

Optimization of sensor field data using a Hybrid Approach

Sreeram Indraneel¹ and Kurra Raja Sekhara Rao²

1. Associate Professor, CSE, St.Anns Engineering College, Chirala, A.P., India.

2. Professor, CSE, K.L. University, Guntur-522502, India.

Abstract: With the vast utilization of wireless sensor networks for various applications including huge data generating applications, there is a need to concentrate to optimizing the sensor field data. In this paper, we proposed a Hybrid approach for optimizing sensor field data. We have assumed a T-nary hierarchical wireless sensor network constituting nodes, cluster Heads & Gateways. In our approach we have applied Fuzzy based Map Reduce technique to the sensor field data at various levels of network hierarchy to reduce the redundant data transmissions which in turn optimizes the power consumption of the nodes in the network.

Key Words: Fuzzy, Hybrid, Map Reduce, Optimization, Wireless sensor Networks etc.

Introduction: With the recent technological advancements including miniaturization of electronics, communication devices, the scope of application of wireless sensor networks in diversified fields is increased. In some applications such as environmental monitoring, battle field surveillance, Geographical information systems etc huge amount of data is generated. Till today, lots of researchers have proposed various data aggregation techniques[1], clustering algorithms[2], to optimize the power consumption by decreasing the amount of data transmission in the network. We have proposed a Hybrid approach to further optimize the data transmissions by reducing the redundant data at various levels of the network hierarchy. For this we have used Fuzzy based Map Reduce technique.

The rest of this paper is organized as follows. Section 2 deals with Fuzzy Logic. Section 3 discuss about Map Reduce technique. Section 4 deals with hierarchical organization of T-nary wireless sensor network. Section 5 Describes the Hybrid optimization procedure using Fuzzy Logic and Map Reduce techniques. Section 6 demonstrates our approach on the assumed sample sensor data using statistical approach. Conclusion and future perspective in section 7.

2. Fuzzy Logic:

The traditional binary logic consists of variables that may take values such as **true** or **false**. But Fuzzy logic is a form of many-valued logic or probabilistic logic. It deals with approximate reasoning instead fixed and exact values. Fuzzy logic variables may have a truth value that ranges in degree between 0 and 1 .

A fuzzy subset F of a set X is a function $F: X \rightarrow L$, where L is the interval $[0, 1]$. This function is also called a membership function. A membership function is a generalization of a characteristic

function or an indicator function of a subset defined for $L = \{0, 1\}$. More generally, one can use a complete lattice L in a definition of a fuzzy subset F .^[2]

$$\int_0^t L(t)dt = 0 \leq L(t) \leq 1$$

3. Map Reduce Routine:

Map Reduce is a programming model and an associated implementation for processing and generating large data sets. Users specify a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key.

This technique helps in parallelized execution on a large cluster of machines. The run-time system takes care of the details of partitioning the input data, scheduling the program's execution across a set of machines, handling machine failures, and managing the required inter-machine communication. Our implementation of Map Reduce runs on a large cluster of member nodes and is highly scalable. A typical Map Reduce computation processes huge data thousands of machines.

The computation takes a set of *input* key/value pairs, and produces a set of *output* key/value pairs. The user of the Map Reduce library expresses the computation as two functions: *Map* and *Reduce*. *Map*, written by the user, takes an input pair and produces a set of *intermediate* key/value pairs. The Map Reduce library groups together all intermediate values associated with the same intermediate key I and passes them to the *Reduce* function.

The *Reduce* function, also written by the user, accepts an intermediate key I and a set of values for that key. It merges together these values to form a possibly smaller set of values. Typically just zero or one output value is produced per *Reduce* invocation. The intermediate values are supplied to the user's reduce function via an iterator. This allows us to handle lists of values that are too large to fit in memory.

4. T-nary wireless sensor network:

We have considered a three level hierarchical network to implement our proposed approach. At the bottom level the sensor nodes will sense the specified parameters in the sensor field application area and generate the corresponding data sets. The neighboring sensor nodes are clustered based on their distance and formed as a cluster with a cluster Head. At the third level of the hierarchy the cluster heads forward the data to the Gateways. The Gateways will transmit the data from sensing field to the remote base station or the end user.

5. Hybrid optimization procedure:

To optimize the data transmissions in the network lot of work was done on data aggregation, clustering of nodes etc. We have followed a novel approach to further optimize the data

transmissions by implementing the Map Reduce technique at the cluster head. The cluster Head specifies a map function that processes a key/value pair to generate a set of intermediate key/value pairs, and a reduce function that merges all intermediate values associated with the same intermediate key. The query related to the sensor field data is parallel processed on all the cluster member nodes. The reduce function uses the outlier evaluation model to reduce the unwanted data before forwarding it to the cluster head. This model uses the parameter evaluation to read the required parameters from the sensor field data and to ignore the rest of huge unwanted data.

The data from the cluster heads is transmitted to the neighboring Gateway node. To optimize the quality of the data pooled at the Gateway and to avoid loss of data due to collision, the time synchronization is achieved between the cluster heads of the Gateway node using Fuzzy logic.

A subset of cluster heads F of a set X is a function $F: X \rightarrow L$, where L is the interval $[0, 1]$. This function is also called a membership function. A membership function is a generalization of a characteristic function or an indicator function of a subset of sensor nodes defined for $L = \{0, 1\}$. All the cluster heads will complete their transmission cyclically within the interval $[0, 1]$. After this the Gate way node will eliminate the duplicate data and forward this to the base station.

6. Experimental Results:

We have synthesized a sample data for a network with 100 nodes with each node generating similar kind of data at different data rates. Based on the geographical distance we have formed 10 clusters out of 100 nodes with each cluster having random no of nodes. We have considered 4 gateways with each gate way connected to random no of cluster heads. By our approach we have obtained the percentage of optimization as follows

S.No	Amount of data at Gateways per m.sec (KB)	
	Before optimization	After optimization
1.	25	20
2.	30	27
3.	32	25
4.	40	25
5.	55	42
6.	100	85
7.	45	40
8.	52	43
9.	38	32
10.	46	36

Table: 6.1: Synthesized Data Rate at Gate Ways

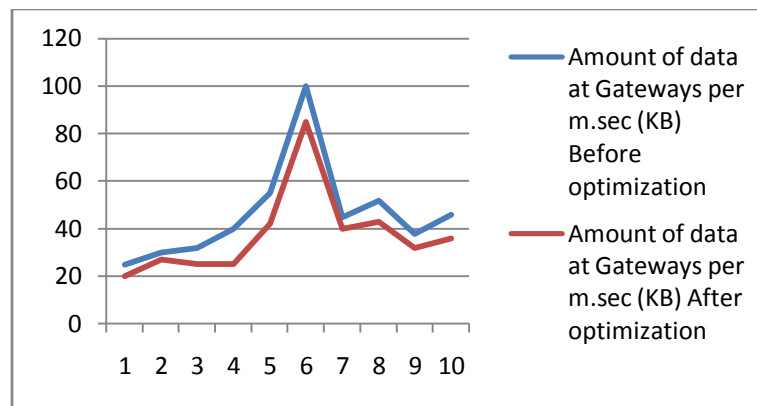


Fig.6.2: Synthesized Data Rate at Gate Ways

The reliability of this technique can be evaluated using the calculation of probability of failure of the generated data samples using the following observation.

$$Probability = \frac{\text{Number of failing samples}}{\text{Total number of samples under consideration}}$$

6. Conclusion and Future Work:

Our study concentrated on handling the huge data generated in sensor field area. In general by its very nature the sensor field data consists of lot of redundancy. Our proposed hybrid approach optimizes the amount of data transmitted in between nodes, cluster heads, Gateways. It decreases the amount of data pooled at Gateways. By using Map Reduce technique we can decentralize the processing of the query related to sensor field data at cluster member nodes. In future we concentrate on further optimizing the amount of data transmitted between the nodes.

References:

- [1] K. Dasgupta, K. Kalpakis, and P. Namjoshi. An efficient clustering-based heuristic for data gathering and aggregation in sensor networks. In in Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC, pages 1948–1953, 2003.
- [2] Vivek Katiyar, Narottam Chand, Surender Soni A Survey on Clustering Algorithms for Heterogeneous Wireless Sensor Network Int. J. Advanced Networking and Applications Volume: 02, Issue: 04, Pages: 745-754 (2011).
- [3] Goguen, J. (1967) "L-fuzzy sets", *J. Math. Anal. Appl.*, 18, 145-174.
- [4] Zadeh, L. A. (1965) "Fuzzy sets", *Information and Control*, 8, 338–35.
- [5] J.Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan. Building efficient wireless sensor networks with low-level naming, 2001.
- [6] H. Luo, Y. Liu, and S. K. Das. Distributed algorithm for en route aggregation decision in wireless sensor networks. *IEEE Transactions on Mobile Computing*, 8, January 2009.
- [7] H. Morcos, G. Atia, A. Bestavros, and A. Matta. An Information Theoretic Framework for Field Monitoring Using Autonomously Mobile Sensors. In Proceedings of DCOSS'08: The 4th IEEE/ACM International Conference on Distributed Computing in Sensor Systems, Santorini, Greece, June 2008.
- [8] Christine Jardak, Janne Riihijärvi, Frank Oldewurtel, and Petri Mähönen. Parallel Processing of Data from Very Large-Scale Wireless Sensor Networks
- [9] Azer Bestavros, Dóra Erdős, Vatche Ishakian, Andrei Lapets, Evimaria Terzi The Filter-Placement Problem and its Application to Content De-Duplication Computer Science Department, Boston University Boston, MA 02215, USA.
- [10] Jeffrey Dean and Sanjay Ghemawat, Map Reduce: Simplified Data Processing on Large Clusters.