Optimal Number of Clusters in Wireless Sensor Networks with Mobile Sink

C.P. Gupta and Dr. Arun Kumar

Abstract-Wireless Sensor Networks (WSNs) employ tiny sensor nodes to observe field and send data from the field to a Base Station. The base station is generally at a fixed location away from the observed field. However, such networks have limited life due to draining of batteries of nodes near the base station quickly in comparison to the nodes away from the base station. Further, data from nodes in close proximity have very little or no variation. Sending data from all the nodes to base station will require larger energy due to longer distances or multiple hops. Clustering has been found to be an effective solution for such networks. However, efforts to find optimal number of clusters are limited to WSNs with static sink. In this paper, we present an analysis of optimal number of cluster in such networks. The theoretical analysis is supported by simulation results.

Index Terms: *Wireless Sensor Networks, Base Station, clustering*

————————————— ◆ —————————————

1. Introduction

In recent years, sensor networks have been deployed for a variety of applications. In few cases viz. fire safety in high rise buildings, the sensors are connected through wires. However, wired networks are rare and do not find large applications. In most cases, the deployment prescribes for a wireless environment. As a consequence, the sensor networks use wireless communication to send the collected data and are thus termed as Wireless Sensor Networks (WSN). Formally, a wireless sensor network (WSN) in its simplest form can be defined as a *network of (possibly low-size and low-complex) devices denoted as nodes that can sense the environment and communicate the information gathered from the monitored field through wireless links; the data is forwarded, possibly via multiple hops relaying, to a sink that can use it locally, or is connected to other networks (e.g., the Internet) through a gateway* [1].

As presented by Akyildiz [3], the total radio power consumption P_c is:

$$P_c = N_T(P_T(T_{on}+T_{st}) + P_{out}(T_{on})) + N_R[P_R(R_{on}+R_{st})]$$

where, $P_{T/R}$ is power consumed by the transmitter/receiver; P_{out}, the output power of the transmitter; T_{on}/R_{on}, the transmitter/receiver on time; T_{st}/R_{st}, the transmitter/receiver start-up time and $N_{T/R}$, the number of times transmitter/receiver is switched on per unit time, which depends on the task and medium access control (MAC) scheme used.

Designing routing protocols for WSNs is challenging due to resource constraints [1, 2]. Research on WSNs has found that clustering the nodes in WSNs reduces the energy consumption significantly as the member nodes transmit

data only to a node within the cluster. The node receiving data from all the members within the cluster is termed as Cluster Head (CH). As the CH is within the cluster, attenuation is only of the order of square of distance as compared to higher power over longer distances. In such networks, CH performs data aggregation on the received data exploiting temporal and spatial proximity thus reducing the total data to be transmitted to the BS. Reduction in distance and amount of data to be transmitted increases network lifetime considerably.

Fixed location of base station or a single base station causes the nodes close to BS die out quicker than rest of the nodes. The phenomenon known as "energy hole problem [3]" makes the network unusable in spite of majority of nodes being still live. Replacing batteries of dead nodes or energy scavenging is generally not possible in WSNs. It suggests that life of network can be prolonged by making the BS mobile. Mobile base station then can travel closer to the node to collect the data from the nodes resulting in energy savings and also avoiding creation of energy holes.

In this paper we present, analysis of optimal number of clusters in networks with mobile base station. In our model, the BS travels very close to the CH. This results in zero attenuation during data collection from CH by the BS. The theoretical results were verified using cluster evaluation techniques. Silhouette Coefficient, Cophenetic Correlation Coefficient, and Spearman Correlation Rank were computed to evaluate the formed clusters. Results establish the theoretical analysis.

Related work is presented in section 2. Mathematical formulation is presented in the next section. Results are presented in section 4. Paper is concluded in section 5.

2. Related Work

Clustering in WSNs was first proposed under the protocol LEACH [4]. In LEACH, number of clusters to be formed is determined a priori by assigning a predetermined value of probability of becoming a CH. Rumor [5], TEEN[6], PEGASIS [7] are other well known clustering algorithms.

———————————————

- *C.P. Gupta is currently working as associate professor at Rajasthan Technical University and pursuing his Ph.D. E-mail: guptacp2@rediffmail.com*
- *Dr. Arun Kumar is working as Lecturer in Mathematics at Govt. College, Kota, INDIA. He has authored several papers in the area of mathematical modeling of biological systems. E-mail: arunkr71@gmail.com.*

However, none of these deliberated upon finding the optimal number of clusters.

Bandyopadhyay et al. [8] investigated the optimal number of clusters in hierarchical WSNs. For a network with n nodes deployed as per a homogeneous spatial Poisson process of intensity λ in a square area of side 2a, using multihop transmission between CH and member nodes; optimal probability of becoming a CH is

$$p = \left[\frac{1}{3c} + \frac{\sqrt[3]{2}}{3c\left(2 + 27c^2 + 3\sqrt{3}c\sqrt{(27c^2+4)}\right)^{1/3}} + \frac{3c\left(2 + 27c^2 + 3\sqrt{3}c\sqrt{(27c^2+4)}\right)^{1/3}}{3c} \cdot \frac{1}{\sqrt[3]{2}} \right]$$

where, $c = 3.06a\sqrt{\lambda}$.

Ranjan et al. [9] computed the optimal number of clusters in random and grid networks considering single hop and multi hop communication between CH and BS.

FCA [10] uses residual energy parameter with distance to the base station metric of the sensor node to calculate competition radius. In every clustering round, each sensor node generates a random number between 0 and 1. If the random number for a particular node is smaller than the predefined threshold T which is the percentage of the desired tentative cluster-heads, then that sensor node becomes a tentative cluster head. The competition radius of each tentative cluster-head changes dynamically. FCA decreases the competition radius of each tentative cluster-head as the sensor node battery level decreases.

FCM [11] computes the optimal number of cluster heads. Xie and Beni's (XB) index was used to validate the partitioning. The total energy consumption in transmitting L bits by Ns number of sensors over a distance of d_{toCH} by the sensor to cluster head and then to the sink at distance dtoBS having a free space loss coefficient ε is

$$E_{Tot} = L(2N_s E_{elect} + N_s E_{DA} + \varepsilon_{fs}(kd_{toBS}^2 + N_s d_{toCH}^2))$$.

The optimal number of clusters K$_{opt}$ is

$$K_{opt} = \sqrt{\frac{N_s}{2\pi}} \frac{2}{0.765}$$,

and average distance to the base station

$$d_{toBS} = 0.765\frac{M}{2}$$

(M = side of deployment area).

If the distance of significant number of nodes is more than

$$d_o = \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}$$ then,

$$K_{opt} = \frac{\sqrt{N_s}}{\sqrt{2\pi}} \sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}} \frac{M}{d_{toBS}^2}$$

However, none of the above works consider the BS to be mobile.

3.   Energy Consumption Model

Energy available with the node is consumed in

(i)   Data Transmission from Cluster Members (CM) to their respective Cluster Head

(ii)   Data Reception at Cluster Heads

(iii)   Data Aggregation at Cluster Heads

(iv)   Data Transmission between Cluster Heads and Base Station

(v)   Computation of Total Energy Consumption

Assumptions:

- Each cluster member sends a data packet of constant length of L bits.

- The cluster head aggregates the data received from all the members with its own data into a frame of L bits.

- The control frames are small in comparison to data frames. Thus, energy consumed in transmitting control frames is negligible in comparison to energy consumed in transmitting data frames and hence, neglected.

Let

k= number of clusters

m_i= number of nodes in the ith cluster

n= Total number of nodes

$d_{i \rightarrow CH}$ = Distance between i[th] node and its cluster head.

$E_{ele}c$= Energy Consumed in transmitting/ Receiving one bit depends upon coding and modulation

ε_{fs}= Free Space Loss Coefficient

E_{fusion}=Energy consumed in data aggregation and computation per bit.

Average Number of nodes in each Cluster= $\frac{n}{k}$

3.1.1   *Energy consumed in transmission*

All the cluster members communicate directly with their respective cluster heads. Further, this communication takes

place only in the assigned time slot and thus no collision occurs. Hence, energy consumed in each cluster,

$$E_{Tx} = \sum_{i=1}^{m-1}(L * E_{elect} + L * \varepsilon_{fs} d_{i \to CH}^2) \qquad (1)$$

In the proposed protocol, the base station being mobile, visits the cluster heads to collect the aggregated data. Because of this, Energy is consumed only in transmitting L bits of information by k CHs to the base station. No energy loss happens and thus, energy consumption due to this is not considered. Hence, energy consumed in transmitting L bits by k cluster heads is given by

$$E_{Tx-CH} = \sum_{j=1}^{k} L * E_{elec} \qquad (2)$$

Total energy consumed in k clusters

$$E_{Tx-Total} = \begin{bmatrix} \sum_{j=1}^{k}\sum_{i=1}^{m_j-1}(L*E_{elect}+L*\varepsilon_{fs}*d_{i \to CH}^2) \\ + \sum_{j=1}^{k} L*E_{elec} \end{bmatrix} \qquad (3)$$

$$= (n * L * E_{elec}) + \sum_{j=1}^{k}\sum_{i=1}^{m_j-1}(L*\varepsilon_{fs}*d_{i \to CH}^2)$$

3.1.2 Energy Consumed in Receiving at k cluster heads

Each cluster head receives data from all the cluster members. Thus, ith cluster with mi number of members receives data from m$_i$-1 members. The energy consumed in receiving at each CH is

$$E_{Rx} = \sum_{i=1}^{m_i-1} LxE_{elec}$$

Thus, the total energy consumed in receiving at k cluster heads

$$E_{Rx-Total} = \sum_{j=1}^{k}\sum_{i=1}^{m_{ij}-1} L * E_{elec} \qquad (4)$$

3.1.3 Energy Consumed in data aggregation and computation at the cluster head

The CH aggregates data from all the members and also its own data. Thus the total energy consumed in aggregating at k cluster heads

$$E_{com} = E_{fusion} * L * n \qquad (5)$$

Thus, total energy consumed in each round

$$E_{Total} = E_{Tx-Total} + E_{Rx-Total} + E_{com}$$
$$=$$

$$\begin{bmatrix} (n*L*E_{elec}) + \sum_{j=1}^{k}\sum_{i=1}^{m_j-1}(L*\varepsilon_{fs}*d_{i \to CH}^2) \\ + \sum_{j=1}^{k}\sum_{i=1}^{m_{ij}-1} L*E_{elec} + (E_{fusion}*L*n) \end{bmatrix} \quad (6)$$

3.2   Optimal Number of Clusters

Let the area observed field is A and the cluster forms a circle with cluster head as center, area of each cluster can be approximated to *A/k*, one may argue that each cluster can be approximated by a circle of radius $\sqrt{\dfrac{A}{\pi k_{opt}}}$, where the cluster head resides at the center.

Since the expected squared distance between a random point in a circle of radius S and its center is $\dfrac{S^2}{2}$, the value of $d_{i \to CH}^2$ can be determined with respect to A and k. This can be obtained straightforwardly by drawing a hypothetical circle around the center (cluster head) and equating the area of this circle to half of the area of the original circle (area of the cluster). Solving the resulting equation for squared radius of the hypothetical circle, which is in fact equal to $d_{i \to CH}^2$, yields:

$$E[d_{i \to CH}^2] = \frac{A}{2\pi k} \qquad (7)$$

Further, assuming that energy required in fusing one bit is equal to the energy required for transmission/receiving one bit i.e.

$$E_{fusion} = E_{elec} \qquad (8)$$

Simplifying equation (6) and substituting the equation (7) and (8) in (6), we get

$$E_{Total} = (3n-k) * L * E_{elec} + \sum_{j=1}^{k}(m_j-1) * L * \varepsilon_{fs} * \frac{A}{2\pi k} \quad (9)$$

$$= (3n-k) * L * E_{elec} + \frac{(n-k)}{2\pi k} * L * A * \varepsilon_{fs}$$

To minimize the total energy consumption, the optimal value of clusters works out to $k_{opt} = \sqrt{-\dfrac{nA\varepsilon_{fs}}{2\pi E_{elec}}}$.

The sign of the expression under square root is insignificant and thus is neglected. Hence, optimal number of clusters is obtained as

$$k_{opt} = \sqrt{\frac{nA\varepsilon_{fs}}{2\pi E_{elec}}} \qquad (10)$$

Bandyopadhyay et al. [8] found the number of clusters considering multihop communication within cluster and also between CH and BS. Anwar et al. [12] considered re-transmission of lost packets during communication between CH and BS. The above results are not suitable for our model due to following reasons:

Our network model assumes single hop communication between CM and CH.

In our network model, the sink is mobile and moves very close to the CH for PULLING the aggregated data from CH. Thus, path loss between CH and sink is insignificant. Further, due to proximity of sink with CH at the time of data collection, probability of error is negligible not requiring retransmission.

4.   Simulation Results

Simulations were carried out on a network of 100 nodes deployed in a square area of 100 x 100. 100 samples were generated and the network was clustered using Hierarchical Agglomerative Clustering [13-15]. The network was clustered using for techniques namely Single Linkage, Average Linkage, Complete Linkage and Ward's Linkage. Evaluation is based on computing three parameters namely (i) Silhouette Coefficient (SC) (ii) Cophenetic Correlation Coefficient (CPCC) and (iii) Spearman Correlation Rank (SCR). Results are shown in Fig. 1 to 3.
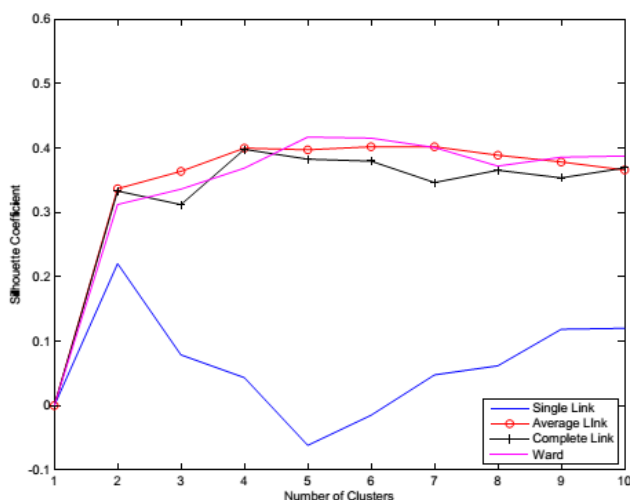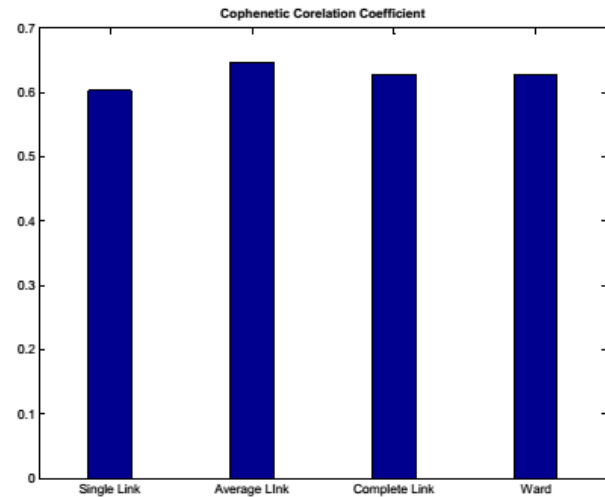

Fig 1: Average Silhouette Coefficient

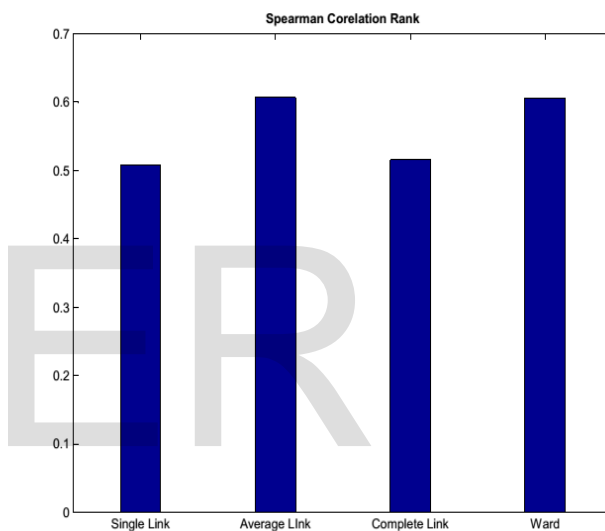Fig 2: Average Cophenetic Correlation Coefficient

Fig 3: Average Spearman Correlation Rank

5.   Conclusion

Theoretical value for optimal number of clusters as computed works out to be 6. Value of Silhouette Coefficient for average linking is the highest for this number of clusters. As per [13, 14] optimal number of clusters is represented by the highest value of the coefficient. Further, CPCC is used to evaluate the optimal linkage in the given data set. The higher value of coefficient suggests the linkage method for clustering the data. For randomly generated points in our simulation, for all the runs, value was the highest for average linking. The SCR was computed between Euclidean Distance and Cophenetic Distance. Cophenetic Distance is measured as the height of the dendrogaram at which the node was first merged into another cluster. Higher value point well clustering. Our results suggest that for randomly deployed WSNs, average linkage results in optimal clustering.

References

[1] Gupta, C P, Kumar Arun: "Wireless Sensor Networks: A Review", International Journal of Sensors, Wireless Communications and Control, 2012, Vol. 2, No. 3, Pg 1-12.

[2] Akyildiz, Ian F., et al. "Wireless sensor networks: a survey." Computer networks 38.4 (2002): 393-422.

[3] Ahmed, Nadeem, Salil S. Kanhere, and Sanjay Jha. "The holes problem in wireless sensor networks: a survey." ACM SIGMOBILE mobile Computing and Communications Review 9.2 (2005): 4-18.

[4] Heinzelman, Wendi Rabiner, Anantha Chandrakasan, and Hari Balakrishnan. "Energy-efficient communication protocol for wireless micro-sensor networks." System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on. IEEE, 2000.

[5] Braginsky, David, and Deborah Estrin. "Rumor routing algorithm for sensor networks." Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications. ACM, 2002.

[6] Manjeshwar, Arati, and Dharma P. Agrawal. "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks." 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing. Vol. 22. 2001.

[7] Lindsey, Stephanie, and Cauligi S. Raghavendra. "PEGASIS: Power-efficient gathering in sensor information systems." Aerospace conference proceedings, 2002. IEEE. Vol. 3. IEEE, 2002.

[8] Bandyopadhyay, Seema, and Edward J. Coyle. "An energy efficient hierarchical clustering algorithm for wireless sensor networks." INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies. Vol. 3. IEEE, 2003.

[9] Ranjan, Ravi, and Subrat Kar. "A novel approach for finding optimal number of cluster head in wireless sensor network." Communications (NCC), 2011 National Conference on. IEEE, 2011.

[10] Godbole, Vaibhav. "Performance Analysis of Clustering Protocol Using Fuzzy Logic For Wireless Sensor Network." IAES International Journal of Artificial Intelligence (IJ-AI) 1.3 (2012).

[11] Raghuvanshi, Ajay Singh, et al. "Optimal number of clusters in wireless sensor networks: a FCM approach." International Journal of Sensor Networks 12.1 (2012): 16-24.

[12] Sadat, Anwar, and Gour Karmakar. "Optimum Clusters for Reliable and Energy Efficient Wireless Sensor Networks." Network Computing and Applications (NCA), 2011 10th IEEE International Symposium on. IEEE, 2011.

[13] Tan, Pang-Ning, "Introduction to data mining". Chapter 8, Pearson Education India, 2007.

[14] Tryfos, Peter," Methods for Business Analysis and Forecasting: Text and Case", Chapter 15, John Wiley, ISBN: 0-471-12384-6, 1998.