A META HEURISTIC ALGORITHM FOR OPTIMAL DATA STORAGE POSITION IN WIRELESS SENSOR NETWORKS

R. Mohanasundaram\textsuperscript{1} and P. S. Periasamy\textsuperscript{2}

\textsuperscript{1}School of Computing Science and Engineering, VIT University, Vellore Tamilnadu, India
mohanasundaramr@vit.ac.in
\textsuperscript{2}Department of ECE, K.S.R. College of Engineering, Tiruchengode Tamilnadu, India
psperiasamy.ps@gmail.com

ABSTRACT

Micro-Electro-Mechanical System (MEMS) is a wireless networking technique in which the background data is monitored and collected using wireless sensor networks. An additional requirement in this technique is to store the data for future retrieval and analysis purpose. This makes data storage become a more fascinating issue in wireless sensor networks. In data storage approach using storage node in the network is to choose suitable positions for storage a node becomes an essential problem. In this paper the above mentioned problem is addressed in wireless sensor network without any of the topology assumptions. Firstly, modeled the problem about data storage positions in theoretical manner and the non efficient heuristic algorithm called Particle Swarm Optimization is proposed to find the suitable positions for k storage nodes while the total energy cost of data transmission is minimized. The algorithm is implemented in a wireless sensor network simulator. The results of the experiments show the feasibility and efficiency of the proposed heuristic algorithm.

Keywords: Heuristic Algorithm, WSN, Micro-Electro-Mechanical System, Particle Swarm Optimization, Data Storage

1. INTRODUCTION

By means of modern advances in MEMS, wireless networking and in embedded technique the wireless sensors are have a permit in the growth of minimum cost with low power, sensor nodes which undergoes multiple operations and they are in small in size and converse release in short distances. These minute sensor nodes consist of sensing, data processing, communicating and controlling the idea of sensor networks based on joint effort of a large number of nodes. Such wireless sensor networks have emerged as a typical \textit{ad hoc} network, which is composed of many sensors (Nakamura \textit{et al.}, 2007, Tang and Gupta \textit{et al.}, 2007). Various areas like health, military and security the sensor network is widely used (Prabh and Abdelzaher 2005). For instance, the physiological data about a patient can be monitored distantly by a doctor, at the same time as this is more convenient for the patient. Also that it allows the doctor to understand the patient’s current condition better. In the air and the water components the sensor networks are used to detect the strange chemical agents purposely. In the applications a large amount of data is continuously collected by sensors and needed to be stored by a data storage strategy for future retrieval and analysis process, which has become one of the important challenges in wireless sensor networks (Wang \textit{et al.}, 2009, Dimokas \textit{et al.}, 2010, Huang \textit{et al.}, 2010).

In wireless sensor networks the data storage approach is significant is described in (Sheng \textit{et al.}, 2010). A bad data storage approach will increase data transmission; depletion of valuable energy (especially for some sensor nodes) and it also reduce the lifespan of all wireless sensor networks. In well-designed data storage approach, data storage position must be carefully considered to minimize the cost of data storage and communication.

In before time the WSN (Wang \textit{et al.}, 2009) sensed the data first and it is stored in sense node itself or transferred to the sink node for storage purpose. The previous approach requires sense node which is comparatively a high storage capacity and hence the sense node is failure at once it will cause the data loss. The latter approach the data transmission is needed, which consumes a lot of energy, especially for some nodes near the sink node. It is very easy to make some nodes run out of energy ahead of time, leading to failure of the entire wireless sensor network.

Due to the rapid advances in sensors, more and more sensing data is collected, the early data storage approaches have been difficult to meet their requirement and some new data storage approaches are well designed to make longer the lifespan of wireless sensor networks. One of the famous approach of them is a new two-tier data storage strategy shown in Figure 1, which the use of some special sensor node, storage node, with much larger permanent storage capability (e.g., flash memory) and more battery power, and like the sink node to store data collected by the sensors in its proximity (Wang \textit{et al.} 2009, Sheng \textit{et al.} 2010).

![Fig 1. A Two-Tier Data Storage Strategy with Four Storage Nodes](image-url)
The most significant issue in the data storage approach is how to gain suitable positions for a limited number of storage nodes in all nodes to make energy efficient, thus extending the lifespan of all wireless sensor networks. Some previous works are carried out on this issue, but most of them found positions with some topology assumptions, such as fixed communication tree.

In this paper, consider a more general case and formulate the storage node position problem. Then propose a heuristic algorithm called Particle Swarm Optimization (PSO) to gain the suitable positions for storage nodes in wireless sensor network based on the energy cost of data transmission. At the same time, some simulated experiments are designed to illustrate the possibility and efficiency of the proposed heuristic algorithm.

2. MODEL AND PROBLEM

Network Model

A two-tier data storage approach for wireless sensor networks is presented in this paper, which compose of three types of nodes. Fig 1 gives the corresponding network model with Fig 1. Sensor node gains sensing data and sends it to the adjacent storage nodes or sink node. Sink node accepts the queries of user, diffuses them to all storage nodes, and aggregates all replies.

![Wireless Sensor Network Model Shown in Figure 1](image)

Shown in Fig 2. There are three types of nodes in this data storage approach, defined as follows:

- **Sink node**: There is only one sink node in the entire wireless sensor network, which accepts user request and returns the suitable reply.
  - **Storage node**: Storage node receives raw data collected by the near sensor nodes and stores them. When sink node diffuses query message, storage node gives a reply by the raw data stored in it. Storage node also collects the sensing data in environments and forwards query message and reply.
  - **Sensor node**: Sensor node gains sensing data and sends them, and it also forwards the raw data gained by near sensor nodes to storage node or other sensor node.

Energy Model

In data storage approach, consider only the energy consumption of sensor node.

For two reasons:
1. Compare with sink node and storage node, sensor node has much less energy and more likely failure because of energy depletion;
2. In the actual wireless sensor networks, there is only one sink node and the number of storage node much less than that of sensor node.

Assume that the total energy consumption of data transmission consists of two parts: sending energy cost and receiving energy cost. For the energy cost of sending data, it is related with the data size and the distance between two nodes. If data is sent from node $i$ to node $j$, the sending energy cost of node $i$ can be computed as follows:

$$E_s = \lambda s d_{ij}$$

where $\lambda$ is a relative weight and $d_{ij}$ is the distance between $i$ and $j$, and $s$ is the data size.

The receiving energy cost includes energy cost of data receiving and storage, and is only related with the data size. In the above example, the receiving energy cost of node $j$ can be represented by

$$E_r = ps$$

where $p$ is a receiving weight, and $s$ is the data size.

**Problem Formulation**

The wireless sensor network can be represented as a graph $G=(V, E)$, where $V$ is the set of nodes, including sink node, storage node, and sensor node. For each sensor node, assume that it generates $n_k$ readings per time unit and the data size of each reading is $s_k$. $E=V \times V$, the set of edges, is the set of wireless communication between a pair of sensor nodes. If the distance from sensor node $i$ to node $j$ is less than the maximum distance that sensor node can send, there is an edge $e \in E$ between node $i$ and $j$.

As the sink node and storage node are very similar in this paper and do not consider energy consumption of the two types of node, so take use of $S$ to represent the sink node and storage node. Therefore total energy cost of the entire wireless sensor network can be calculated by

$$E(S) = \sum_{s \in S} (E_s(s) + E_r(s))$$

**Definition 1 (Data Storage Position Problem):**

Given a wireless sensor network $G$, the data storage position problem is how to find $k$ storage nodes for set $S \subseteq V$ ($|V| = k + 1$), such that the entire wireless sensor network has minimal total energy cost $E$, which is given by (3).

3. **A METAHEURISTIC ALGORITHM FOR DATA STORAGE POSITION**

Assuming the optimal set for data storage position problem is $S$, total energy cost $E$ of the entire wireless sensor network can be derived as follows:

$$E(S) = \sum_{s \in S} (E_s(s) + E_r(s))$$
where \( n(v) \) is the node which accepts data sent from node \( v \), \( s \) represent the data size sent by node \( v \), \( \text{path}(v,u) \) is the optimal path by which node \( v \) sends raw data to one storage node \( u \) of set \( S \).

When the number of storage node is \( k \), \( |\text{path}(v,u)|=k \) is a constant. By formula (4), in order to make the total energy cost \( E(S) \) minimum, obviously \( \sum v \in S \text{es}\sum (a_1 c_1+\rho) \) need to minimize.

Thus, solving the data storage position problem is similar with a \( k \)-median problem.

According to the above derivation, an optimization algorithm is designed to solve the data storage position problem.

**Particle Swarm Optimization**

PSO (Kennedy and Eberhart 1995). Algorithm is forced by the social behavior of a collection of migrating birds trying to reach their destination that is an unknown destination. In PSO, each solution is representing as a ‘bird’ in the flock and is known to as a ‘particle’. A particle is equivalent to a chromosome that is a population member in Genetic Algorithms (GAs) (Tabatabai and Alex 1999). The PSO algorithm does not produce a new birds from parent ones. Instead of that the birds in the population only evolve their social behavior and as a result their movement towards a destination (Eberhart 1998).

A group of birds communicate together when they fly. Each bird appears in a particular direction, and they communicating collectively and recognize the bird that is in the best location. Consequently, each bird speeds in the direction of the best bird through a velocity that is based on its current position. All birds inspect the search space from its new local location, and the process is repeated until the flock arrives at a favored destination. It is to be observed that the procedure comprises both social interaction and intelligence so that birds discover from their own experience called as local search and also from the experience of others around them called as global search.

The process is initiated with a collection of random particles, \( N \). The \( i \)th particle is denoted by its position as a point in \( S \)-dimensional space, where in which \( S \) denotes the number of variables. During the process, each particle \( f \) observes three values namely its current position \( \mathbf{X}_f \), the best position it arrived in previous cycles \( \mathbf{X}_f^* \), its flying velocity \( \mathbf{V}_f \). These three values are denoted as follows:

\[
\begin{align*}
\mathbf{X}_f &= (x_{f1}, x_{f2}, \ldots, x_{fD}) \\
\mathbf{X}_f^* &= (x_{f1}^*, x_{f2}^*, \ldots, x_{fD}^*) \\
\mathbf{V}_f &= (v_{f1}, v_{f2}, \ldots, v_{fD})
\end{align*}
\]

Step 1: Computes a relative weight of edges of \( G \) by following formulas:

\[
\begin{align*}
d &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\
\sigma_{ij} &= \begin{cases} 
1 & (d \leq \delta) \\
\frac{d - \delta}{\alpha} & (d > \delta)
\end{cases}
\end{align*}
\]

In each time interval (cycle), the position \( \mathbf{X}_f \) of the best particle \( g \) is computed as the best fitness of all particles. Thus, each particle updates its velocity \( \mathbf{V}_f \) to get closer to the best particle \( g \), as follows (Chen et al., 2003).

\[
\mathbf{V}_f = w \times \mathbf{V}_f + c_1 \times \text{rand}() \times (\mathbf{X}_f^* - \mathbf{X}_f) + c_2 \times \text{rand}() \times (\mathbf{X}_f^* - \mathbf{X}_g) \quad (2)
\]

As such, using the new velocity \( \mathbf{V}_f \), the particle’s updated position becomes:

\[
\mathbf{X}_f = \mathbf{X}_f + \mathbf{V}_f \quad (3)
\]

where \( c_1 \) and \( c_2 \) represent two positive constants named learning factors (usually \( c_1 = c_2 = 2 \)); \( \text{rand}() \) and \( \text{Rand}() \) denotes two random functions in the range \([0, 1]\). \( V_{\text{max}} \) is an upper limit on the maximum change of particle velocity (Ghandeharizadeh and Shayandeh 2007). \( w \) denotes an inertia weight employed as an enhancement proposed by (Shi and Eberhart 1998) to manage the influence of the previous history of velocities on the current velocity. The \( \omega \) balances the global search and the local search; and it is introduced to minimize linearly with time from a value of 1.4–0.5 (Chae et al., 2002). For itself global search is initiates with a large weight and then decreases with time to favor local search over global search (Turrini and Panzieri 2002).

It is observed that the second term in equation (2) indicates cognition or the private judgment of the particle when comparing its current position to its own best position. The third term in equation (2), denotes the social collaboration between the particles and compares a particle’s current position to that of the best particle (Kennedy and Eberhart 1995). Furthermore, in order to control the change that occur in the particle velocities, upper and lower bounds for velocity change is limited to a user-specified value of \( V_{\text{max}} \). Once the new position of a particle is computed using equation (3), the particle, then, flies towards it (Shi and Eberhart 1998). Therefore, the main parameters used in the PSO are the population size (number of birds); number of generation cycles; the maximum change of a particle velocity \( V_{\text{max}} \) and \( \omega \).

**PSO in Data storage Position**

**Step 1:** Computes a relative weight of edges of \( G \) by following formulas:
where \((x_i, y_i), (x_j, y_j)\) are the coordinates of node \(i\) and node \(j\) in the field and \(d_{ij}\) is the maximum distance which sensor node can send;

Step 2: Computes minimum distance matrix \(D\) of \(G\), where \(d_{ij}\) is the minimum distance between node \(i\) and node \(j\), and minimum distance \(D\) can be optimized by PSO algorithm.

Step 3: Let storage node set \(S = \{p\}\) where \(p\) is the sink node. Computing Set \(S\) by the following code:

```
Begin;
Generate random population of \(N\) solutions(particles);
For each individual \(i \in N\): calculate fitness \((i)\); Initialize the value of the weight factor, \(\alpha\);
For each particle:
Set \(p\) as the best position of particle \(i\);
If fitness \((i)\) is better than \(p\); 
\(p_{\text{Best}}(i) = \text{fitness}(i)\);
End;
Set \(g\) as the best fitness of all particles;
For each particle;
Calculate particle velocity according to Eq. (3);
Update particle position according to Eq. (4);
End;
Update the value of the weight factor, \(\alpha\);
Check if termination=true;
End;
```

The random numbers of nodes are initialized and it is denoted by \(N\) or \(k\). Compute a relative weight of edges of \(G\) by initialize the value of the weight factor \(\alpha\). Then calculate the fitness function for all the nodes. For each distance of the sensor node a best position is determined as \(p_{\text{Best}}\). If fitness \((i)\) is better than \(p_{\text{Best}}\), function is satisfied and the execution ends. The data is stored in the nodes. If the condition is not satisfied assign an energy cost to the fitness value then calculate the temporary energy cost for each node. The calculated temporary energy cost and the energy cost is satisfies each other the algorithm ends and the data is stored in the nodes are successfully stored for future retrieval process.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of PSO algorithm, a wireless sensor network is implemented in simulator to carry out some experiments for the data storage position. In the simulator, sensor node and storage node are randomly deployed in a \(100 \times 100\) square field, and the sink node is in the center. Some other parameters are defined as next: \(n=20\), \(S_2 = 3\), \(\alpha = 1.0\), \(\beta = 0.1\), \(\delta = 25\).

There are two experiments implemented in simulator, the first is to construct communication model for wireless sensor network with the proposed heuristic algorithm. The total node number in all wireless sensor networks is 50. Figure 3 and Figure 4 respectively give the corresponding communication models when storage node number are 0 and 2. In Figure 3, only the sink node can store data, and the network model is built into a tree. In Figure 4, the sink node and storage nodes can store data, so three trees are built.

There are two experiments implemented in simulator, the first is to construct communication model for wireless sensor network with proposed heuristic algorithm. The total node number in all wireless sensor networks is 50. Figure 3 gives the corresponding communication models when storage node 2. In Fig 3 the sink node and storage nodes can store data, so three trees are built.

It can observe that compared with other two algorithms, the curve of random algorithm less smooth, and the performance of the proposed algorithm is very close to that of the optimal algorithm, especially when the storage node number is small. From the figure it can be seen that the proposed algorithm is more efficient compared with other algorithms.

In the second experiment, compare the heuristic algorithm two other algorithms:

(1) Random algorithm. \(k\) storage nodes \(10 \times k\) times are randomly selected, and the set with the smallest energy cost is regarded as the set \(S\);

(2) Heuristic algorithm: This algorithm computes the minimum distance matrix through heuristic approach such as Floyd algorithm or Dijkstra algorithm.

(3) Optimal Algorithm: The optimal storage node set \(S\) is the storage node set with minimum energy cost in all possible \(k\) nodes set in the wireless sensor network.
The proposed research work uses Metaheuristic nature based optimization algorithm using PSO which is compared with the above mentioned algorithms.

Fig 5. shows the energy cost with different storage node number, and the total node number in all wireless sensor network is 100. First calculate a maximum energy cost $E_{max}$, which is the energy cost when there is no storage node ($k = 0$) and every sensor sends data to the sink by the communication tree. The energy cost in Fig 5. is given as a ratio relative to $E_{max}$. It can observe that compared with other two algorithms, the curve of random algorithm less smooth, and the performance of the proposed heuristic algorithm is very close to that of the optimal algorithm, especially when the storage node number is small. The difference between the proposed algorithm and optimal begins to increase when the storage node number becomes larger. Thus, the simulated experiments show the feasibility and efficiency of the proposed heuristic algorithm.

**Data Reduction and Energy Cost**

Figure 6 illustrates the impact of data reduction rate to the energy cost. This time, data reduction, $\alpha$ becomes an important parameter, because every storage node is in charge of many forwarding nodes. A small decrease of $\alpha$ will reduce energy cost greatly. It is observed that the proposed PSO based metaheuristic approach provides better performance when compared with the other approaches taken for consideration.

**Network Lifetime**

Fig 7. shows the network life time with the different number of storage nodes. It is observed that the proposed PSO based metaheuristic algorithm greatly influences the life time of the entire network. The life time of the network is prolonged by the application of the proposed PSO based metaheuristic based algorithm. The metaheuristic approach lengthens the lifetime a lot with a small number of storage nodes, although the objective of our algorithm is to minimize the total energy cost. For example, with 10 storage nodes, the lifetime is increased significantly.

5. CONCLUSION

Data storage has become an important issue in sensor networks as a large amount of collected data need to be archived for future information retrieval. This paper considers the storage node placement problem aiming to minimize the total energy cost for gathering data to the storage nodes. This paper introduces a metaheuristic nature inspired optimization algorithm to address the data storage problem in wireless sensor network. In this data storage strategy, storage node has been introduced to reduce the transmission of raw data, and to prolong the lifespan of all wireless sensor network. Formulate the storage node position problem and propose a heuristic algorithm for it and implemented the algorithm. Hence, the experimental results show that the proposed algorithm is more efficient and feasible.

REFERENCES


