become a CH should have its energy threshold fall between 0.0 and 1.0. If the correlated degree is below the energy threshold value, the sensor node will periodically broadcast message to its neighboring nodes to inform that it will be the CH. Once a node is selected as the CH, it advertises its decision to the nodes in its close vicinity. Then, started the simulation with 20 nodes, and then, each time, that removed one of the end nodes and repeated the simulation till are left with one node in the network. As a result, Table 5.4 shows the correlated degree of malicious nodes behaviour of the network with respect to the total number of nodes in the network. Accordingly, conclude that there is a certain number of nodes 20 nodes in our case in the network that allow the network clustering to achieve.

5.3.2.4 Scenario 4: Failure Node

The scenario simulated the failure node behaviour. The Enhancement Clustering Algorithm proposed in this chapter was implemented to form the network cluster. Hence, CH is needed to determine the cluster region. Based on the algorithm, the CHs were selected using three parameters; distance D , packet β , and energy ℓ . In the failure scenario, 20 nodes were selected randomly with their energy level set below 0.5 joules. The CH was selected based on

the data sensed from the nodes within the radius distance of 200 *meters*. The CH was selected based on the packet delivery ratio from the nodes within the packet size of 100 *bytes*. The CH later formed the clusters based on the calculated correlated degree. Figure 5.9 illustrates the formation of ten clusters. The yellow node reflects the failure node. The failure node did not take part in the formation because the cluster for failure node was almost zero, which implies that the network is disconnected almost surely; while an almost sure connectivity is achievable only with failure node. The failure neighbours were reluctant to broadcast the route request for CH.

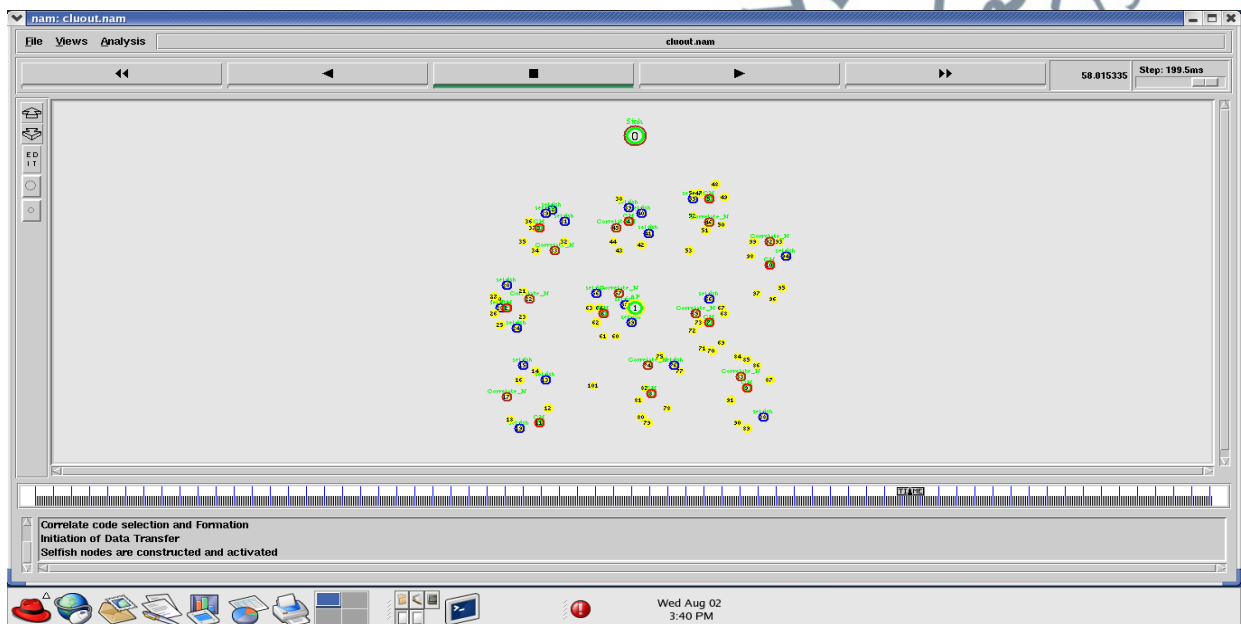


Figure 5.9: Failure Node

The effect of failures node was assessed with an experimental setup with 20 nodes. The failure nodes were enactive both in route discovery and launching attacks. If the failure nodes launch DoS attacks on the network layer with cooperative node in the routing stage, they become reluctant in forwarding data packets, or disrupting legitimate path selections by broadcasting fake route replies. Thus, if the node has a node as the next hop, communication

would be lost with the nodes at least two-hop away. When all node neighbours are failure nodes, the node would be isolated by the failure neighbours (see Table 5.5).

Table 5.5: Correlated degree of failure nodes

No node	Distance (Meters)	Packet (bytes)	Energy (Joule)	Correlated degree
1	250	16	0.052	0
2	217	24	0.012	0
3	263	20	0.025	0
4	210	35	0.087	0
5	310	13	0.018	0
6	250	16	0.031	0
7	317	24	0.035	0
8	245	37	0.027	0
9	350	26	0.049	0
10	241	13	0.038	0
11	270	16	0.025	0
12	279	24	0.042	0
13	369	26	0.051	0
14	439	33	0.059	0
15	230	21	0.037	0
16	350	16	0.024	0
17	273	24	0.129	0
18	258	45	0.013	0
19	244	30	0.009	0
20	310	11	0.015	0

Here, investigated failure nodes to find the CH based on parameters, the correlated value of the selected failure node will not become a correlated cluster because the effect of a

failure nodes rate of the cluster heads in the network based on correlated degree of the network clustering. Considered the same network shown in Figure 5.9 along with the simulation parameters the same as. The only parameter that is different here is the maximum hops. There is a maximum hop count and the total number of CHs. As the maximum hop count decreases, the total number of CHs decreases. The number of CHs and a total node of the clustering. There is an optimum number of CHs which lead the network to die the number node of the cluster possible .which having partitioning.

Then, started the simulation with 20 nodes, and then, each time, that changed the node may select itself to be a CH if its energy threshold value ranges between 0 and 1. The correlated degree is below the energy threshold value, the sensor node will periodically broadcast message to its neighboring nodes to inform that it will be the CH. and repeated the simulation for various correlated degree value of the selected failure node may be a correlated cluster. As a result, Table 5.5 shows the failure behaviour node of the network clustering with respect to the correlated degree of the nodes in the network.

Accordingly, conclude that as the correlated degree of the failure node decreases on the network clustering. This is an expected result, because, as the correlated degree of failure node decreases, lose packets are not sent between the nodes thus more energy is consumed. Also, random selection of CH which results from giving all sensors in the network the same chance to be selected as CH will cause that; a node having less energy may not become CH and will be out of the cluster heads.

5.4 Summary

This chapter presents the development of Clustering Algorithm of correlated nodes behaviour. The algorithm focuses only on energy consumption to improve network connectivity. Correlation clustering offers a method for clustering as a set of objects into the

optimum number of clusters without specifying that number in advance. In order to keep the connectivity of the sensor network, the parameters formulated in Chapter 4 were embedded into the Enhancement Clustering Algorithm to enhance connectivity under correlated node behaviour. The main objectives of this research is to enhance the existing clustering algorithms to improve network connectivity under the event of correlated node behaviour. This research proposes the Enhancement Clustering Algorithm based on 3D Euclidean Distance.

Generally, the correlated degree had managed to further improve the obtained solution. The experiment found that the Enhancement Clustering Algorithm had been able to enhance the network connectivity of the solution. However, the distance between the different nodes that may affect the values could be addressed by implementing the self-adaptive Enhancement Clustering Algorithm. The next chapter further details the results and discussion.