Routing and data aggregation in wireless sensor networks by using clusters

Kanegonda Ravi Chythanya¹, Komuravelly Sudheer Kumar², Bonthala Prabhanjan Yadav³, Madhuri P.M.⁴ and Rajesh Mothe⁵

¹Department of CSE, S R University, Warangal, Telangana, India - 506371
²Department of CSE, S R Engineering College, Warangal, Telangana, India - 506371
³Sumathi Reddy Institute of Technology for Women, Ananthasagar, Hasanparthy, Warangal, Telangana, India - 506371
⁴Department of CSE, GokarajuRangaraju Institute of Engineering and Technology, Bachupally, Hyderabad -500090

E-mail: chythu536@gmail.com

Abstract. Due to the non-unique node distribution energy consumption exists difference on the basis of clusters networks for wireless sensors. Based on this problem, we propose an effective communication and routing data aggregation tree based on the previous clustering architecture. Using fuzzy logic technology, cluster heads are chosen based on residual capacity, node density and heads for load clusters. Via modifying inter-cluster energy consumption, the algorithm for the energy usage of cluster heads between cluster routing balancing. Using the data correlation model, data compression is then applied to minimise energy consumption.

1. Introduction

1.1 Wireless Sensor Networks (WSN)
Using sensor modules, Wireless Sensor Networks are made up of intelligent small Sensor hubs can detect different types of anomalies that transmit the data into an end node. All data is collected and computed by WSNs and different sensing information is given to individual users. Typically, the sensor nodes are mounted in large numbers in areas where manipulation is difficult or impossible (from a few to a thousand). Consequently, sensors must be distributed randomly and using energy limited, for example battery packs, resources. Therefore, sensor nodes must work together and provide energy-efficient transceivers and protocols for energy savings and a long grid existence in order to create a self-organized network[1].

One of WSN’s main missions is to accumulate and relay the surrounding parameters to the Station base. Station base. In hostile conditions, sensors are usually used and batteries cannot substitute, making energy consumption a crucial factor in the design of a protocol. In the WSN, transmitted information is transmitted by a sensor by means of a wireless signal, and all neighbours are transmitted by a sensor. Data aggregation is one of the most effective ways of eliminating overhead communication, and several systems have been implemented to eliminate duplicated broadcasting. [2, 3].

There are thousands to millions of indicators inside a networking sensor, which can be spread across a geographic computer area, networking, and sensing capabilities. Their restricted computational capacity, bandwidth and memory space limit the usage of conventional algorithms for data processing and the space of Medium outcomes on sensor nodes which can be handled. A well organised level of routing in WSNs is also needed[4, 5].
1.2 Cluster based data aggregation in WSN

In the wireless sensor world, data aggregation is an important technique, as data packets are reducing emissions, the service provider life will improve and the efficient data transfer ratio will be increased. In terms of goal monitoring, the principle of clustering can be used to improve efficiency of cluster formation in a patriarchal order. The static and the other dynamic category are the two styles. The network is proactively clustered into multiple clusters in static form. In order to incorporate data rapidly and efficiently into sink portal, aggregation of information through the static cluster uses the pre-selected cluster heads. Networks clustered dynamically reactively developing a cluster near the event detecting vertices. All data are obtained by means of the cluster chief, aggregated and envied at the sink. The vorteil of the technology that’s what it is. the aggregation of data involves only the appropriate nodes and greatly saves energy for other sensor nodes [1].

Cluster-based data aggregation provides many advantages: reliability, information precision, lower redundancy, reduced transport load and energy conservation. Also known as the data aggregator nodes are cluster heads and combine the data to be sent to the base station[6, 7]. A compression model based on a data correlation and a cluster routing algorithm based on relays was proposed and implemented in the WSN cluster to achieve these objectives. This extends our previous work [8] by proposing to optimise life span on WSN with a fuzzy-based optimal clustering protocol. The remaining paper will be arranged accordingly. Section II offers a collection of research relevant to our work. The design approach and the proposed method are explained in Section III. Section IV displays the outcomes and the debate about our possible framework and Section V the closure of the methodology change suggested by comparing the previous techniques.

2. Proposed solution

2.1 Overview

We propose a condensed cluster-based approach in this paper focused on the previous communications clustering architecture. It encompasses the clustering and routing of compact information between groups. The Information integration and data compression process is used during compressed combination as a power-saving method. This calculates the statistics and grouped size compaction ratios. The Cluster Head (CH) gathers the received readings and supplies them to the sink with one message. The volume of the message is dependent for a cluster mutual entropy.
A routing tree has been developed for the elected CHs in the intercluster routing process. The CH selects the next CH with a relay. This transmitter is determined by CH’s residual energy, Cluster volume, bandwidth consumption and the stable node resources. This mechanism balance power consumption between cluster heads by changing energy consumption within the cluster and energy consumption in the inter cluster.

2.2 Optimal clustering protocol

High residual power and low strain should be present in a cluster head (CH) because many processes of data transmission are involved. There should also be a Minimum sink spread and additional nodes. The CH is chosen on the basis of residual (Eres) capacity, sink distance (D), node density (NDi) and charge. The logic of Fuzzy is often used chosen CH because of the fact that all possible combinations of these contradictory steps are considered and CH and the cluster size can be a node correctly evaluated.

The following steps are taken in the clustering protocol[8]

1) At that time t, numerous interim PCHs are randomly chosen to compete for the ultimate heads of the cluster.
2) With the exception of the preferred PCH, the other nodes contain a set HELLO message.
3) Each PCH receiving HELLO message computes the gradient and fills in its data path with the node ID.
4) The parameters including residual energy (Eres), node density (NDi) and load are then collected for each PCH.
5) Uses estimated parameter values each node uses a smooth logic technique to evaluate the parameter condition.

![Figure 1: Block diagram on WSN aggregation and routing of proposed cluster dependent data.](image-url)
6) Whenever a PCH predicts that it can become the head of the clubs, a D Mes message is broadcast that confirms that it wants to be a CH.

7) However, when PCH decides that there is one additional PCH to be specified as CH, it simply announces its DCL Mes post.

8) However, when PCH decides there is still one additional PCH to be specified as CH, it simply announces its DCL Mes post.

2.3 Data correlation model

Statistically equivalent sources of information with assume that the findings are normal with good precision and variance are known as sensor nodes within the WSN. The assumption is that the dependence of the sample readings is exponential, with a gap at a fixed rate, for the grade of data correlation (DDC). Thus, the connexion would be lower if the gap is longer.

Therefore, we describe the following two metric magnitude (MG) and distance (D) for the evaluation of the DDC of the two-time series.

The readings Ni and Nj are said to be two sensor nodes correlated, if

(i) \(MGI = MGj\)

where MGI and MGj are the scale of the Ni and Nj readings

(ii) \(Dij < Dth\)

the gap threshold is where Dij is between Ni and Nj and Dth

The two nodes can be interpreted as 2-dimensional points. The coordinate node Ni is given (MGI, Dij), and the coordinate node Nj is given (MGj and DJ). The DDC between the two nodes is calculated on the basis of the Euclidean Distance estimate.

Euclidean distance assesses the square root of discrepancies between a duo of point positions.

The Euclidean gap between Ni and Nj nodes is therefore determined by [6],

\[
DDC_i = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2}, \quad n = 2.
\]  

Where, \(x_{i1} = MGI, x_{i2} = Dij\) and \(x_{j1} = MGj, x_{j2} = DJ\).

This model is used to assign the optimum data rate set to sensor nodes. The Power Exponential model is used for this model of the data correlation as a covariance function. The exponential power model for covariance with distance \(d\) is indicated [6]:

\[
K_{\mu}^{PE}(d) = \exp \left( -\left( \frac{d}{1} \right)^{\mu} \right),
\]  

Here \(K_{\mu} = (||u-v||) = \text{corr} \{S(u), S(v)\}\) denotes an isotropic correlation function with \(\mu = (t_1, ..., tc) \in \Theta \subset \mathbb{R}^c\) as a set of correlation and smoothness parameters in random area \(S(u)\).

The covariance matrix aspects is sometimes conveyed in the Squared Exponential Correlation model [2]

\[
\sigma_{ij} = \sigma^2 \exp(-\alpha d_{ij}^2)
\]  

Where, \(a = t^2\) is the exponent of the connexion and \(d_{ij}\) marking the distance from the Euclidean nodes Ni to Nj.

2.4 Data compression model

The structure of the message is determined by the joint invariance with the DC model. The combined data streams are automatically compressed and then decrypted on a single recipient.

The data flux in the Ni and Nj nodes should be associated, so that data flux in the Ni node is X and the data flux in the Nj node is Y.

The associated X & Y data flows come out of the joint conditional probability \(P(X = x, Y = y)\) by having n separate designs. In fig. 2 Encoder 1 gets X stream data and then delivers the decoder with a hidden message, each of X’s characters is encrypted with RX bits.
Likewise, Encoder 2 obtains the Y data stream, then transfers the decoder with the hidden message where a number of RY bits are encoded for every Y character. Once these two coded messages are sent, the transmitter produces two $X^*$ and $Y^*$ vectors, which are the estimates of the initial $X$ and $Y$ data streams.

![Diagram](image.png)

**Figure 2.** Two compatible data streams $X$ and $Y$ separate encoding & coding.

The fuzzy numbers $X$ and $Y$ with correlation matrix are calculated by the permissible rate region of,

$$
P(X = x, Y = y) = \sum_x P(X = x) \sum_y P(Y = y)$$

(4)

$2.5$ Inter-cluster routing algorithm

Each CH transmits the networking message with details like the ID, the transmission range, the companies alike and the distance from the BS. The chance with next hop CH is provided by [17]

$$PR_{CHN} = a \frac{E_j}{E_{max}} + (1-a) \frac{1}{(NOCM)_j}$$

(5)

- Where $E_j$ is the residual energy of cluster head $S_j$.
- $Emax$ is the early increasing computational energy.
- $(NOCM)_j$ is the number of $S_j$ cluster members.
- $a$ is the constant of which $[0,1]$ is a value.

In the given equation, inter-cluster travel is defined.

- RT MSG is a pathway response consisting of $I_d$ ($I_d$), residual energy (RE), cluster members number (No CM) and length to BS (Dist BS), respectively.
- DIST_TH - threshold distance.
- $Tc$ - Routing timer for tree building.
- CHNT – Cluster head neighbor table.
- $PRCHN$ – Probability of next hop CH.
Algorithm:

1. CH broadcast a RT_MSG
   \text{CH} \rightarrow \text{RT\_MSG [Id, RE, No\_CM, Dist\_BS]}
2. if Dist\_BS < DIST\_TH, then
   Data can be transmitted directly to BS
3. else
   while (timer Tc not expired)
     Receive RT\_MSG
     Compute the value of PRCHN for each neighbor CH.
     Update CHNT
   End
   For each CHj in CHNT
     If PRCHN = Maximum(PRCHN), then
       Choose CHj as the next hop CH.
     End if
   End for
   If j > 1, then
     Choose CHj with maximum(Dist\_BS)
   End if
4. End if

\textbf{Figure 3.} Representation of inter-cluster routing with the encoded data packets

Each CH selects in this algorithm the following Hop CH with more residual energy, fewer companies alike and less distance from the sink. If the highest payout of PRCHN is more than few CH, the CH with greater distance to the BS is opted for the fast drain of CHs close to the BS, because of the transmission of a large amount of observations.
Upon receipt of data from the members of its cluster, CH compresses the data correspondence model as defined in the previous section, and then sends the aggregate data to the next hop CH. If a CH receives information of another CH, the data is passed directly without aggregation to the chosen next hop. Although it has a high capacitive coupling and a compressed configuration and analyzes data from its affiliates, energy consumption within the cluster has indeed been optimized. The next Hop CH is chosen in the same way, by balancing the inter-cluster energy consumption with high residual energy and a low load. The inter cluster routing of the coded data packets is shown in Figure 3.

3. Simulation Results

3.1 Simulation model and parameters
In simulating the proposed architecture the network simulator (NS-2)[19] is employed. The sensor nodes are randomly used for this simulation around a region of 500 m x 500 m with a transmission range of 40 m. The sink or the BS is located at the rightmost topology alley. Evaluated regarding is used to generate CBR (Constant Bit Rate), and the sensor data are considered to be periodic and constantly sized. On average, the results are 10 runs. Table 1 summarises the simulation and parameters[20-22].

| PARAMETER             | VALUE                       |
|-----------------------|-----------------------------|
| NO. OF NODES          | 25, 50, 75, 100, 125 and 150 |
| TRANSMISSION RANGE    | 250 – 300m                  |
| MAC PROTOCOL          | IEEE 802.11                 |
| AREA SIZE             | 500 Sq. mts.                |
| SIMULATION TIME       | < 1 Min.                    |
| PACKET SIZE           | 1024                        |
| TRAFFIC SOURCE        | CBR                         |
| INITIAL ENERGY        | 10 J                        |
| TRANSMISSION POWER    | 0.660                       |
| RECEIVING POWER       | 0.395                       |
| RATE                  | 300 Kb                      |

3.2 Performance metrics
Compared to the cluster-based correlated data collection (CCDG) technology[2], the proposed protocol for aggregation and routing of compressed data (CCDAR) based on cluster is compared. Package distribution and average energy use evaluation of performance are assessed.
- **Packet Delivery Ratio**: It is the ratio of the number of received packets to the number of sent items.
- **Packet Drop**: It determines the number of packets that have fallen during propagation.
- **Energy Consumption**: The energy consumed by the nodes is the amount to pass data packets to the recipient.
- **Delay**: It is the time the nodes spend transmitting data packets.

3.3 Results
The random number of nodes ranges from 25, 50, 75, 100, 125 and 150 and the performance indicators described above are assessed. Network is formed in a way to make 10 percent of the total number of nodes the size distribution of each cluster.

Figure 4 indicates the delay in the number of nodes in both the protocols. For a high node, the delay increases linearly as the length of the Euclidian distance path is high. However, for inter-cluster routing, the delay is 32% less compared to the CCDG when selecting the shortest routes between each CH. For both protocols, when the number of nodes is increased, the distribution ratio is shown in Figure 5. Packet drop CCDG rises linearly from 7000 to 12800 to 75 nodes, compared to 4600 and 6300 for CCDAR. The delivery ratio for CCDG is similarly reduced from 0.56 to 0.37 and for CCDAR from 0.6 to 0.4. This is
because the load in CH increases, leading to more packet losses, as there are more nodes. Since in inter-cluster routing CCDAR chose the least loaded CH, a packet fall is 43 per cent less and a network throughput rate of 10% is higher than CCDG.

The results of both protocols, when the nodes are lifted, are shown in Figure 6. The figure shows that energy consumption is consistently reaching 25 CCDAR and CCDG nodes. CCDAR has 11% less energy consumption than CCDG, because the CH strategy is based on high cluster head for both intra and inter cluster routing. It crosses approximately 12,500 at 100 CCDAR knots and approximately 23,000 CCDG at 100 knots. It should be noted that CCDAR does not include additional overheads. The overhead for CCDAR is 45% lower in comparison with CCDG.

4. Conclusion

We have proposed in this paper an effective data aggregation tree on the basis of the previous clustering architecture for communication and routing. The tool for energy saving is used here as data correlation and data compression. Then a routing tree is created using the routing algorithm based on the cluster. This routing algorithm on a cluster-based basis balances energy consumption between the cluster heading by changing energy consumption within the cluster. With we showed that the proposed strategy tends to improve network life, performance parameters with a higher supply ratio and lower energy consumption. We intended to deliver a fuzzy and optimising algorithm in future for optimal clustering so that the clustering efficiency could be improve and device performance improved.

5. References

[1] G S Chhabra and D Sharma 2011 International Journal of Soft Computing and Engineering (IJSCE) 1(1) 27-31.

[2] W S Jung, K W Lim, Y B Ko and S J Park 2009 Third International Conference on DigitalSociety, 2009 ICDS'09(1) 112-7.

[3] S Mishra, HThakkar, A Chakrabarty and DKimtani 2012 International Journal of ElectronicsCommunicationand ComputerTechnology (IJECCT) 2(5) 227-30.
[4] P PramodKumar, S Naresh Kumar and Ch Sandeep, 2019 Journal of Mechanics of Continua and Mathematical Sciences 15(6) 201-12.
[5] KKinoshita, TOkazaki, H Toedand KMurakami 20085th IEEE Consumer Communications and Networking Conference CCNC 719-23.
[6] Yua, Y Qia, G Wangb and X Gu 2012 International Journal of Electronics Communication 66(1) 54-61.
[7] A Dabirmoghaddam, M Ghaderi and C Williamson 2010 IEEE International Symposium on Modelling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS) 163-71.
[8] Kumar BV, Chanti Y, Yamsani N, Aluvala S and Bhaskar B 2019 International Journal of Recent Technology and Engineering (IJRTE) 8(2S3) 1121-29.
[9] S. Jothi and M. Chandrasekaran 2014 Research Journal of Applied Sciences, Engineering and Technology, Maxwell Scientific Organization 8(6) 714-725.
[10] Arabelli R R and Revuri K 2019 International Journal of Innovative Technology and Exploring Engineering (IJITEE) 8(5) 1051-54.
[11] S Khalid, A Mason, F B Hussain, H Abbas and A Ghafoor 2014 “Load Balanced Routing for Lifetime Maximization in Mobile Wireless Sensor Networks”, International Journal of Distributed Sensor Networks.
[12] A Takeda, T Oidey and A Takahashi 2012 BWCCA 684-689.
[13] J G and D Chen 2013 Advanced Materials Research 756 1126-30.
[14] S Mathapati and S R Patil 2012 IEEE International Conference on Computing Sciences (ICCS) 153-8.
[15] T Fukabori, H Nakayama, H Nishiya, N Ansari and N Kato 2010 IEEE International Conference on Communications (ICC) 2010 1-5.
[16] MM R Mozumdar, N Guofang, F Gregoretti and L Lavagno 2009 Journal of Networks 4(7), 598-606.
[17] M H Anisi, A H Abdullah and S A Razak 2011 Wireless Sensor Network 3(10) 329.
[18] Shin, M Kim, M WMutka, H Choo and T J Lee 2009 Sensors 9(8) 6028-45.
[19] Ravi Kumar R, Babu Reddy M & Praveen P 2019, International Journal of Scientific and Technology Research 8(12) 2071-74.
[20] Rajasri I, Guptha AVSSKS and Rao YVD 2016 Generation of Egt’s: Hamming Number Approach Procedia Engineering 144 537-542 10.1016/j.proeng.2016.05.039
[21] Mahender K, Kumar TA and Ramesh KS 2017 Performance study of OFDM over multipath fading channels for next wireless communications International Journal of Applied Engineering Research 12(20) 10205-10210
[22] Seena Naik K and Sudarshan E 2019 Smart healthcare monitoring system using raspberry Pi on IoT platform ARPN Journal of Engineering and Applied Sciences 14(4) 872-876.