Australian Journal of Basic and Applied Sciences, 9(10) Special 2015, Pages: 187-194

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Data Aggregation Technique using Genetic Algorithm

[1]S. Rajanarayanan and [2]Dr. C. Suresh Gnana Dhas

[1]*Research Scholar, Department of Computer Science and Engineering, St.Peters University, Chennai, Tamil Nadu, India,*
[2]*Professor, Department of Computer Science and Engineering, Vivekandha college of engineering*

ARTICLE INFO

Article history:
Received 28 January 2015
Accepted 25 February 2015
Available online 6 March 2015

Keywords:
Wireless Sensor Network (WSNs), Network Lifetime, mobile sinks, single or multiple hops and Genetic Algorithm (GA).

ABSTRACT

For Wireless Sensor Network (WSNs), conservation of available energy at each sensor node is the major design issues. Sensor nodes monitor some surrounding environmental phenomenon, process the data obtained and forward this data towards a base station located on the periphery of the sensor network. Base station(s) collect the data from the sensor nodes and transmit this data to some remote control station. The data collection approaches in WSNs with path constrained mobile sinks and path-controllable mobile sinks, which can be sub classified according to the communication mode (single or multiple hops) and the number of mobile sinks. This paper explores the idea of exploiting the mobility of data collection points (sinks) for the purpose of increasing the lifetime of a wireless sensor network with energy-constrained nodes. Genetic Algorithm (GA) is proposed to collect data efficiently and to extend the lifetime of sensor network, andto deploy multiple mobile base stations.

To Cite This Article: S. Rajanarayanan and Dr. C. Suresh Gnana Dhas, Data Aggregation Technique using Genetic Algorithm. ***Aust. J. Basic & Appl. Sci.,*** *9(10): 187-194, 2015*

INTRODUCTION

In Wireless Sensor Network (WSNs), sensor nodes are combine the wireless communication infrastructure with the sensing technology. Instead of transmitting the perceived data to the control center through wired links, ad hoc communication methods are utilized, and the data packets are transmitted using multi-hop connections. The problem of data collection in WSNs is encountered in many scenarios such as monitoring physical environments such as tracking animal migrations in remote-areas, weather conditions in national parks, habitat monitoring on remote islands, city traffic monitoring etc. The objective is to collect data from sensors and deliver it to an access point in the infrastructure. These systems are expected to run unattended for long periods of time (order of months). The principal constraint is the energy budget of the sensors which is limited due to their size and cost (Kulkarni, 2011).

The sensors coordinate among themselves to form a communication network such as a single multi-hop network or a hierarchical organization with several clusters and cluster heads. The sensors periodically sense the data, process it and transmit it to the base station. The frequency of data reporting and the number of sensors which report data usually depends on the specific application. Data gathering is defined as the systematic collection of sensed data from multiple sensors to be eventually transmitted to the base station for processing. Since sensor nodes are energy constrained, it is inefficient for all the sensors to transmit the data directly to the base station. Data generated from neighboring sensors is often redundant and highly correlated. In addition, the amount of data generated in large sensor networks is usually enormous for the base station to process. Hence, we need methods for combining data into high quality information at the sensors or intermediate nodes which can reduce the number of packets transmitted to the base station resulting in conservation of energy and bandwidth. This can be accomplished by data aggregation. Data aggregation is defined as the process of aggregating the data from multiple sensors to eliminate redundant transmission and provide fused information to the base station. Data aggregation usually involves the fusion of data from multiple sensors at intermediate nodes and transmission of the aggregated data to the base station (sink) (Rajagopalan, 2006).

Data aggregation attempts to collect the most critical data from the sensors and make it available to the sink in an energy efficient manner with minimum data latency. Data latency is important in many applications such as environment monitoring where the freshness of data is also an important factor. It is critical to develop energy efficient data aggregation algorithms so that network lifetime is enhanced. There are several factors which determine the energy efficiency of a sensor network such as network

Corresponding Author: S. Rajanarayanan, Research Scholar, Department of Computer Science and Engineering, St. Peters University, Chennai,Tamil Nadu, India

architecture, the data aggregation mechanism and the underlying routing protocol.

From a networking. perspective, WSNs generally follow the well-established ad hoc paradigm of communication: Data delivery between any two nodes follows a multi-hop route. Differently from ad hoc networks, where any two nodes can be source and destination of data packets, in WSNs data generated by the sensor nodes are sent to one or more data collection points (the sinks). Sinks are considered resource-rich, i.e., energy, processing power and memory are not considered a limitation for their prolonged functioning and operations. Sensor nodes are instead usually quiteconstrained in terms of battery power, storage, and computational capabilities. Due to the large number of sensor nodes deployed and to their being often placed in hostile, un accessible environments, it is not viable to recharge/replace their batteries. Whenever a node depletes its energy it is considered “dead,” i.e., no longer able to perform its sensing or communication duties.

The efficiency of the sensor network investment is directly related with the length of the reliable monitoring duration of the field. The better energy control mechanisms are used in the sensor nodes’ firmware and in the network management techniques, the longer the network will be serving their investors. Therefore, the limited battery resource of the sensors should be handled efficiently. Current approaches involve forming an ad-hoc network among the sensor nodes to send data. However, this faces the following energy related issues. Firstly, in a network, the energy required for transmitting data over one hop is quite large. This is because sensors may be far from each other and the transmission power required increases as the fourth power of distance. Secondly, in an ad-hoc network sensors have to not only send their data, but also forward data for other sensors. Thirdly, the network has routing hotspots near the access points. Sensors that are near the access points have to forward many more packets and drain their battery much more quickly.

There are multiple ways in which the sensor readings aretransferred from the sensors to a central location. Usually, the readings taken by the sensor nodes are relayed to a basestation for processing using the ad-hoc multi-hop networkformed by the sensor nodes. While this is surely a feasibletechnique for data transfer, it creates a bottleneck in the network.The nodes near the base station relay the data fromnodes that are farther away. This leads to a non-uniform depletionof network resources and the nodes near the basestation are the first to run out of batteries. If these nodesdie, then the network is for all practical purposes disconnected.Periodically replacing the battery of the nodes forthe large scale deployments is also infeasible (Somasundara, 2004).

Exploiting mobility for communication in ad-hoc networks has received much attention recently (Jain, 2006). The work focuses on scenarios in which there is no immediate end-to-end path between two nodes that wish to communicate, usually because of limited radio range. If the nodes are mobile, endto-end connectivity may be achieved by buffering data at the nodes and waiting to transfer until they are in range of accesspoints.

Genetic Algorithm (GA) is one of the most powerful heuristics for solving optimization problems that is based on natural selection, the process that drives biological evolution. The GA repeatedly modifies a population of individual solutions. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation. Over successive generations, the population "evolves" towards an optimal solution. GAs can be applied to solve a variety of optimization problems that are not well suited for standard optimization algorithms, including problems in which the objective function is discontinuous, non-differentiable, stochastic, or highly nonlinear. Several researchers have successfully implemented GAs in WSN design, this led to the development of several other GA-based application-specific approaches in WSN design, mostly by the construction of a single fitness function. However, these approaches either cover limited network characteristics or fail to incorporate several application specific requirements into the performance measure of the heuristic.

This paper explores the idea of exploiting the mobility of data collection points (sinks) for the purpose of increasing the lifetime of a wireless sensor network with energy-constrained nodes. Genetic Algorithm (GA) is proposed to collect data efficiently and to extend the lifetime of sensor network, and to deploy multiple mobile base stations.The remainder of this paper is organized as follows. In Section 2, we discuss some related studies. Section 3 describes the details of the proposed algorithm. Section 4presents and assesses the simulation results. Finally, Section 5 concludes the paper.

Literature survey:

Luo *et al* (2006) investigated the approach that makes use of a mobile sink forbalancing the traffic load and in turn improving network lifetime. MobiRoute, a routing protocol that effectively supports sinkmobility is proposed. Through intensive simulations in TOSSIM with a mobile sinkand an implementation of MobiRoute, the feasibility of themobile sink approach was provedby demonstrating the improved network lifetimein several deployment scenarios.

Jain *et al* (2006) analyzed an architecture based on mobility to address the problem of energy efficient data collection in a sensor network. The proposed approach exploits mobile nodes present in the sensor field as forwarding agents. As a mobile node moves in close proximity to sensors, data is transferred to the mobile node for later depositing at the destination. An analytical model was presentedto understand the key performance metrics such as data transfer, latency to the destination, and power. Parameters for the model include: sensor buffer size, data generation rate, radio characteristics, and mobility patterns of mobile nodes. Through simulation proposed model was verified and showed that it can provide substantial savings in energy as compared to the traditional ad-hoc network approach.

Oyman and Ersoy (2004) focused on the multiple sink location problems in large-scale WSNs. Different problems depending on the design criteria are presented. Locating sink nodes to the sensor environmentwas considered, where given a time constraint that states the minimum required operational time for the sensor network. A sample sink location casewas demonstrated, where the number of sink nodes was known before the deployment phase. Further, the authours implemented the solution for this BSL problem, and presented the corresponding energy and disconnected region maps on a sample sensor network for different snapshots in time.

Basagni *et al* (2008) demonstrated the advantages of using controlled mobility in WSNs for increasing their lifetime, i.e., the period of time the network is able to provide its intended functionalities. In order to determine sink movements, a Mixed Integer Linear Programming (MILP) analytical model was first definedwhose solution determines those sink routes that maximize network lifetime.Further the first heuristics for controlled sink movements that are fully distributed and localizedwas defined. A Greedy Maximum Residual Energy (GMRE) heuristic moves the sink from its current location to a new site as if drawn toward the area where nodes have the highest residual energy. A simple distributed mobility scheme (Random Movement orRM) was also introducedaccording to which the sink moves uncontrolled and randomly throughout the network. The different mobility schemes are compared through extensive ns2-based simulations in networks with different nodes deployment, data routing protocols, and constraints on the sink movements. It was observed that moving the sink always increases network lifetime.

Gandham *et al* (2004) proposed to split the lifetime of a sensor network into equal periods of time referred to as rounds and model the energy constrained routing during a round as polynomial-time solvable flow problems. The flow information from an optimum solution to a flow problem is then used as a basis for an energy-efficient routing protocol. Through simulations, it was demonstrated that the proposed routing algorithm performs significantly better than the shortest path based algorithms and consumes less energy than the ILPbased method.

Wang *et al* (2005) presented a novel linear programming formulation for the joint problems of determining the movement of the sink and the sojourn time at different points in the network that induce the maximum network lifetime. Differently from previous solutions, the objective function maximizes the overall network lifetime (here defined as the time till the first node “dies” because of energy depletion) rather than minimizing the energy consumption at the nodes. For WSNs with up to 256 nodes the model produces sink movement patterns and sojourn times leading to a network lifetime up to almost five times that obtained with a static sink. Simulation results are performed to determine the distribution of the residual energy at the nodes over time. These results confirm that energy consumption varies with the current sink location, being the nodes more drained those in the proximity of the sink. Furthermore, the proposed solution for computing the sink movement results in a fair balancing of the energy depletion among the network nodes.

Liao *et al* (2008) considered the problem of constructing data aggregation tree in a WSN for a group of source nodes to send sensory data to a single sink node. The ant colony system provides a natural and intrinsic way of exploring search space in determining data aggregation. Moreover, we propose an ant colony algorithm for data aggregation in wireless sensor networks. Every ant will explore all possible paths from the source node to the sink node. The data aggregation tree is constructed by the accumulated pheromone. Simulations have shown that our algorithm can reduce significant energy costs.

Al-Karaki *et al* (2009) focused on the *joint* problem of optimal data routing with data aggregation en route such that the above mentioned objective is achieved. Grid-based Routing and Aggregator Selection Scheme (GRASS), waspresentwda scheme for WSNs that can achieve low energy dissipation and low latency without sacrificing quality. GRASS embodies optimal (exact) as well as heuristic approaches to find the minimum number of aggregation points while routing data to the Base-Station such that the network lifetime is maximized. The results show that, when compared to other schemes, GRASS improves system lifetime with acceptable levels of latency in data aggregation and without sacrificing data quality.

Zhu *et al* (2011) intelligent mobile data collector has been explored to collect data for improving the networking facilities in the system. To improve the efficiency of data collecting, an efficient energy-aware distributed intelligent data-gathering algorithm (DIDGA) is proposed to plan the data-gathering path for the mobile collector in WSNs. DIDGA will reduce high energy expenditure in multihop routings and increase the efficiency of the mobile collector to gather data. DIDGA creates and constructs a minimum connected dominating set (MCDS) based on maximal independent sets (MISs) in distributed and localized manner, which will reduce high energy expenditure in multihop routing with the cluster head, and it will select the node with more power to be the cluster head in turn to prolong the network lifetime. Then DIDGA restricts to the hop counts of the sensed data transmission by communicating with the cluster head in one hop. Sensor nodes' data transmission can cooperate with mobile collector's data-gathering path, which will increase the efficiency of the mobile collector to collect data. DIDGA disperses routing hot spots in WSNs. So it extends the network's lifetime.

Methodology:

The main objective of this study is to increase the useful lifetime of sensor networks. This achieved by application of multiple mobile sinks.

Problem formulation:

At the beginning of each round we need to determine the location of the base stations at feasible sites (2003). We refer to this problem as the Base Station Location (BSL) problem. The sensor network is represented as a graph *G(V, E)* where

a. $V = V_s \cup V_f$ where V_s represents the sensor nodes and V_f represents the feasible sites.

b. $E \subseteq V \times V$ represents the set of wireless links.

Let $K_{\max}$ be the maximum number of base stations. Let a round consist of T timeframes. Each sensor node generates one unit of data in every timeframe. At the beginning of a round, let a sensor node *i*have residual energy RE_i A constraint we impose is that during the round, the total energy spent by sensor node *i*is at most αRE_i, where $0 < \alpha \leq 1$ is a parameter.

Flow based Routing:

Sensor nodes can use the flow information obtained by solving the integer linear program to route messages in an energy efficient manner. Consider sensor node A with its incoming and outgoing number of messages as shown in the Figure 1. Once a sensor node has this information it would perform its routing as described below.

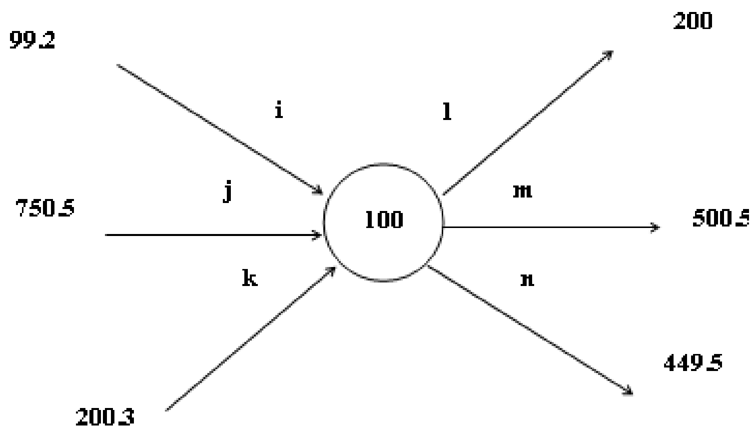

Fig. 1: Flow Based Routing.

(i) For every outgoing link a counter is maintained. The value of the counter is set to the floor of the flow going out on that particular link.
(ii) Whenever a node needs to transmit its packets, it would select one of the outgoing links in a round robin fashion.
(iii) If the counter value of the selected link is greater than the number of packets that have to be transmitted, then all the packets are transmitted on that link and counter value is decremented by the number of packets transmitted; otherwise the number of packets equal to the counter value of the link are transmitted along the link and its counter values is set to 0. To transmit the remaining messages outgoing links are selected in a round robin fashion.
(v) If the counter value of all the outgoing links is zero then a link is selected arbitrarily and all the packets are transmitted on this link.

Proposed Implementation Method:

Network lifetime as the time period for the first node to run out ofits energy reserve (2000). When evaluating this quantity, the problemis convertedof maximizing network lifetime. Some factors to be considered for the implementation of proposed method are gathering of topology information, tracking residual energy of nodes, updating routing information of all nodes and optimizing genetic algorithm.

GA [15] is used to create energy efficient clusters for data dissemination in WSNs and determine the locations of the base stationsand a flow-based routing protocol.Since energy consumption during communication is a major energy depletion factor, the number of transmissions must be reduced to achieve extended battery life. A GA performs fitness tests on new structures to select the best population. Fitness determines the quality of the individual on the basis of the defined criteria. In nature, an individual's fitness is its ability to pass on its genetic material. Anything that contributes to this ability contributes to the organism's overall fitness. This ability includes traits that enable it to survive and further reproduce. In a GA, fitness is evaluated by the function defining the problem. The fate of an individual chromosome depends on the fitness value; the better the fitness value, the better the chance of survival.

GAs have a number of parameters that are problem specific and need to be explored and tuned so that the best algorithm performance is achieved. These parameters are the population size, the probabilities of crossover and mutation and the type of crossover. In the beginning, a number of experiments were carried out to determine the most appropriate population size. Sizes from 100 to 1000 individuals in orders of hundreds were tested. The best performance, by means of maximizing the corresponding fitness function, was achieved with a population size of 300 individuals. Then, several explorations were performed with probabilities of crossover ranging from 0.3 to 0.9 for both one-point and two-point crossover types and probabilities of mutation ranging from 0.0001 to 0.01. The results led to the use of one-point crossover with probability pc = 0.8 and probability of mutation pm = 0.005.

RESULTS AND DISCUSSIONS

In this study, table 1 to 4 and figure 1 to 4 shows the result table and graph for average end to end delay, packet loss rate, number of nodes alive and remaining energy in joules respectively.

Table 1: Average End to End Delay.

Number of nodes	Number of Sinks =3	Number of Sinks =5	Number of Sinks =7
75	0.001269	0.001334	0.001306
150	0.001376	0.001614	0.00138
225	0.013186	0.014906	0.01389
300	0.022284	0.018148	0.022215
375	0.046353	0.041454	0.047885
450	0.04891	0.044692	0.051963

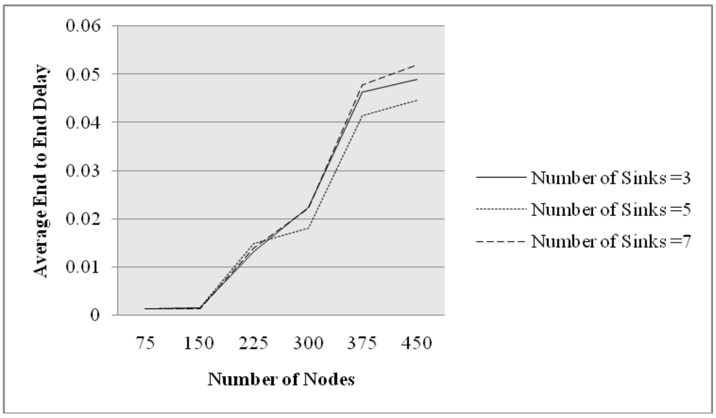

Fig. 2: Average End to End Delay.

From table 3.1 and figure 2 it is observed that the average end to end delay is measured with different number of sinks such 3, 5 and 7. Result show that the delay reduced for number of sinks=3 by 4.99% and 2.87% than number of sinks=5 and 7 respectively at node 75. Also at node 450, the delay reduced for number of sinks=5 by 9.01% and 15.05% than number of sinks=3 and 7 respectively.

Table 2: Average Packet Loss Rate %.

Number of nodes	Number of Sinks =3	Number of Sinks =5	Number of Sinks =7
75	9.58	8.52	7.98
150	14.59	12.89	11.86
225	15.01	14.49	12.16
300	20.62	19.55	17.97
375	28.53	27.04	24.34
450	39.22	37.96	27.98

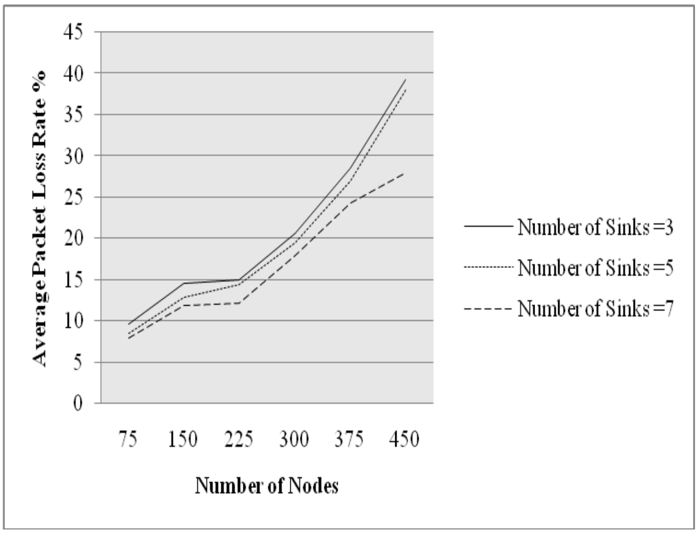

Fig. 3: Average Packet Loss Rate %.

From table 2 and figure 3 it is observed that the average packet loss rate is measured with different number of sinks such 3, 5 and 7. Result show that the packet loss rate reduced for number of sinks=7 by 18.22% and 6.54% than number of sinks=3 and 5 respectively at node 75. Similarly at node 450, the delay reduced for number of sinks=7 by 33.45% and 30.27% than number of sinks=3 and 5 respectively.

Table 3: Percentage of Nodes Alive.

Number of rounds	Number of Sinks =3	Number of Sinks =5	Number of Sinks =7
0	100	100	100
100	100	100	100
200	91	93	99
300	70	86	95
400	74	78	82
500	23	55	66
600	3	19	22
700	0	0	4
800	0	0	0

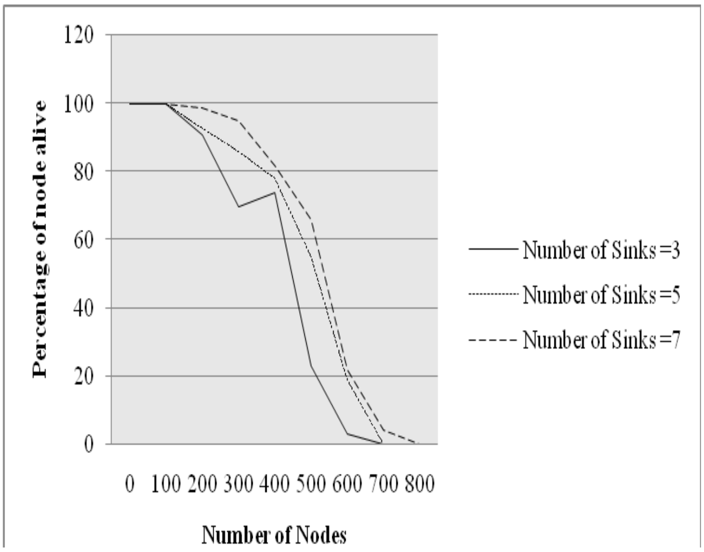

Fig. 4: Percentage of Nodes Alive.

Table 4: Remaining Energy in Joules.

Number of rounds	Number of Sinks =3	Number of Sinks =5	Number of Sinks =7
0	0.5	0.5	0.5
100	0.46	0.51	0.48
200	0.26	0.36	0.39
300	0.2	0.32	0.36
400	0.21	0.3	0.34
500	0.12	0.19	0.29
600	0	0.11	0.25
700	0	0	0.11
800	0	0	0

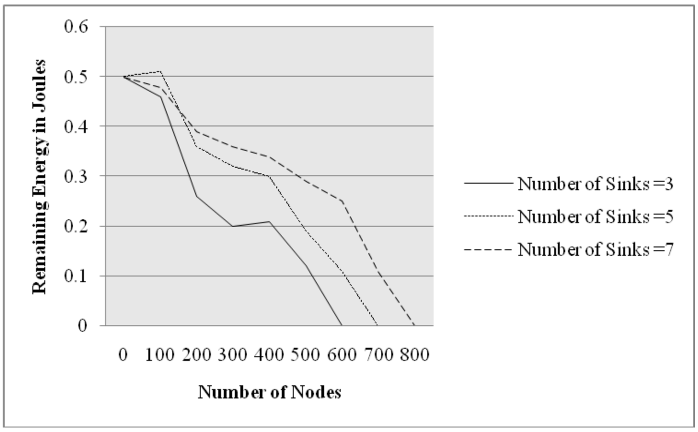

Fig. 5: Remaining Energy in Joules.

From table 3 and figure 4 it is observed that the percentage of nodes alive is better for number of sinks=7 in the range of 0 to 200% than number of sinks=3 and 5.

From table 4 and figure 5 it is observed that the remaining energy in joules is better for number of sinks=7 in the range of 0 to 200% than number of sinks=3 and 5.

Conclusion:

The battery resource of the sensor nodes should be managed efficiently, in order to prolong network lifetime in wireless sensor networks. Moreover, in large-scale networks with a large number of sensor nodes, multiple sink nodes should be deployed, not only to increase the manageability of the network, but also to reduce the energy dissipation at each node. In this work the efficient information gathering and improved network lifetime is aimed at using mobile sink to collect data from the sub-sinks.Genetic algorithm is used to optimize the number of mobile sinks. Result show that the packet loss rate reduced for number of sinks=7 by 18.22% and 6.54% than number of sinks=3 and 5 respectively at node 75. Similarly at node 450, the delay reduced for number of sinks=7 by 33.45% and 30.27% than number of sinks=3 and 5 respectively.

REFERENCES

Al-Karaki, J.N., R. Ul-Mustafa, A.E. Kamal, 2009. Data aggregation and routing in wireless sensor networks: Optimal and heuristic algorithms. Computer networks, 53(7): 945-960.

Basagni, S., A. Carosi, E. Melachrinoudis, C. Petrioli, Z.M. Wang, 2008. Controlled sink mobility for prolonging wireless sensor networks lifetime. Wireless Networks, 14(6): 831-858.

Chang, J.H., L. Tassiulas, 2000. Energy Conserving Routing in Wireless Ad-hoc Networks. In: Proc. of the 19th IEEE INFOCOM.

Gandham, S., M. Dawande, R. Prakash, 2004. An integral flow-based energy-efficient routing algorithm for wireless sensor networks. In Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE, 4: 2341-2346. IEEE.

Gandham, S.R., M. Dawande, R. Prakash, S. Venkatesan, 2003. Energy efficient schemes for wireless sensor networks with multiple mobile base stations. In Global telecommunications conference, 2003. GLOBECOM'03. IEEE, 1: 377-381. IEEE.

Jain, S., R.C. Shah, W. Brunette, G. Borriello, S. Roy, 2006. Exploiting mobility for energy efficient data collection in wireless sensor networks. Mobile Networks and Applications, 11(3): 327-339.

Kulkarni, R.V., A. Forster, G.K. Venayagamoorthy, 2011. Computational intelligence in wireless sensor networks: A survey. Communications Surveys & Tutorials, IEEE, 13(1): 68-96.

Liao, W.H., Y. Kao, C.M. Fan, 2008. Data aggregation in wireless sensor networks using ant colony algorithm. Journal of Network and Computer Applications, 31(4): 387-401.

Luo, J., J. Panchard, M. Piórkowski, M. Grossglauser, J.P. Hubaux, 2006. Mobiroute: Routing towards a mobile sink for improving lifetime in sensor networks. In Distributed Computing in Sensor Systems, pp: 480-497. Springer Berlin Heidelberg.

Oyman, E.I., C. Ersoy, 2004. Multiple sink network design problem in large scale wireless sensor networks. In Communications, 2004 IEEE International Conference on, 6: 3663-3667. IEEE.

Praveena, K., T. Sripriya, 2013. Genetic Algorithm Based Data Aggregation Using Mobile Sink in Wireless Sensor Networks. In *International Journal of Engineering Research and Technology* (Vol. 2, No. 2 (February-2013)). ESRSA Publications.

Rajagopalan, R., P.K. Varshney, 2006. Data aggregation techniques in sensor networks: A survey. Electrical Engineering and Computer Science. Paper, 22: 1-30.

Somasundara, A.A., A. Ramamoorthy, M.B. Srivastava, 2004. Mobile element scheduling for efficient data collection in wireless sensor networks with dynamic deadlines. In Real-Time Systems Symposium, 2004. Proceedings. 25th IEEE International, pp: 296-305. IEEE.

Wang, W., V. Srinivasan, K.C. Chua, 2005. Using mobile relays to prolong the lifetime of wireless sensor networks. In Proceedings of the 11th annual international conference on Mobile computing and networking, pp: 270-283. ACM.

Zhu, R., Y. Qin, J. Wang, 2011. Energy-aware distributed intelligent data gathering algorithm in wireless sensor networks. International Journal of Distributed Sensor Networks, 2011.