

Gathering Correlated Data in Wireless Sensor Networks Using a Heuristic Algorithm

Yuexian Wang, Cheng Chew Lim

School of Electrical and Electronic Engineering, the University of Adelaide, Adelaide, Australia, {jwang, cclim}@eleceng.adelaide.edu.au

Abstract. We propose a routing scheme for data gathering and aggregation in wireless sensor networks. The scheme aims to optimise an aggregation tree in order to minimise the energy dissipation of data aggregation and transmission. A modified particle swarm optimisation algorithm is developed in the proposed scheme. In addition, the routing scheme uses a generic data aggregation model which accommodates different correlation conditions. The performance of the proposed scheme is evaluated and compared with three other existing data gathering algorithms. Simulation results show that the proposed scheme outperforms existing algorithms in terms of energy consumption and that the scheme can adapt to the change of network connectivity and data correlation conditions.

Keywords: Sensor networks, data aggregation, particle swarm optimisation, correlation coefficient, energy consumption

1. Introduction

In wireless sensor networks, a large number of sensor nodes are scattered in an observed region and left unattended after deployment, and all nodes in the network are thus energy-constrained. Energy consumption, scalability, and load balancing are important requirements for many data gathering sensor network applications. The tree-based data gathering algorithm is proposed for achieving development in this area, such as the shortest path tree algorithm [1], the minimum spanning tree algorithm [2] and the shallow light tree algorithm [3]. In a tree-based network, sensor nodes are organised into a tree structure where data aggregation is performed at intermediate nodes along the tree and a concise representation of the data is transmitted to the root node. Therefore, the research problem is to find out a routing tree with data aggregation to achieve minimum energy expense. This problem is an NP-complete problem [4,] because it can be reducible to a weighted set cover problem in graph theory, which has been shown to be NP complete [5].

In this paper, we propose a routing scheme for data gathering and aggregation in wireless sensor networks. The scheme utilises a heuristic algorithm to optimise data traffic and the transmission structure in terms of energy consumption. Using this heuristic algorithm, significant data traffic can be reduced since more reasonable aggregator nodes to eliminate data redundancy can be selected. In order to reduce the energy consumption for data transmission, the heuristic algorithm optimises transmission structure as well. In addition, a generic aggregation model which does not depend on any specific relationship among correlated data is exploited in order to adapt to a variety of applications.

The remainder of this paper is organised as follows: Section 2 gives the introduction to the standard particle swarm optimisation algorithm. The system models of our new scheme are presented in Section 3. The details of the proposed modified particle swarm optimisation algorithm for the routing scheme are described in Section 4. Section 5 discusses the performance evaluation and simulation results and Section 6 concludes the paper.

2. Standard Particle Swarm Optimisation Algorithm

The particle swarm optimisation (PSO) algorithm stems from the simulation of a simplified society model, and it was proposed as a typical heuristic algorithm inspired by the behaviour of bird flocking [6]. The PSO was used to search optimal or near-optimal solutions in large search space. Assume that in a D -dimensional searching space m particles compose a swarm and fly with a certain velocity, where the particle i denotes a D -dimensional vector $x_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$, $i=1, 2, \dots, m$. The performance of x_i is assessed by its fitness value which is calculated through substituting x_i into the objective function. The motion of particles changes according to the following equations:

$$v_{id}^{t+1} = \omega v_{id}^t + c_1 \text{rand}() (p_{ibest}^t - x_{id}^t) + c_2 \text{rand}() (p_{gbest}^t - x_{id}^t), \quad (1)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1}, \quad (2)$$

Where t is the iteration number, v_{id} denotes the velocity of the particle i in the d th dimensional ($1 \leq d \leq D$) particle swarm space, x_{id} represents the particle i current position, p_{ibest} is its best previous position, and p_{gbest} is the best position among all particles in the population. $\text{rand}()$ is a random function with a range $[0, 1]$. c_1 and c_2 are learning factors used to accelerate the motion speed of particles. The inertia weight, ω is a user-specified parameter that controls,

with c_1 and c_2 , the impact of previous historical values of particle velocities on its current one. A large inertia weight facilitates global exploration while a small inertia weight facilitates fine-tuning local search.

3. System Models

A.Network Model. One base station and n sensors are assumed to be distributed uniformly in a field and quasi-stationary. The base station sends a query and k ($k \leq n$) sensors respond to that query. Wireless links are considered bidirectional and symmetric so that any two nodes can communicate using the same transmission power levels only if they are within range of each other. Sensor nodes are homogenous and arbitrarily allocated with equal initial energy. In addition, nodes are not location-aware and left unattended after deployment.

B.Data Correlation and Aggregation Model. We assume that there is correlation among the data generated by the k sensors. To possibly accommodate a wide range of scenarios, we abstract data redundancy among two sensor nodes using a single value ρ , termed the correlation coefficient. Assume that data amount after aggregation is not less than any of its inputs and not more than the sum of all inputs. Let R_{v_i} and R_{v_j} be the amount of data generated by the node v_i and v_j in response to a query. Without loss of generality, if $R_{v_i} \geq R_{v_j}$, the output $R(v_i, v_j)$ after data aggregation is:

$$R(v_i, v_j) = R_{v_i} + (1 - \rho)R_{v_j}, \quad (3)$$

C.Energy Model. We adopt both the free-space propagation model to approximate path loss sustained because of wireless channel transmission [7]. The energy expended by the radio for transmitting and receiving an l bit message over a distance d , is given by:

$$E_{Tx}(l, d) = lE_{elec} + l\epsilon_{fs}d^2, \quad (4)$$

$$E_{Rx}(l) = lE_{elec} \quad (5)$$

Where E_{Tx} is the energy dissipated in the transmitter, E_{elec} is the per bit energy dissipation for running the transceiver circuitry, ϵ_{fs} is the amplifier parameter for the free space propagation model, and E_{Rx} is the radio expends of receiver.

4. Problem Formulation

Given the set of source node S and the base station d in wireless sensor networks $G(V, E)$, our objective is to find a connected subgraph $G'(V', E') \subseteq G$, where $S \subset V'$, and $d \in V'$, in order to minimise the energy consumption for transmitting data from all source nodes to the base station. The objective function is formulated as follows:

$$\min \sum_{v_i \in V'} R_{v_i} \omega(v_i, d), \quad (6)$$

Where $\omega(v_i, d)$ is the total weight of edges on path from node v_i to base station on constructed aggregation tree. In wireless networks, the weight of each edge can be considered as the energy expended for per unit data communication and data aggregation. R_{v_i} is the amount of data generated by node v_i .

5. PSO Modified by Genetic Algorithm for Routing with Aggregation

The standard PSO algorithm is the real valued PSO, and it cannot operate addition or subtraction directly on the path. Hence, the standard PSO algorithm should be extended to deal with the discrete optimisation problems which require the ordering or arranging of discrete elements, eg. the routing tree with aggregation problem. Facing the problems, we propose a PSO algorithm which is modified by the genetic algorithm [8] in order to address the discrete nature of our optimisation problem. The core of this algorithm is an equivalent form of velocity and displacement formulae combining the thought of genetic algorithm. Hence, Eq. 1 and Eq. 2 are replaced by Eq. 7 and Eq. 8.

$$v_{id}^{t+1} = \omega v_{id}^t \oplus (p_{ibest}^t \otimes x_{id}^t) \oplus (p_{gbest}^t \otimes x_{id}^t), \quad (7)$$

$$x_{id}^{t+1} = x_{id}^t \oplus v_{id}^{t+1}, \quad (8)$$

Where the operator \oplus represents a composition, and the operator \otimes indicates information exchange which is implemented by crossover.

A.Encoding. Encoding involves coding paths serial numbers into a feasible solution (or a position) in the search space. We represent the individual, for a specific aggregation tree, as a string of node numbers. The length of each individual is always equal to the number of relay nodes. A routing scheme for a network with 7 relay nodes, and one base station, is shown in Fig. 1(a) and the corresponding particle is shown in Fig. 1(b).

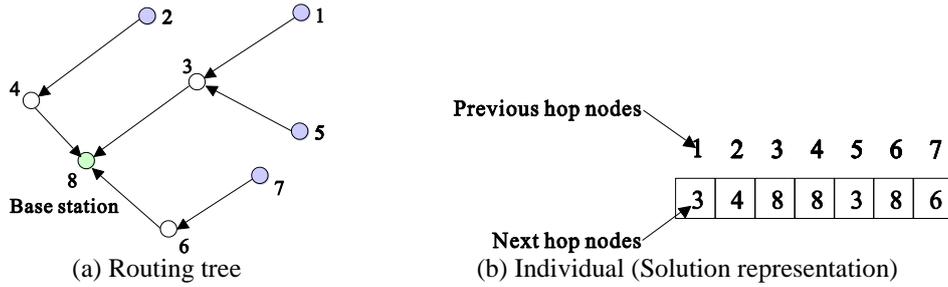


Fig.1 Encoding a routing tree as an individual

B.Population Initialisation. In general, there are two ways to generate the initial population, heuristic initialisation and random initialisation. For most large scale problems, such as network communication design, random initialisation has better effect on global optimal solutions. The main idea of our work is that the correspondence of nodes behaviour through local optimisation is greater than the correspondence of nodes behaviour in random initialisation. The SPT algorithm and the MST algorithm have good results in the initial population, when correlation coefficient is equal to 0 and 1 respectively. Hence the initial population should include the shortest path tree and the minimum spanning tree.

C.Fitness Function. We define the fitness value as the energy consumption of the network, and our metric for total energy dissipation takes into consideration including the transmit energy, the receive energy and aggregation energy.

D.Selection. The function of selection operator is to select individuals which have relative large fitness values with relative large probabilities from their parent generation. The PSO algorithm then puts the chosen individuals in the offspring generation and waits for further evolution implemented by crossover and mutation. We adopt the ranking selection [9] as a part of implementation of the modified PSO algorithm in this paper.

E.Crossover. The operator \otimes in Eq. 7 indicates information exchange which is similar to the operation of crossover in the genetic algorithm, so we use the crossover as the equivalent form of the \otimes operation. Because of the requirement of encoding, source nodes are known at the stage of crossover. First, we randomly select a gene in the locus of the same source node between the optimal individual and a generic individual and put it into the same locus of the offspring. We then make the same selection at the locus of the relay node which is indicated by the previous chosen gene. This process is repeated until the base station is found. For the rest nodes which are not used, genes from the same locus between the two individuals are selected randomly. Fig. 2 is an example for the crossover operation.

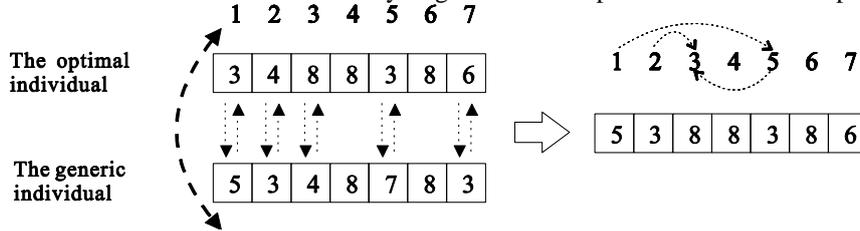


Fig.2 Crossover between the optimal individual and the generic individual.

F.Mutation. Multiplying ω by v_{id}^k in Eq. 7, indicating that a particle searches toward a new space, is similar to the operation of mutation in the genetic algorithm, so we use the mutation as the equivalent form of ωv_{id}^k . We select a node 3 to substitute for node 6 from set Ω where the node 3 has the same previous hop node 7 and next hop node 8 as the node 6. This process is illustrated by Fig. 3.

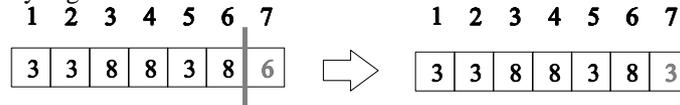


Fig.3 Examples of the mutation operation.

G.Addition. The addition operation indicates the sum of several processing steps. Through orderly executing the equivalent subitems in the PSO algorithm, we can obtain the effect of addition. In the equivalence of the addition, the outputs of previous steps are the inputs of latter steps. According to Eq. 7 and Eq. 8, the computation sequence is from right to left. Applying two crossover operations and one mutation operation to a generic aggregation tree, we can acquire an offspring and then put it in the population for the next iteration.

6. Simulation Results and Discussions

In our simulation, average energy dissipation and the maximum run (the first node dies) are metrics used to compare the PSO algorithm with three existing tree-based data gathering algorithms: the shortest path tree (SPT) algorithm, the minimum spanning tree (MST) algorithm, and the shallow light tree (SLT) algorithm.

A square region of size 50m×50m is uniformly divided into 100 grids, and each sensor is deployed in one grid. Since the coordinate of each sensor is determined by a random function which can generate uniformly distributed random numbers within the interval [0, 50], the sensor nodes are randomly distributed. Each node is equipped with 1J initial energy at the beginning of the simulation. Every node transmits a 4000 bit message per round to the base station.

The calculation of energy dissipation in the simulations is based on Eq. 4 and Eq. 5. We use the same parameters as that in [10]: $E_{elec}=50nJ/bit$ and $\varepsilon_{fs}=10pJ/bit/m^2$. Moreover, the per bit energy dissipation for aggregating data is set as $E_{da}=15nJ/bit/signal$ and the transmission range is set as $R=20m$.

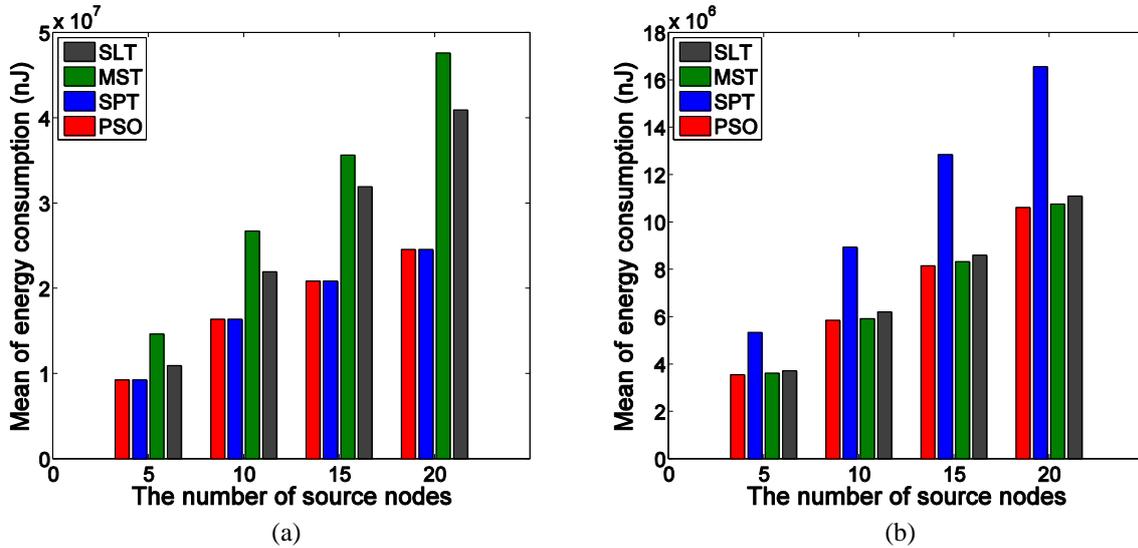


Fig.4 Influence of the number of source nodes on (a) mean of energy consumption versus the number when ρ approaches 0 and (b) mean of energy consumption versus the number of source nodes when ρ approaches 1

Fig.4 (a) shows the mean of energy consumption per round versus the number of source nodes using an aggregation tree with the PSO algorithm, the MST algorithm, the SPT algorithm, and the SLT algorithm. Clearly, when ρ approaches 0 the PSO algorithm performs as well as the SPT algorithm and reduces energy consumption significantly compared with the MST algorithm and the SLT algorithm. A reduction of average energy consumption of about 40% and 50% can be obtained by the PSO algorithm as well as the SPT algorithm over the SLT algorithm and the MST algorithm, respectively. In a network with poor correlation, the PSO algorithm reduces to the SPT algorithm, and data should be transmitted directly to the base station via the shortest paths instead of aggregator nodes by detouring, since there is no redundancy among the data, and aggregation at any relay node is not efficient in reducing the data amount.

When ρ approaches 1 as illustrated in Fig.4 (b), the PSO algorithm performs better than all other algorithms with regard to the mean of energy consumption. A reduction of average energy consumption of about 5%, 5%, and 40% can be obtained by the PSO algorithm over the MST algorithm, the SLT algorithm, and the SPT algorithm respectively. The performance of the MST algorithm is the closest to that of the PSO algorithm. However, since some nodes within one hop to the base station would waste energy for detouring to aggregator nodes rather than transmitting directly to the base station, the MST algorithm cannot achieve the optimal performance. Because the nodes can not sufficiently take advantage of data aggregation to reduce the transmission task, the SPT algorithm expends the most energy compared with the three other algorithms. The SLT algorithm can balance between data aggregation and direct transmission, but it gets the benefit implicitly. Hence the energy consumption of the PSO algorithm increases more slowly than the SLT algorithm with the increase in the number of source nodes.

Fig.5(a) illustrates the average energy consumption of the four algorithms when the number of source node $k=15$. The costs of all algorithms decrease with the increase in ρ , the correlation coefficient. This exemplifies that data aggregation in sensor networks can greatly benefit the routing performance by reducing redundancy among correlated data. In addition, the PSO algorithm evidently outperforms other algorithms in the energy consumption with the increase in the correlation coefficient. A reduction of average energy consumption of about 25%, 30%, and 35% can be obtained by the PSO algorithm over the SLT algorithm, the MST algorithm, and the SPT algorithm, respectively. When ρ is small, the SPT algorithm performs well. However, it does not benefit from the increase in ρ as it cannot efficiently perform data aggregation to eliminate redundancy among data. In the contrast, the MST algorithm gets the worst performance when ρ is small since it pursues data aggregation but data are actually low correlated. The PSO algorithm performs much better than the SLT algorithm since the PSO algorithm recalculates total energy dissipation in every stage to get perfect matching and thus can adapt to the correlation among nodes.

We can see from Fig. 5(b) that the maximum run of all algorithms grows with the increase in ρ . The main reason is that data aggregation can significantly reduce traffic by eliminating redundancy among correlated data and hence consume less energy. Since aggregator nodes receive data from different nodes and perform data aggregation, they always die earlier than others due to the heavy load. When $\rho < 0.8$, the SPT algorithm outperforms other algorithms markedly because data are opportunistically aggregated and the amount of data for aggregation is not large. The SPT algorithm does not benefit well when ρ increases from 0.8 to 1 due to insufficient data aggregation. Because the MST algorithm and the SLT algorithm prefer to detour to pursue data aggregation, large amount of data may be aggregated on nodes near the sources, and these aggregator nodes will exhaust energy earlier than that in the PSO algorithm. In addition, the maximum run of the PSO algorithms drops dramatically when ρ varies from 0 to 0.2 because the PSO

algorithm can find out more aggregator nodes to share the same paths and the load of aggregator nodes becomes heavier.

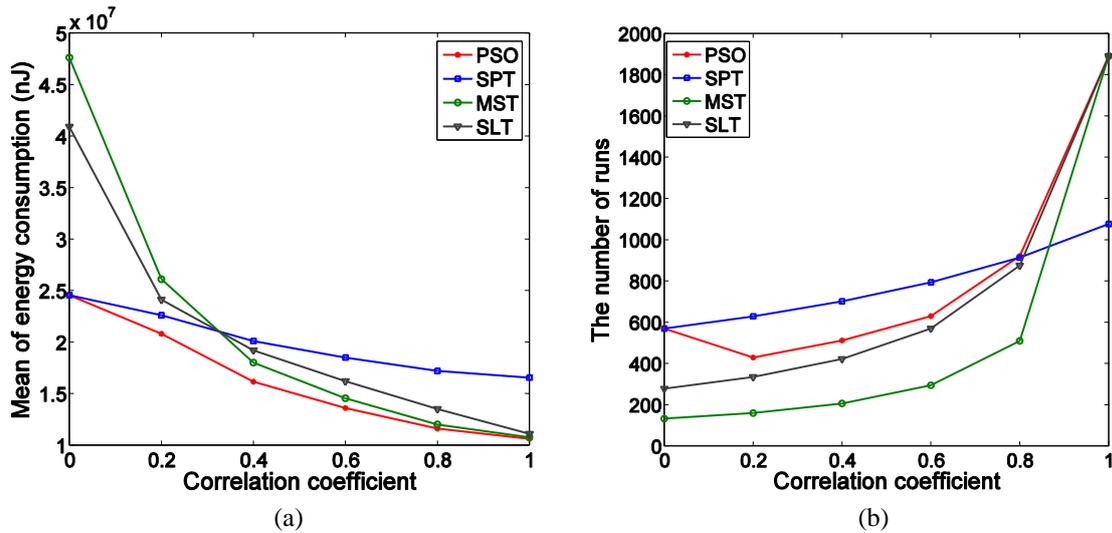


Fig.5 Influence of correlation coefficient on (a) mean of energy consumption versus the correlation coefficient when $k=15$ and (b) The number of run versus the correlation coefficient when $k=15$

7. Conclusions

We propose an energy-efficient data gathering and aggregation routing scheme that caters for the energy challenges in wireless sensor networks. The particle swarm optimisation algorithm is utilised to optimise a joint objective between data traffic and transmission structure. The scheme also adopts a generic data aggregation model which accommodates different correlation conditions. Simulation results show that the particle swarm optimisation algorithm outperforms its counterparts in terms of energy consumption. Although there is a small sacrifice in the maximum run due to the heavy load of aggregator nodes, the particle swarm optimisation algorithm can be regarded as a good tradeoff since it can save energy and adapt to the change of network connectivity and data correlation conditions.

References

- [1] W. Heinzelman, J. Kulik, and H. Balakrishnan, "Adaptive protocols for information dissemination in wireless sensor networks," in Proc. ACM MobiCom, Aug. 1999, pp. 174-185.
- [2] R. C. Prim, "Shortest connection networks and some generalization," Bell System Technical, vol. 36, no. 6, pp. 1389-1401, 1957.
- [3] A. Goel and K. Munagala, "Balancing steiner trees and shortest path trees online," in Proc. 11th ACM-SIAM Symp. on Discrete Algorithms, Jan. 2000, pp. 562-563.
- [4] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in Proc. IEEE INFOCOM, vol. 4, Mar. 2004, pp. 2571-2582.
- [5] M. R. Garey and D. Johnson, Computers and intractability: A guide to the theory of NP-Completeness. W. H. Freeman, 1979.
- [6] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization," in the IEEE Conf. on Neural Networks, IV, 1995, pp. 1942-1948.
- [7] H. Xia, H. L. Bertoni, and L. R. Maciel, "Radio propagation characteristics for line-of-sight microcellular and personal communications," Antenna and Propagation, vol. 41, no. 10, pp. 1439-1447, Oct. 1993.
- [8] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, 3rd ed., Springer-Verlag, 1996.
- [9] J. E. Baker, "Adaptive selection methods for genetic algorithms," in Proc. the 1st International Conference on Genetic Algorithms, 1985, pp. 101-111.
- [10] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," Wireless Communications, vol. 1, no. 4, pp. 660-670, October 2002.