ISSN XXXX XXXX © 2017 IJESC

Research Article Volume 7 Issue No.5

# Aggregation of Data Based on Clusters: A Bayesian Approach

Rachana.D.N
M. Tech
Department of Computer Science & Engineering
UBDTCE, Davanagere, India

**Abstract:**
Wireless sensor networks are increasingly used in many real-time applications now-a-days for collection of data from remote areas. The efficient approach for aggregation of data is of at most concern as the sensor nodes are of limited energy and computational power. In this paper we discuss the cluster based technique along with the Bayesian network approach to overcome this problem. The proposed system adaptively uses a clustering technique. The sensed data from every sensor node is forwarded to the cluster heads which in turn forwards the collected data to the base station. The cluster heads forward the collected data to the base station by an optimal path which is chosen probabilistically by the Bayesian approach.

**Keywords:** Wireless sensor network, Bayesian Approach.

## I. INTRODUCTION

A Wireless Sensor Network (WSN) can be defined as a set of spatially scattered intelligent sensors designed to obtain measurements from the environment, abstract relevant information from the data gathered, and to derive appropriate inferences from the information gained. Wireless sensor networks depend on multiple processors to simultaneously gather and process information from many sources. Currently, there has been an increasing interest in the development of WSNs for the process of information gathering. Availability of new technology makes these networks economically feasible. The increased complexity of today's information gathering tasks has created a demand for such networks. These tasks are usually time -critical and rely on the reliable delivery of accurate information.

Thus, the search for efficient, fault -tolerant architectures for WSNs has become an important research area in computer science. All sensor nodes have limited power supply and have the capabilities of information sensing, data processing and wireless communication. WSN has various characteristics like Ad Hoc deployment, Dynamic network topology, Energy Constrained operation, Shared bandwidth, large scale of deployment. Despite of these characteristics routing in WSN is more challenging. Firstly, resources are greatly constrained in terms of power supply, processing capability and transmission bandwidth. Secondly, it is difficult to design a global addressing scheme as Internet Protocol (IP). Furthermore, IP cannot be applied to WSNs, since address updating in a large-scale or dynamic WSN can result in heavy overhead. Thirdly, due to the limited resources, it is hard for routing to cope with unpredictable and frequent topology changes, especially in a mobile environment. Fourthly, data collection by many sensor nodes usually results in a high probability of data redundancy, which must be considered by routing protocols. Fifthly, most applications of WSNs require the only communication scheme of many-to-one, i.e., from multiple sources to one particular sink, rather than multicast or peer to peer. Selecting the optimum sensors and wireless communications link requires knowledge of the application and problem definition.

### 1.I COMMUNICATION IN WSN

In a typical WSN, each node needs to fuse the local information with the data collected by the other nodes so that an updated assessment is obtained. Current research involves fusion based on a multiple hypothesis approach. Maintaining consistency and eliminating redundancy are two important considerations. The problem of determining what should be communicated is more important than how this communication is to be effected. An analysis of this problem yields the following classes of information as likely candidates for being communicated: information about the WSN, information about the state of the world, hypothesis, conjectures and special requests for specific actions. It is easy to see that different classes of information warrant different degrees of reliability and urgency. Fig 1. shows the communication occuring in WSN.

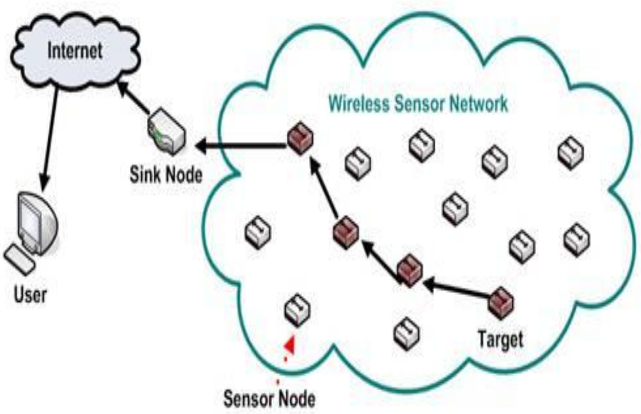


**Figure.1. Communication in WSN**

### 1.2 DATA AGGREGATION

A data aggregation function is taken that aggregates the individual sensor readings. The Bayesian approach is used for the aggregation of the sensor readings .For computation of the aggregate functions, the following requirements are to be satisfied: (i) Every node should know all the other nodes which are at one hope distance from it, (ii) the number of messages transmitted within the WSN for the purpose of data aggregation should be kept at a minimum, and (iii) the aggregation result should be as accurate as possible. Data aggregation in intermediate nodes (called aggregator nodes) is an effective approach for optimizing consumption of scarce resources like bandwidth and energy in Wireless Sensor Networks (WSNs).

### I.3 BAYESIAN APPROACH

The Bayesian classifcation algorithm is the good theory for Bayesian approach. It is a statistical classification algorithm. This algorithm efficiently helps in solving the problem of large scale distributed Wireless Sensor Networks which are deployed randomnly and uniformly. The nodes sense the data and the data must be transmitted from the sensed nodes to the sink. For this the routing must be done in a effective manner. The quality of the routing path is evaluated by this approach. Energy consumption is maintained in a effective manner. Once the nodes sense the data, the data is transmitted to the sink node taking different paths which is based on probability. Fig 2 shows the cluster formation in environment based on Bayesian approach.

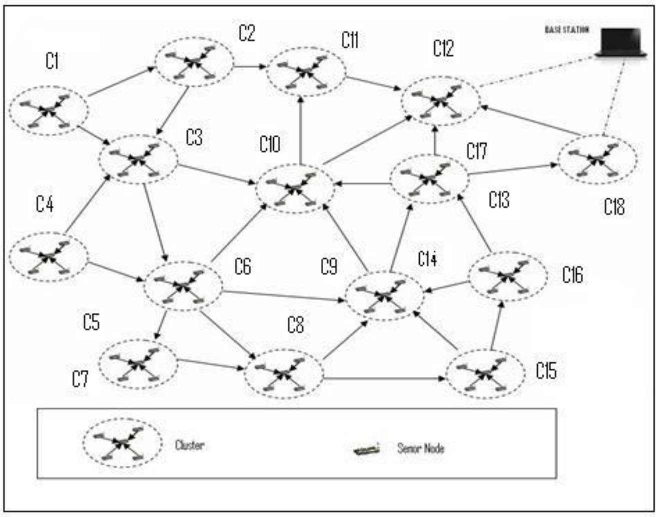


**Figure. 2. Cluster formation in environment based Bayesian Approach**

## II. LITERATURE SURVEY

[1] Discusses the energy hole problem which is caused by turning off the sensor node which is near to sink node which includes in frequent transmission particularly in MAC based WSN. Author proposed an Energy-Aware Hybrid Data Aggregation Mechanism (EHDAM) which controls data transmission with burst length by adjusting threshold value in reciprocal proportion to the remaining energy state of the node. This mechanism increases the lifetime of the nodes that are affected by energy whole problem.

[2] Author proposed two aggregation schemes called single-hop-length (SHL) and Multiple-hop-length(MHL) where the WSN are randomly deployed with some node density. It mainly proposes on the study of tradeoffs between aggregation throughput and gathering efficiency. Author showed that a set of symmetric function called Divisible Perfectly Compressible (DPC), mean, max and various kinds of indicator functions, the data can be aggregated to the sink the throughput of a constant order. The author has proved the MHL is scalable.

[3] States that data aggregation significantly reduces the communication overhead and energy consumption. The author proposed a loss-resilient aggregation framework called synopsis diffusion, which uses duplicate insensitive algorithms on top of multipath routing to calculate the aggregates. By doing this there comes a problem of false sub aggregate values to contributed by compromised nodes. This may cause a large errors in aggregate computed at sink node. The author has proved through attack-resilient computation algorithm which guarantees the successful computation of aggregate at sink node.

[4] Proposes a theory due to limited computational power and energy resources, data aggregation is done only on the basis of simple methods like averaging. If a node is compromised then it is vulnerable. Thus trustworthiness of data is crucial part in collecting the data. Iterative filtering algorithms can be trusted for this purpose. These algorithms simultaneously aggregate data from multiple sources and provide trust assessment of these sources, usually in a form of corresponding weight factors assigned to data provided by each source. The author discussed several algorithms that are existing and proposed an improvement for iterative filtering technique by providing initial approximation and making them collision robust, accurate and faster converging.

[5] Addressed the problem of localizing an unknown number of energy emitting nodes. They came across Bayesian solution for joint estimation of unknown sources as well as their parameters based on SMC sampler. Author derived posterior Cramer-Rao bound which helped in estimating the characteristics of these multiple energy emitting resources. Author derived the equation which can be used as a criterion to minimize in order to design efficiently the network parameters.

[6] Discusses the application over the principal component analysis(PCA) as one of the solution to data aggregation in power and computational limited environments. The author proposed a method known as Jacobi Eigen Value algorithm. Signaling has been reduced by means of exchanging the parameters in order to keep projection basis synchronized. The algorithm is able to sustain the number of projections by a defined threshold value. Author proved that with the proper adjustments of Jacobi updates, there is a minimum loss of energy consumption.

## III. METHODOLOGY

The proposed system involves data aggregation by cluster formation. This is done using the Bayesian approach. The components of the system are sensor nodes, base station or sink, clusters, cluster heads.

i. Sensor nodes: The sensor nodes sense the environment for data.

ii. Base Station: The sensor nodes send the sensed data to the base station. The base station sends all the data to the user.

iii. Clusters: The sensor nodes group among themselves to form clusters. Collection of data is done among the cluster.

iv. Cluster Head: The collected data in the cluster is managed by the cluster head. Fig 3. shows the system design of the proposed system.

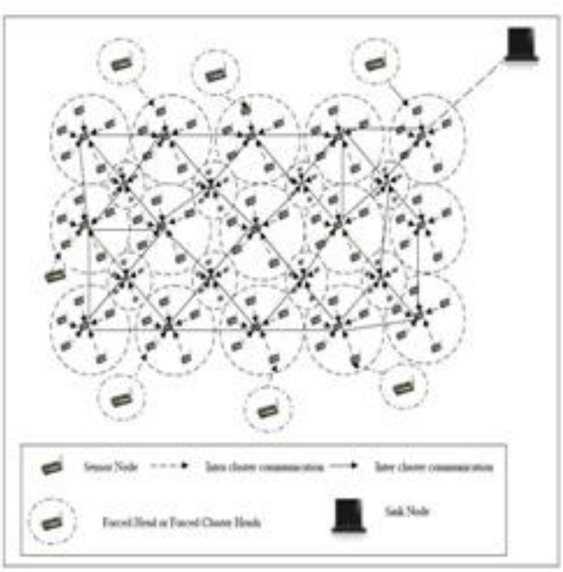

**Figure. 3. System design**

Each node resides in a cluster. The sensor nodes communicate with the cluster head only. The cluster head collect the data sensed by all the sensor nodes in its cluster. Then the cluster head forwards the collected data to the base station. One cluster head communicates with other cluster heads during data transmission. There occurs communication between cluster heads. The base station communicates only with the cluster heads of all the clusters. Hence the energy consumed in this process is comparatively less. Fig 4. Shows the general architecture of the proposed system.

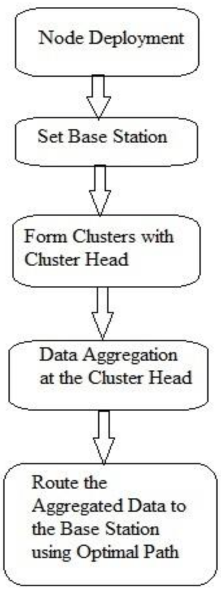


**Figure.4. General Architecture of the Proposed System**

**The operation of the proposed system is as follows:**

1. Nodes are deployed randomly in the environment.
2. Set the base station in the environment.
3. Create the cluster among the sensor nodes.
4. Assigning a cluster head for each cluster.
5. Data aggregation at cluster heads.
6. Data transmission from cluster heads to base station probabilistically an optimal path using the Bayesian approach.
7. Acknowledgement from the base station.
8. Further action is performed by the base station after getting the sensed data.

## 3.I FUNCTION SCHEME

### 3.2 Node Deployment

The sensor nodes are deployed randomly and uniformly in the large scale environment. They may be static or moving very slowly. There is a bidirectional link between the cluster heads and the base station or sink. Every node will know other nodes which are present at one-hop distance.

### 3.3 Initialization

After the node deployment is done, the base station sends the broadcast message to all the nodes. The message consists of two parts namely the base station ID and the hop count from the base station. Fig.5 shows the initialization message for the initialization phase.

| | |
|---|---|
| Base station ID | Hop count |

**Figure. 5. Initialization message**

Every node gets the information about the base station and at what distance each node is from the base station. Every node updates its routing table after they get the message from the base station.

### 3.4. Cluster Formation

During this phase, the query server Q triggers a HELLO message. As shown in fig.6, the sensor nodes elect a cluster based on pre-defined probability when they receive the HELLO message. When a node becomes a cluster head, it forwards the HELLO message to its neighbor. When the node gets the HELLO message, they tend to join the cluster by sending the broadcast JOIN message back to the cluster head. This process is repeated in the WSN environment until all the sensor nodes join different clusters. Every cluster will have its individual cluster heads. The number of nodes differs from cluster to cluster.

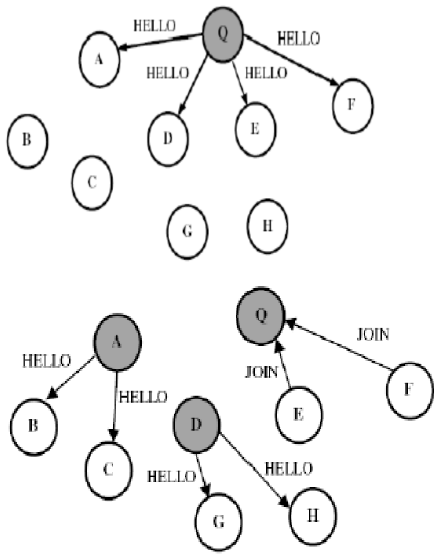


**Figure. 6. Process of cluster formation in the proposed system**

### 3.5. Data Aggregation and Transmission

The sensor nodes sense the data from the environment. A certain amount of energy is consumed by the nodes for sensing the data. The sensed data is forwarded to the cluster head in which the node resides. A certain amount of energy is consumed for transmission of the data. All the nodes send the data to the cluster head in the cluster. The cluster head does the function of data aggregation. After the data is aggregated at the cluster head, it has to be forwarded to the base station. For forwarding the data from the cluster head to the base station an optimal path must be used. If the cluster head is near to the base station the data is directly sent to the base station else the cluster head sends the data to the nearest cluster head and then the data is forwarded to the base station. After the base station receives the data, it sends the data to the user. The base station also sends the acknowledgment back to the cluster head about the status of the data.

### 3.6. Route Maintenance

During the transmission of data there is consumption of energy by the sensor nodes. When the energy of the sensor nodes depletes, the sensor nodes broadcast the ALARM message in the environment. When the cluster head gets the message, the sensor node is removed from the cluster and the routing table is updated.

## IV. RESULTS

Fig 7 shows the energy level comparison by the proposed method versus the other algorithms. As we can see the proposed Bayesian network approach is more efficient than the other two algorithms.

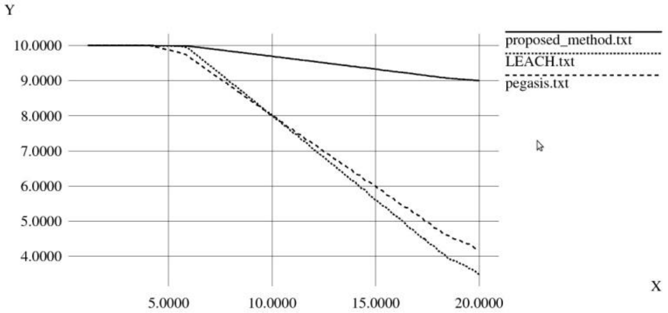


**Figure.7. Energy level comparison of our proposed method vs different algorithms.**

## V. CONCLUSION

The proposed method is energy efficient and responsive to network. The proposed system selects efficient Cluster Heads, optimization of routing and increase the network lifetime. It has been observed that Bayesian Network is used to probabilistically select Cluster Heads in achieving better results as compare with LEACH and PEGASIS approach. As compared to LEACH and PEGASIS protocol the proposed model performed better performance in the network. A simulation results obtained from a proposed algorithms are more efficient than LEACH and PEGASIS algorithm.

## VI. REFERENCES


[1]. M. G. Kim, Y. T. Han and H. S. Park, "Energy-Aware Hybrid Data Aggregation Mechanism Considering the Energy Hole Problem in Asynchronous MAC-Based WSNs," in IEEE Communications Letters, vol. 15, no. 11, pp. 1169-1171, November 2011. doi: 10.1109/LCOMM.2011.092911.110792 URL:http://ieeexplore.ieee.org/stamp/ stamp.jsp?tp= &arnu mbe r= 6034398&isnumber=6083590

[2]. C. Wang, C. Jiang, S. Tang and X. Y. Li, "SelectCast: Scalable Data Aggregation Scheme in Wireless Sensor Networks," in IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1958-1969, Oct. 2012. doi: 10.1109/TPDS.2011.312 URL: http://ieeexplore.ieee. org/ stamp/stamp.jsp?tp=&arnumber=6122022&isnumber=6285945

[3]. S. Roy, M. Conti, S. Setia and S. Jajodia, "Secure Data Aggregation in Wireless Sensor Networks: Filtering out the Attacker's Impact," in IEEE Transactions on Information Forensics and Security, vol. 9, no. 4, pp. 681-694, April 2014. doi: 10.1109/TIFS.2014.2307197 URL: http://ieeexplore.ieee. org/stamp/stamp.jsp?tp=&arnumber=6746043&isnumber=6755552

[4]. M. Rezvani, A. Ignjatovic, E. Bertino and S. Jha, "Secure Data Aggregation Technique for Wireless Sensor Networks in the Presence of Collusion Attacks," in IEEE Transactions on Dependable and Secure Computing, vol. 12, no. 1, pp. 98-110, Jan.-Feb. 1 2015. doi: 10.1109/TDSC.2014.2316816 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6786996&isnumber=7008601

[5]. T. L. T. Nguyen, F. Septier, H. Rajaona, G. W. Peters, I. Nevat and Y. Delignon, "A Bayesian Perspective on Multiple Source Localization in Wireless Sensor Networks," in IEEE Transactions on Signal Processing, vol. 64, no. 7, pp. 1684-1699, April1, 2016. doi: 10.1109/TSP.2015.2505689 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp= &arnumber=7347468&isnumber=7410177

[6]. A. Morell, A. Correa, M. Barceló and J. L. Vicario, "Data Aggregation and Principal Component Analysis in WSNs," in IEEE Transactions on Wireless Communications, vol. 15, no. 6, pp. 3908-3919, June 2016. doi: 10.1109/ TWC. 2016.2531041 URL: http://ieeexplore.ieee.org/stamp/stamp .jsp?tp=&arnumber=7410064&isnumber=7485904