ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 1, Issue 9, November 2012

A Comparative study of Data Gathering algorithms for a Mobile Sink in Wireless Sensor Network

Jeba Princess Jafrin.J, Anitha Christy Angelin.P

***Abstract*— As Wireless Sensor Networks (WSN) has become a rapidly growing field of interest, it is essential to know the various aspects, functionalities and the methodologies involved in this field. Data gathering in WSN involves collection of data from widely distributed set of sensor nodes. There are different algorithms to gather data in WSN based on the nature of the network topology and the nodes involved. Though there are many different techniques to achieve data gathering, not all of them provide optimized result. Also, in data gathering it is important to recognize a scheme that will ensure fairness in the data gathered. This paper provides a detailed understanding to the various data gathering methodologies for mobile sinks in WSNs.**

***Index Terms*— Cluster, Sensor node, Sink.**

I. INTRODUCTION

Wireless sensor networks are a trend of the past few years, and they involve deploying a large number of small nodes. The nodes then sense environmental changes and report them to other nodes over flexible network architecture. Sensor nodes are great for deployment in hostile environments or over large geographical areas.

The sensor nodes leverage the strength of collaborative efforts to provide higher quality sensing in time and space as compared to traditional stationary sensors, which are deployed in the following two ways:

- *Sensors can be positioned far from the actual phenomenon. Here, large sensors that use some complex techniques to distinguish the targets from environmental noise are required.*
- *Several sensors that perform only sensing can be deployed. The position of the sensors and interactions topology is cautiously engineered. They transmit time series of the sensed incident to central nodes where computations are performed.*

Manuscript received Nov 27, 2012.
***J. Jeba Princess Jafrin**, Department of Computer Science and Engineering, Karunya University, C oimbatore, India.*
***P. Anitha Christy Angelin,** Asst. Prof., Department of Computer Science and Engineering, Karunya University, Coimbatore, India.*

A sensor network is designed to collect information from a physical environment. In many applications, it is more appropriate to address nodes in a sensor network by physical properties, such as node locations or proximity, than by IP addresses. How and where data is generated by sensors and consumed by users will affect the way data is compressed, routed, and aggregated. Because of the peer-to-peer connectivity and the lack of a global infrastructure support, the sensors have to rely on discovery protocols to construct local models about the network and environment.

II. KEY CONCEPTS

A Wireless Sensor Network is a collection of nodes organized into a cooperative network. Each node consists of processing capability (one or more microcontrollers, CPUs or DSP chips), may contain multiple types of memory (program, data and flash memories), have a RF transceiver (usually with a single Omni-directional antenna), have a power source (e.g., batteries and solar cells), and accommodate various sensors and actuators. The nodes communicate wirelessly and often self-organize after being deployed in an ad hoc fashion.

A sensor node might vary in size from that of a shoebox down to the size of a grain of dust, although functioning "motes" of genuine microscopic dimensions have yet to be created. The cost of sensor nodes is similarly variable, ranging from a few to hundreds of dollars, depending on the complexity of the individual sensor nodes. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed and communications bandwidth.

The topology of the WSNs can vary from a simple star network to an advanced multi-hop wireless mesh network. The propagation technique between the hops of the network can be routing or flooding.

III. COMPARATIVE STUDY

A detailed comparative study has been done in the various schemes involved in data gathering. We will discuss some of those schemes in this section.

A. DIRECTED DIFFUSION

Chalermek proposed Directed diffusion [2], which is datacentric in that all communication is for named data. All nodes in a directed diffusion-based network are application-aware. This enables diffusion to achieve energy savings by selecting empirically good paths and by caching and processing data in-network (*e.g.*, data aggregation).

Directed diffusion is significantly different from IP-style communication where nodes are identified by their end-points, and inter-node communication is layered on an end-to-end delivery service provided within the network. With directed diffusion one can realize robust multi-path delivery, empirically adapt to a small subset of network paths, and achieve significant energy savings when intermediate nodes aggregate responses to queries.

The use of directed diffusion is evaluated for a simple remote-surveillance sensor network analytically and experimentally.

Fig.1, shows the simplified scheme followed by the directed diffusion. It is composed of three main steps, interest propagation, initial gradients set up and the data delivery along the reinforced path. A sensor node that detects a target searches its interest cache for a matching interest entry. In this case, a matching entry is one whose rect encompasses the sensor location, and the type of the entry matches the detected target type.

When it finds one, it computes the highest requested event rate among all its outgoing gradients. The node tasks its sensor subsystem to generate event samples at this highest data rate using various aspects such as Gradient establishment, reinforcement, multiple sinks and Repair.

The directed diffusion paradigm for designing distributed sensing algorithms has several lessons we can draw from the preliminary evaluation of diffusion.

First, directed diffusion has the potential for significant energy efficiency. Even with relatively unoptimized path selection, it outperforms an idealized traditional data dissemination scheme like omniscient multicast. Second, diffusion mechanisms are stable under the range of network dynamics considered here. Finally, for directed diffusion to achieve its full potential, however, careful attention has to be paid to the design of sensor radio MAC layers. The evaluation indicates that directed diffusion can achieve significant energy savings and can outperform idealized traditional schemes (*e.g.*, omniscient multicast) under the investigated scenarios.

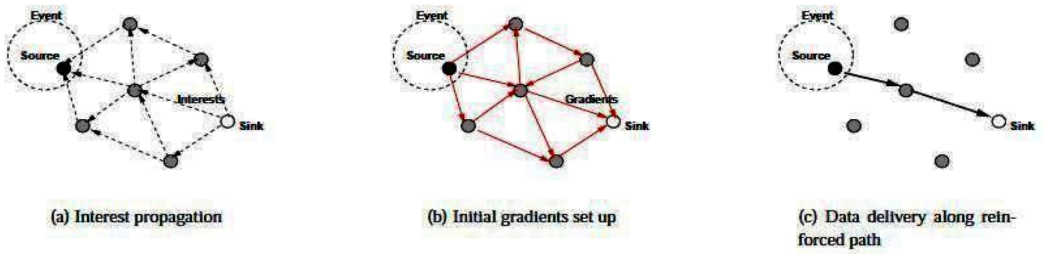


Fig1. A Simplified Schematic for directed diffusion

B. MULE ARCHITECTURE

Shah, proposed an analytical model [4], to 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.

The key idea of this approach is to exploit the presence of mobile nodes in the environment by using them as forwarding agents. This approach extends the lifetime of the network by minimizing the communication responsibility of the resource-constrained sensors.

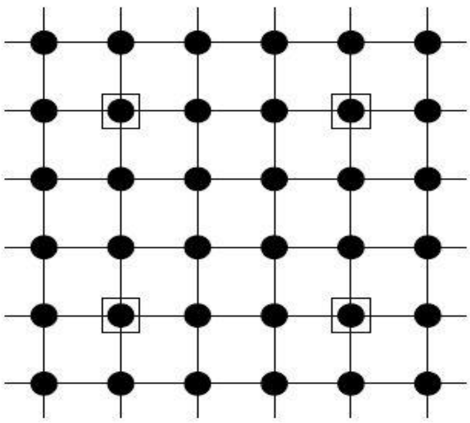


Fig2. A two dimensional grid with of the access points

However, the MULE architecture is limited to non real time applications which have mobility. Thus, the MULE architecture is not always the method of choice, but for certain applications it may be the most effective option.

The Data MULE architecture, exploits the presence of mobile entities (called MULEs) present in the environment. MULEs pick up data from the sensors when in close range, buffer it, and drop off the data to wired access points as shown in Figure2. This can lead to substantial power savings at the sensors as they only have to transmit over a short range. This method focuses on a simple analytical model for understanding performance as system parameters are scaled. The model assumes two-dimensional random walk for mobility and incorporates key system variables such as number of MULEs, sensors and access points. The performance metrics observed are the data success rate (the fraction of generated data that reaches the access points) and the required buffer capacities on the sensors and the MULEs. The modeling along with simulation results can be used for further analysis and provide certain guidelines for deployment of such systems.

The architecture connects sparse sensor networks at the cost of higher latencies. The main idea is to utilize the motion of the entities that are already present in an environment to provide a low power transport medium for sensor data. After introducing the architecture, the focus of the paper was on presenting a simple analytical model based upon two-dimensional random walks to provide insight into various performance metrics (data success rate and buffer sizes). The key observations made in this method are:

1) When the sensor buffer is large the buffer capacity on each MULE can be traded-off with the number of MULEs to maintain the same data success rate.

2) The change in the buffer capacity on each sensor needs to be greater than the change in the number of MULEs to keep the same data success rate.

From the protocol point of view, MULE-to-MULE communication and reliability using acknowledgments are some of the interesting issues. Another limitation of current work is the assumption that the sensors have to continuously listen in order to identify a MULE's presence. Approaches to increase the sleep time for sensors, such as reduced duty cycle, need to be explored along with their effect on system performance.

Through simulation this model is verified and it is shown that this approach can provide substantial savings in energy as compared to the traditional ad-hoc network approach.

C. SINK TRAIL PROTOCOL

Two energy-efficient proactive data reporting protocols, SinkTrail [5] and SinkTrail-S, for mobile sink based data collection are designed. These protocols feature low-complexity and reduced control overheads. Two unique aspects distinguish this approach from earlier ones: (1) Sufficient flexibility in the movement of mobile sinks is allowedto dynamically adapt to various terrestrial changes; and (2) without requirements of GPS devices or predefined landmarks, SinkTrail establishes a logical coordinate system for routing and forwarding data packets, making it suitable for diverse application scenarios.

The impact of several design factors in these algorithms are analyzed systematically. Both theoretical analysis and simulation results demonstrate that these algorithms reduce control overheads and yield satisfactory performance in finding shorter routing paths.

The SinkTrail and its improved version, SinkTrail-S protocol are two low-complexity, proactive data reporting protocols for energy-efficient data gathering. SinkTrail uses logical coordinates to infer distances, and establishes data reporting routes by greedily selecting the shortest path to the destination reference.

In addition, SinkTrail is capable of tracking multiple mobile sinks simultaneously through multiple logical coordinate spaces.It possesses desired features of geographical routing without requiring GPS devices or extra landmarks installed. SinkTrail is capable of adapting to various sensor field shapes and different moving patterns of mobile sinks.

Further, it eliminates the need of special treatments for changing field situations. The energy consumptions of SinkTrail and other representative approaches were systematically analyzed and validated the analysis through extensive simulations.

The results reveal that SinkTrail finds short data reporting routes and effectively reduces energy consumption. The impact of various design parameters used in SinkTrail and SinkTrail-S are investigated to provide guidance for implementation.

D. KAT MOBILITY

Nakayama, proposed K-means and TSP-based mobility (KAT mobility) scheme [1] based on clustering and routing optimization algorithms. This is a mobility model of mobile sinks that can efficiently collect usable sensed data in a wireless sensor network, even if some sensors are compromised or annihilated. The scheme consists of two modules: the K-means clustering algorithm and the approximate solution for TSP (Traveling Salesman Problem). Sensors are first clustered by using the K-means clustering algorithm, from which the cluster centers are determined as anchor points. The migration route of mobile sinks is determined as an approximate solution of TSP.

Here, it is assumed that an administrator distributes sensors to monitor the targeted area, and sensors are scattered at random positions and do not move afterwards. The administrator is assumed to be able to localize the actual coordinates of scattered sensors, which acquire the monitored data at their own positions. The amount of data which they can acquire per unit time is fixed, and buffers in which data can be stored temporarily are equipped.

The sensors are randomly scattered with different densities; some areas are sparse, some are dense. In addition to the underlying nature of scattered sensors, when the events of compromised sensors are invoked artificially, the sensors are distributed disproportionately. From the point of view of energy efficiency, if the administrator can control mobile sinks like a pilotless drone plane, the trajectory of mobile sinks can be optimized with relation to the distribution dynamically. Consequently, the KAT mobility scheme for a mobile sink, is based on the clustering algorithm and the route optimization.

After clustering the sensor nodes, this method navigates the mobile sink to traverse through the cluster centers according to the trajectory of an optimized route.

The mobile sink then collects the data from sensors at the visited clusters. Simulation results have demonstrated that this scheme can provide not only better energy efficiency as compared to those obtained by conventional methods which assume random waypoint for the mobile sink, but also fault-resilience in case of malfunctions of some sensors due to attacks.

The Mobile sinks use a certain mobility pattern in the sensing area. The novelty of this scheme in comparison with the conventional schemes is that here the mobile sinks could be considered as independent sensors from regular sensor nodes, and therefore they can be recharged and reprogrammed to acquiesce to the updated trajectory. The trajectory of the migration of a sink is assumed random in order to mitigate malicious attacks. Realizing the fact that the conventional random waypoint mobility would not necessarily be energy efficient, in the KAT mobility scheme, the K-means clustering algorithm and the TSP-derived migration route is used for the mobile sinks.

The tradeoff between the throughput and energy consumption is considered as the efficiency metric in the evaluation. Meanwhile, the KAT mobility can calculate the optimal route for the sink to circumvent the damaged area or malfunctioned sensors caused by attacks while still preserving its random behavior, i.e., the mobile sinks move at random speeds so that the arrival timing at each centroid cannot be easily inferred by attackers. Simulation results demonstrated that this scheme can provide better energy efficiency and fault-resilience compared with conventional methods that assume random waypoint model for the mobile sink.

E. SET COVER ALGORITHM

Sasaki.Y., proposed the Set Cover algorithm [3], which focuses on the fairness issue of data gathered by the mobile sinks while also considering the efficiency of data gathering.

A new mobility scheme based on a new clustering method and the Set Cover Algorithm was proposed to ensure that the mobile sinks can gather data from all of the nodes, and simulation results show that fairness of data gathering by this mobility scheme is greatly improved as compared to conventional KAT mobility scheme.

First of all, it is assumed that there are many sensor nodes which can sense environmental data around them in the field. Since these sensor nodes are deployed randomly, some areas are sparse and other areas are dense. Although the KAT mobility scheme aims to gather data efficiently, it exhibits some problems.

This mobility scheme uses a new clustering method, which is based on the communication range of each node, to gather data from all of the sensor nodes. Then, the clusters, which cover all of the sensor nodes with the least number of clusters, are further selected by the set cover algorithm.

Finally, the algorithm solves the traveling salesman problem to determine the path of the mobile sink, and it navigates the mobile sink to the destination along the TSP-path among the cluster heads.

In the clustering algorithm, sensor nodes are partitioned into clusters by forming the set of nodes to which the *i*th node can communicate with as the *i*th cluster. The *i*th node is assigned as the *i*th cluster head since only the *i*th node is guaranteed to communicate with all of the nodes in the *i*th cluster.

Therefore, the cluster head is usually located near the centroid of its cluster. Actually, we we can form these clusters by using the information of the physical communication range, the location of each node, and the maximum hop counts of communication. However, these clusters were formed actually by observing the nodes to which each node can communicate with.

The Set Cover Algorithm (SCA) can solve the SCP (Set Cover Problem), which is one of the oldest and most studied NP-hard problems. The SCA produces the set of clusters covering all of the sensor nodes with the least number of clusters.

There are many different combinations of sets of clusters that can cover all of the sensor nodes. The exact SCA produces a set of clusters covering all of the sensor nodes with the least number of clusters.

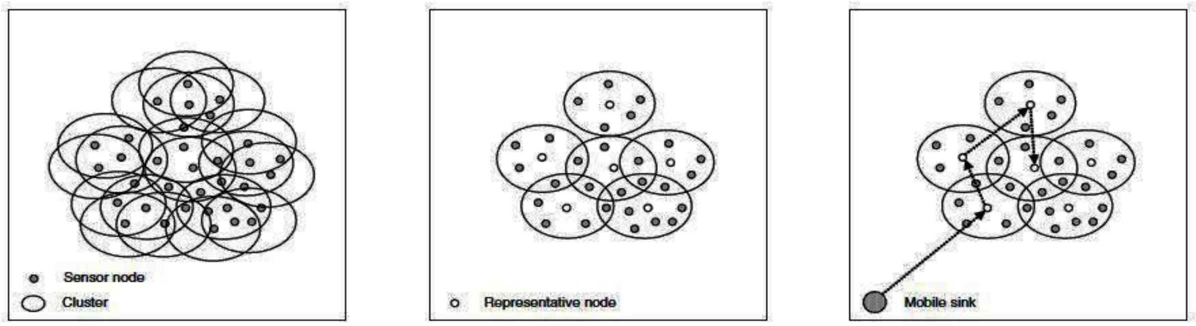
Figure3 depicts the methods involved in data gathering using Set covering algorithm. Fig. 3(a) represents the primary clustering, Fig. 3(b) shows the chosen clusters using the Set cover algorithm and Fig. 3(c) depicts the calculation of the TSP path.

The objective of this mobility scheme is to achieve both higher fairness of gathered data and higher efficiency of data gathering, thus ensuring that the mobile sink can gather data from all of the nodes. Therefore, this scheme can even gather data from the nodes, which are isolated from the neighboring nodes.

Simulation results show that this scheme acquires higher fairness than that of the KAT mobility scheme, and exceeds or achieves the same efficiency as that of the KAT mobility scheme, which has achieved higher efficiency than conventional ones. Therefore, the Set Cover algorithm can gather data from all of the nodes fairly without compromising efficiency.

F. SET PACKING ALGORITHM

The design considerations of the SPAT (Set Packing and TSP) [6], scheme guarantees the complete data gathering from all of the sensor nodes over the sensing field in the target WSAN. In addition, SPAT ensures fairness in terms of data gathering frequency.



(a) There are n clusters. (b) Chosen clusters by the algorithm. (c) The trajectory along the TSP path.

Fig3. The overview of the Set Cover algorithm

This idea is derived from one of our earlier mobility methods called Set Covering Algorithm and TSP (SCAT) [3], which aims at achieving the same objective, but lacks fairness due to inability of the set covering algorithm to guarantee coverage of all sensor nodes without overlapping clusters.

The reason why the SPAT method adopts the set packing algorithm is basically similar to that of adopting the set covering algorithm. The set packing problem is also a NP-hard one.

However, it is capable of reducing duplicated data gathering. In contrast with the set covering algorithm, the output of the set packing algorithm does not produce duplicated sensor nodes. As a consequence, the set packing algorithm does not produce duplicated data. Instead, though some sensor nodes may be beyond the communication range of the mobile sink, the SPAT algorithm is designed to accommodate for few additional clusters for those nodes.

The SPAT scheme functions through four basic steps like clustering, Removal of Redundant clusters, Addition of minimal redundancy clusters and calculating the TSP path for data gathering.

Many clustering techniques can be adopted to classify the given data in WSANs. For instance, in a static sink scheme, LEACH adopts a clustering mechanism, which assigns cluster heads, i.e., representative sensor nodes of respective clusters.

On the other hand, in a mobile sink scheme, KAT mobility adopts a simpler clustering technique referred to as k-means method, without resorting to assignment of cluster heads. In the SPAT scheme, we first extract as a cluster whereby a sensor node can communicate with, and regard this as the cluster head of cluster.

First, the cost between the initial node and each of the other nodes is calculated. Next, the lowest cost node, joins as a member of cluster. The cluster members depend upon various parameters such as adopted protocols of sensor networks and the communication range of a sensor node. From the viewpoint of each cluster head, the mobile sink can communicate with all members belonging to the cluster in question. This clustering applies to every sensor node. Therefore, in this first stage, there are clusters as well as cluster heads. We can see that this clustering is basically based on the Dijkstra's algorithm.

As evident from Fig. 4(a), there are many sensor nodes in the considered WSAN topology that belong to more than a single cluster. These nodes are referred to as the "duplicated" ones. In order to obtain the number of clusters with the smallest number of duplicated nodes, we apply the set packing algorithm as follows. First, the algorithm assigns a score, which equals the number of clusters that the node belongs to, to each node of the network. At this moment, the list of possible clusters contains all of the clusters; i.e., clusters.

This score indicates the number of duplications because it indicates the number of times a node is covered by clusters. Second, the algorithm chooses the node, which has the minimum score, from the list of possible clusters. Third, the algorithm eliminates the clusters, which contain the node already selected in the previous step (Step 2), from the list of possible clusters. Finally, the algorithm repeats these procedures until the list of possible clusters becomes empty.

Continuing from the example depicted in Fig. 4(a), the result at the end of this stage is shown in Fig. 4(b). It is worth noting that there is a possibility that some sensor nodes may still exist in the target WSAN that are not covered by any cluster. The reason behind is that the set packing algorithm only generates the smallest number of clusters without duplicated nodes.

In order to guarantee data gathering from all of the sensor nodes, clusters are added until all of the sensor nodes are covered. Indeed, there are various options to choose the additional clusters. If the fairness is focused in gathering data from the deployed sensor nodes, the additional clusters with the least number of duplicated sensor nodes has to be taken into account.

The number of duplications is defined with respect to a sensor node as the number of clusters to which the sensor node belongs. Indeed, this is a good metric for evaluating the fairness.

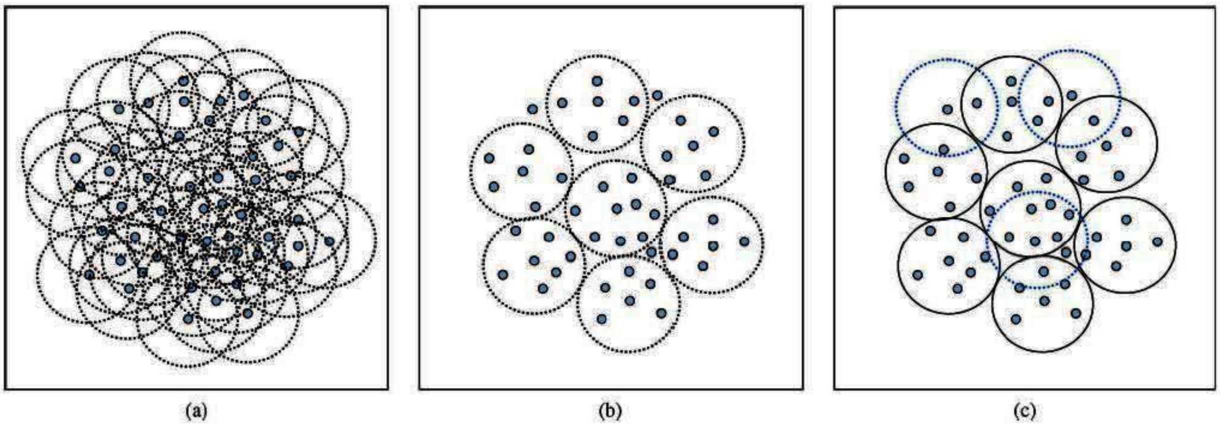


Fig4. SPAT clustering scheme

Eventually, the SPAT mechanism is capable of choosing additional clusters with the least number of duplications, as shown in Fig. 4(c).

In the final stage of this method, the trajectory among the cluster heads of selected clusters is calculated by using the traveling salesman algorithm. Similar to the set packing problem, it is also difficult to find the optimal solution for the traveling salesman problem. After having determined the trajectory, the mobile sinks can move along the calculated path, taking turns to gather data from the sensor nodes.

IV. Conclusion

It is important to enable reliable and efficient data gathering by the mobile sinks in WSNs. The discussed schemes have tried to realize this goal. However, not all the schemes are successful in proving reliable data gathering. This paper has overviewed most commonly used data gathering techniques. Moreover, specific applications require a rethinking of some of the basic paradigms with which communication protocols are engineered. As wireless sensor networks are still a young research field, much activity is still on-going to solve many open issues in reliable data gathering.

REFERENCES

[1] H. Nakayama, N. Ansari, A. Jamalipour, and N. Kato, ”Fault-resilient sensing in wireless sensor networks,” *Comput. Commun. archive*, vol. 30, no. 11–12, pp. 2375–2384, Sep. 2007.

[2] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, “Directed diffusion for wireless sensor networking,” *IEEE/ACM Trans. Networking*, vol. 11, no. 1, pp. 2–16, Feb. 2003.

[3] Y. Sasaki, H. Nakayama, N. Ansari, Y. Nemoto, and N. Kato, “A new data gathering scheme based on set cover algorithm for mobile sinks in wsns,” in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM’08)*, 2008, pp. 725–729.

[4] R. Shah, S. Roy, S. Jain, and W. Brunette, ”Data mules: modeling and analysis of a three-tier architecture for sparse sensor networks,” *Ad Hoc Networks*, vol. 1, no. 2–3, pp. 215–233, Sep. 2003.

[5] H.Zhao, X. Yang and X. Li, “Sink-Trail: A proactive data reporting protocol for Wireless Sensor Networks”, in *IEEE Transactions on Computers*, Issue 99, 2011.

[6] H. Nakayama, Z. Fadlullah, N. Ansari and N. Kato, “ A Novel Scheme for WSAN Sink Mobility based on Clustering and Set Packing Techniques,” *in IEEE Transactions on Automatic Control, Vol.56, no.*10,2011.

[7] X. Fan and Y. Song, “Improvement on leach protocol of wireless sensor network,” in *Proc. Int. Conf. Sensor Technol. Appl. (SENSORCOMM’ 07)*, 2007, pp. 260–264.

[8] V. Hsu, J. Kahn, and K. Pister, “Wireless communications for smart dust,” UC Berkely Electronics Research Laboratory Memorandum M98/2 1998, pp. 168–171.

[9] J. Kahn, R. Katz, and K. Pister, “Next century challenges: Mobile networking for smart dust,” in *Proc. 5th Annu. ACM/IEEE Int. Conf. Mobile Comp. Netw. (Mobicom’99)*, 1999, pp. 271–278.

[10] R. Jain, D. Chiu, and W. Hawe, “A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems,” DEC, Tech. Rep. DEC-TR-301, Sep. 1984.

AUTHORS

Jeba Princess Jafrin.J., is pursuing her M.Tech (CSE) in Karunya University, Coimbatore, India. Her research area is Wireless Sensor Networks.

Anitha Christy Angelin.P., is working as a Assistant Professor in Karunya University. She has completed her M.Tech degree in CSE. She has published various papers in National and International conferences. Her research interests include Wireless Sensor Networks.