Energy Efficient Wireless Sensor Network using Genetic Algorithm based Association Rules

T. Abirami, Ph. D
Assistant Professor (SRG),
Department of Information Technology
Kongu Engineering College
Perundurai, Erode, Tamilnadu, India

P. Priakanth, Ph. D
Professor,
Department of Computer Application
Kongu Engineering College
Perundurai, Erode, Tamilnadu, India

ABSTRACT
Wireless Sensor Networks (WSN) usually contains thousands or hundreds of sensors which are randomly deployed. Sensors are powered by battery, which is an important issue in sensor networks, since routing consumes a lot of energy. Such nodes are deployed in thousands to form a network with capacity to report to a data collection sink (base station). An efficient routing scheme in sensor network is also important. Networking unattended sensor nodes are expected to have significant impact on the efficiency of many military and civil applications such as combat field surveillance, security and disaster management. Genetic algorithm (GA) based data aggregation trees are used where the sensors receive data from neighboring nodes, aggregate the incoming data packets, and forward the aggregated data to a suitable neighbor. GA is used to create energy efficient data aggregation trees. In this work, the amount of data sent to sink is reduced using association rule mining and in turn to further reduce the energy consumption of the network; optimal routes are chosen to transmit data to the sink based on energy consumption. The proposed method is able to discover the association rules to make predictive analysis on node failure, asymmetric links. The rules found form the basis for coding solutions in the proposed genetic algorithm. GA is applied to generate balanced and energy efficient data aggregation spanning trees for wireless sensor networks. E-Span, which is an energy-aware spanning tree algorithm and Lifetime-Preserving Tree (LPT) are used to create data aggregation trees. The proposed GA extends network lifetime.

Keywords
Wireless sensor networks, genetic algorithm, energy efficient, data aggregation Trees, Association Rules

1. INTRODUCTION
The challenge of improving sensor energy consumption is a key issue in wireless sensor networks (WSNs). This is due to the fact that sensor network lifetime is directly related to operational lifetime. If sensors can operate for longer periods of time and the frequency of node failures can be reduced, the reliability and adaptability of the sensor network will consequently improve. Data transmission is very costly in terms of energy. Sensor which collect data hand them over to the sink which is followed by offline data analyses to extract patterns. The existence of a large communication overhead affects sensor network performance negatively. This large overhead becomes a hurdle for the deployment of long term large scale sensor networks. Association mining is used to discover frequent patterns in the data. As the association mining is applied in-network, Patterns and not the raw data streams are forwarded to the sink when association mining is applied to the network which thereby reduces communication overhead significantly.

Sensor nodes that are used to form a sensor network are normally operated by a small battery which has small amount of energy. Therefore, in wireless sensor networks reducing energy consumption of each sensor node is one of the prominent issues to address in the network lifetime, since wireless communications consume significant amount of battery power, sensor nodes should be energy efficient in transmitting data. Protocols can reduce transmitted power in two ways. First where nodes can emit to short distances such as data sinks or cluster nodes. The cluster node can then send the data over a larger distance preserving the power of the smaller nodes. The second is by reducing the number of bits (amount of data) sent across the wireless network.

In this work, the amount of data sent to sink is decreased using association rule mining and to further reduce the energy consumption of the network, optimal routes are chosen to transmit data to the sink based on energy consumption.

A genetic algorithm (GA) is implemented to generate balanced and energy efficient data aggregation spanning trees for wireless sensor networks. In a data gathering round, a single best tree consumes lowest energy from all nodes but assigns more load to some sensors. As a result, the energy resources of heavily loaded nodes will be depleted earlier than others. The proposed GA extends network lifetime.

2. MATERIAL AND METHODS
2.1 Dataset
To evaluate the proposed methods, the intel lab sensor dataset is used [5]. This dataset contains data collected from 54 sensors installed in the Intel Berkeley Research lab. The sensors installed were Mica2Dot sensors with weather boards. The data collected were timestamped. The data collected consists of topology information, humidity, temperature, light and voltage value. The data was collected once every 31 seconds. The dataset includes a log of about 2.3 million readings from the sensors. The data is represented as shown in Fig 1.

Date	Time	Epoch	Moteid	Temp	Humidity	Light
Yyyy-mm-dd	hh:mm:ss.	int	int	real	real	real

Fig 1: The data captured in the sink

In this paper thirty minutes of data collected between first march 2004, 9:00 AM to first March 2004, 9:30AM consisting of over 9000 data messages received in the sink was studied to find the association between the sensor motes

2.2 Association Rules

Association rules are if/then statements that uncover relationships between apparently unrelated data in a relational database/information repository. An example would be "If a customer purchases a dozen eggs, he is also 80% likely to buy milk." An association rule contains two parts, an antecedent (if) and consequent (then). An antecedent is an item within the data and a consequent is that is found in combination with an antecedent [6].

Association rules are created by data analysis for regular if/then patterns and use criteria *support* and *confidence* to identify important relationships. *Support* indicates how frequently the item appears in a database whereas *Confidence* indicates the number of times if/then statements were seen to be true. Association rules extracts interesting correlations, frequent patterns, associations or casual structures among item sets in transaction databases or data repositories [7]. Association rules are used in varied areas including telecommunication networks, market and risk management, inventory control etc. Association rule mining is to locate association rules which satisfy predefined minimum support and confidence from a database. Numbers of algorithms are proposed for discovering association rules in literature [8, 9].

2.3 Rule Extraction

The user defines *maxscope* which is the upper limits of distance in which an event occurs and *maxhistory* which is a time frame. The sensors collect event notifications occurring within *maxscope* and keep a history with size *maxhistory* of these events. Association mining rule is applied on the data collected to discover patterns. Every node mines patterns in the following form:

$$I_1 \wedge \ldots \wedge I_m \Rightarrow E[S,C]$$

where event E occurs at node n with support S and confidence C given that antecedents A_i holds true. Antecedents for a dataset D, set of transaction, are in the form of

$$I_i = (E_i, D_i, T_i, N_i)$$

Every node sends a subset of discovered patterns to the sink, thus, reducing the communication overhead.

2.4 Genetic Algorithms for Mining Association Rules

Genetic Algorithms are chaotic and non-determinist search methods, which use real life models to solve complex and at times intractable problems [20]. It functionality is based on the Darwin’s theory of biological evolution through natural and sexual selection. In fact, these feasible solutions are chromosomes. The reasonable solutions are selected after the population is studied in each iteration under genetic operators. But there are some problems of the genetic algorithm, such as the slow convergence rate, premature convergence and occasionally local optimal solution.

The value of evaluating the path is taken as the fitness value in GA, which contains all constraint conditions in searching optimum path. The factor such as path length, energy consumption and the network energy equilibrium should be considered for feasible path fitness value. The feasible path fitness function of n nodes is defined as follows:

$$f(X) = w_d d(X) + w_e e(X) + w_l l(X)$$

w_d, w_e and w_l are respectively length, energy consumption and the network energy equilibrium. d(X) is the path length, e(X) is energy consumption of path(20), l(X) is consumption of network energy equilibrium, the total length of path is as follows:

$$d(X) = \sum_{i=1}^{n-1} d^2(m_i, m_{i+1})$$

$d(m_i, m_{i+1})$ is the distance between the node m_i and m_{i+1}. Network energy equilibrium consumption is as follows: A chromosome shows a path of transmitting data, which is divided into feasible path and unfeasible path. Different chromosome consists of different sensor node sequence, so the length of chromosome is variable and it can be showed by two-dimensional array or chain structure.

$$l(X) = \sum_{i-1}^{n-1} \alpha(l_i, l_{i+1})$$

Selection, crossover and mutation

According to the practice need of WSN path optimization, the three operators were used. And the fitness value of the optimum path is less than the second one. At first, the number of best individuals is retained before crossover. They chose best individual transitioned directly to offspring group and the selection on proportion method with other individual. Two individuals (paths) p1, p2 are chosen at random from the group to exchange the path in the path p1, p2 from the start node to target node. Pm is mutation probability; one individual (path) Pi is chosen to generate a probability of random p.

3. RESULTS

The proposed GA is simulated for performance evaluation. The associations are identified using the proposed genetic association rule mining. Table 3.1 shows the associations found.

Table 3.1 Association identified using Fuzzy-Genetic Association Rule Mining

m3	m6	==>	m2	
m4		==>	m3	
m2	m4	==>	m3	m6
m2	m11	==>	m9	
m4	m6	==>	m2	m3
m6	m9	==>	m16	
m8	m9	==>	m1	m11
m8	m11	==>	m1	m9
m10	m11	==>	m7	
m2	m6	==>	m3	

The proposed method is able to discover the associations to make predictive analysis such as node failure, asymmetric links. The rules found forms the basis for coding solutions in the proposed Genetic Algorithm. Intel lab sensor dataset is used for evaluating the proposed GA. The genetic parameters are given in Table 3.2.

Table 3.2 Genetic Algorithm Parameters

GA Population	500
Selection Rule	Roulette wheel
Cross over probability	0.6
Mutation Probability	0.004
Fitness Function	As defined
NGmin	10
NGmax	60

The schedule of data aggregation tree is created at the base station using E-Span and LPT algorithm. Two sets of experiments are conducted. The simulation parameters for the sensor network are as follows: a) network deployment areas are 50×50 and 100m×100m, b) initial energy of each sensor node is 1 Joule, c) sensor nodes are randomly deployed in the given area, d) each tree is used for 10 rounds, e) each experiment is conducted for 3 simulation scenarios and the average is used for documentation, f)base station is located at the center of the sensor network, g) MAC layer, 802.11, is used in simulation. Table 3.3 gives the simulation parameters.

For performance evaluation, the proposed GA is compared only with energy efficient data aggregation spanning tree-based approaches such as Power Efficient Data gathering and Aggregation Protocol- Power Aware (PEDAPPA) and E-Span. Although EESR dynamically adjusts frequencies of trees usage, in this thesis, the frequencies of trees usage is constant for all the trees, as given in PEDAPPA.

Table 3.3 Simulation Parameters

Parameter	Remarks
Network Area	50×50 m2 and 100×100m2
Number of nodes	25, 50 and 75
Base station location	a) at the center of network field, and b) outside network field at the distance of 100 m.
data packet length	It is assumed that each sensor generates fixed length data packet of size 1000 bits.
Initial Energy	Each sensor is initialized with 1 J

In other words, the algorithms are evaluated based on the creation of energy efficient data aggregation trees only. The optimal frequency of usage of each tree could be a separate topic of research and can be applied independently on any algorithm. Table 3.4 shows the simulation results for PEDAPPA, E-Span and GA algorithm for dense network field. The results show that GA is better than PEDAPPA and E-Span algorithms for most of the cases.

Table 3.4 Network lifetime for Dense network 50m×50m

Nodes	Proposed GA with ARM	PEDAPPA	E-Span
25	521	497	499
50	451	432	428
75	405	387	388

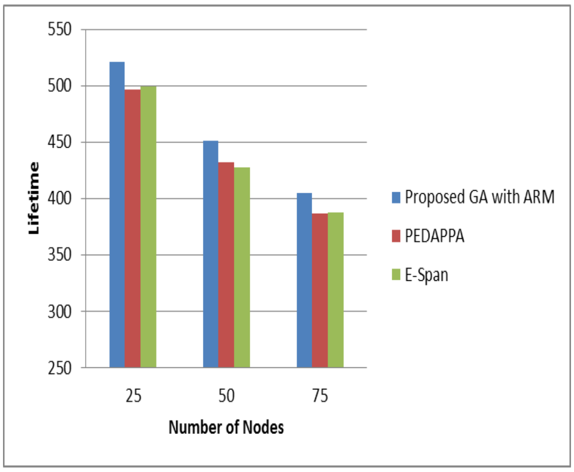

Fig 3.1 Network lifetime for Dense network 50m×50m

Table 3.5 shows the simulation results for PEDAPPA, E-Span and GA algorithm for sparse network field. It is perceived from Figure 6.5 that the proposed GA with ARM increases the network lifetime by 4.42% when compared with PEDAP-PA and 4.5% when compared with E-Span. The results show that GA is better than PEDAPPA and E-Span algorithms for most of the cases.

Table 3.5 Network lifetime for sparse network 100m×100m

Nodes	Proposed GA with ARM	PEDAPPA	E-Span
25	481	451	445
50	416	375	378
75	386	354	352

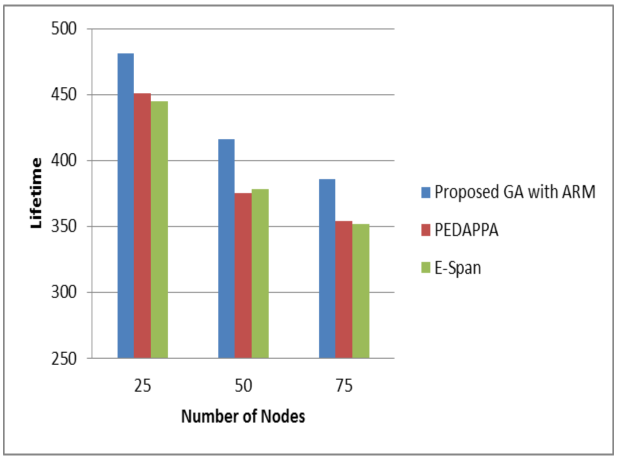

Fig 3.2 Network lifetime for sparse network 100m×100m

It is perceived from Fig 3.2 that the proposed GA with ARM increases the network lifetime of sparse network by 8.13% when compared with PEDAPPA and 8.48% when compared with E-Span. The proposed method is more effective in sparse networks.

4. CONCLUSION

GA is used to create energy efficient data aggregation trees. In this study, the amount of data sent to sink is reduced using association rule mining and to further reduce the energy consumption of the network, optimal routes are chosen to transmit data to the sink based on energy consumption. The proposed method is able to discover the associations to make predictive analysis such as node failure, asymmetric links. The rules found forms the basis for coding solutions in the proposed Genetic Algorithm. Intel lab sensor dataset is used for evaluating the proposed GA. GA is implemented to generate balanced and energy efficient data aggregation spanning trees for wireless sensor networks. E-Span and LPT are used to create data aggregation trees. In a data gathering round, a single best tree consumes lowest energy from all nodes but assigns more load to some sensors. As a result, the energy resources of heavily loaded nodes will be depleted earlier than others. The proposed GA extends network lifetime.

5. REFERENCES

[1] F. Zhao and L. J. Guibas. Wireless Sensor Networks: an Information Processing Approach. Morgan Kaufmann publisher, 2002.

[2] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. "Wireless Sensor Networks: A Survey" IEEE Transactions on Systems, Man and Cybernetics (B), Vol. 38(4), pp:394-422, 2002.

[3] C. S. R. Murthy, and B. S. Manoj. Ad Hoc Wireless Networks Architecture and Protocols. Prentice Hall PTR, New Jersey, 2004.

[4] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva. "Directed Diffusion for Wireless Sensor Networking", IEEE/ACM Transactions on Networking, Vol. 11, no. 1, pp: 2-16, 2003.

[5] Intel Lab Sensor Data. http://berkeley.intel-research.net/labdata/.

[6] Sotiris Kotsiantis, Dimitris Kanellopoulos, 2006, "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 71-82

[7] Agrawal, R., Imielinski, T., And Swami, A., 1993, Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 207-216.

[8] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. In *VLDB 1994*, Santiago de Chile, Chile, September 1994.

[9] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In *SIGMOD 2000*, Dallas, USA, May 2000

[10] Zaki, M. J.: Scalable Algorithms for Association Mining, IEEE Transactions on Knowledge and Data Engineering, 12(3), 2000, 372–390.

[11] Mata, J., Alvarez, J., Riquelme, J.: An Evolutionary Algorithm to Discover Numeric Association Rules, ACM Symposium on Applied Computing, Madrid, Spain, March 2002.

[12] Mata, J., Alvarez, J., Riquelme, J.: Mining Numeric Association Rules with Genetic Algorithms, 5th International Conference on Artificial Neural Networks and Genetic Algorithms, Taipei, Taiwan, April 2001.

[13] Alippi, C., Anastasi, G., Di Francesco, M. and Roveri, M. "An Adaptive Sampling Algorithm for Effective Energy Management in Wireless Sensor Networks with Energy-Hungry Sensors", Instrumentation and Measurement, IEEE Transactions on, Vol. 59, No. 2, pp. 335-344, 2010.

[14] Alippi, G., Anastasi, C., Galperti, F., Mancini, M. and Roveri "Adaptive Sampling for Energy Conservation in Wireless Sensor Networks for Snow Monitoring Applications", Proc. of IEEE International Workshop on Mobile Ad-hoc and Sensor Systems for Global and Homeland Security (MASS-GHS 2007), Pisa (Italy), October 8, 2007.

[15] Boukerche, A. and Samarah, S. "An Efficient Data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks", Communications, ICC '07. IEEE International Conference on, Vol. 2, No. 2, pp. 3936 - 3941, 24-28 June 2007.

[16] Chong, S.K., Gaber, M.M., Krishnaswamy, S. and Loke, S. W. "Energy Conservation in Wireless Sensor Networks: A Rule-Based Approach", Knowledge and Information Systems, Vol. 28, No. 3,pp. 579-614, 2011.

[17] Fernando Berzal, Juan Cubero, C., Nicolas Marín and José-María Serrano "TBAR: An Efficient Method for Association Rule Mining in Relational Databases", Data and Knowledge Engineering, Vol. 37, No. 1, pp. 47- 64, 2001.

[18] Kandris, D., Tsioumas, P., Tzes, A., Nikolakopoulos, G. and Vergados, D. D. "Power conservation through energy

efficient routing in wireless sensor networks", Sensors, Vol. 9, No. 9, pp. 7320-7342, 2009.

[19] Maraiya, K., Kant, K. and Gupta, N. "Wireless Sensor Network: A Review on Data Aggregation", International Journal of Scientific and Engineering Research, Vol. 2, No. 4, pp. 1-6, 2011.

[20] Venturini, G. "Sia: a Supervised Inductive Algorithm with Genetic Search for Learning Attrib-ute based Concepts", in: Proceedings of the European Conference on Machine Learning, of Lecture Notes in Computer Science, Viena, Austria, Vol. 667, pp. 280–296, April 1993.

[21] Vergados, D.J., Pantazis, N. and Vergados, D. Energy-Efficient Route Selection Strategies for Wireless Sensor Networks", Mobile Netw. Appl., Vol. 13, pp. 285–296, 2008.