A Game-theory Based Clustering Approach for Wireless Sensor Networks

Zhanyang Xu[1,2], Yue Yin[1] , Xiaotian Chen[2], and Jin Wang[1]

[1]Computer and Software School, Nanjing University of Information Science & Technology, Nanjing 210044, China
[2]Institute of Information Network Technology, Nanjing University of Posts and Telecommunications, Nanjing 210003, China.

Abstract. Clustering has long been regarded as an efficient technique for data collection in wireless sensor networks. The approach divides network into clusters and makes cluster heads (CHs) responsible for data aggregation. However, due to the constraints on available resources, a sensor node is likely to be selfish and refuse to declare itself as a cluster head. In this paper, we propose a game-theory based clustering approach for wireless sensor networks. A game-theoretic model is built for CH selection. This paper adopts data replication to reduce possible network disconnection. The selection of a candidate CH is discussed under a second price sealed auction. Simulation results show that the throughput of the sink can still be guaranteed if any CH fails to work.

Keywords: wireless sensor networks, clustering, game theory, data replication

1 Introduction

Clustering is an efficient data collection method in Wireless sensor networks (WSNs) [1], where the entire network is divided into multiple clusters. Each cluster has one cluster head (CH) and it is responsible for data aggregation. Instead of direct communication with the sink, all the member nodes in one cluster send data to the cluster head. In this way, the traffic load can be reduced.

In general, we usually assumed that all nodes are cooperative and are willing to provide network services such as relaying data for others. However, this assumption is not strictly true in WSNs. Some rational nodes turn to decline the requirement of relaying data for other nodes. Such nodes are noted as the selfish nodes. As cluster heads usually shoulder heavier traffic burden compared with normal sensor nodes, a selfish node may refuse to declare itself as a cluster head. As the battery, capability of computing, storage and data processing of a sensor are limited, how to ensure the throughput in case any network disconnection occurs is one of the key problems. To solve the problem, game theory [2] is introduced, which studies the conflict and cooperation between intelligent rational decision-makers. The selfishness of sensor nodes can be restricted by game theory.

The communication link between a CH and the sink is of vital importance. Any disconnection of these links at certain point could result in a loss of all sensed data in its cluster region. A possible solution is to adopt data application [3]. The idea is to keep copies of data in more nodes, so that if any failure occurs to the node that owns the original data, its information is not lost and can be retrieved through its copies. Here, we simply replicate the data in a CH to another node. So even network division occurs and separate the connection from CH to the sink, we can still have another node to send the cluster's data.

The rest of the paper is organized as follows. Related work is discussed in section 2. In Section 3 we address CH election on the basis of game theory and adopt data replication in case of possible disconnection between an original CH and the sink. Performance evaluation is given in Section 4 and Section 5 concludes this paper.

2 Related Work

Many energy-efficient routing algorithms have been proposed based on the hierarchical topology. Clustering has the advantages of low energy consumption, simple routing scheme and good scalability, and it is especially suitable for WSNs. Various well-known clustering approaches have been proposed, such as LEACH[4] and PEGASIS[5]. Ref. [6] aims to solute the energy hole problem. For the clusters, the closer they are to the sink, the smaller size they are formed. However, many assumptions do exist in such approaches. For example, nodes should have much information about other nodes, which is not practical in reality. Besides, most clustering protocols do not consider the selfishness of nodes.

Selfishness in wireless networks is a popular study. Incentive mechanisms have been proposed. In Tit-For-Tat (TFT) [7], the player cooperates on the first stage and does what its opponent did in the previous stage. However, a perceived defection may be unjustly punished due to packet collisions. GTFT [8] improves TFT by providing a tolerance threshold. Limited number of defections will not be punished. Ref. [9] studies the impact of packet collisions on the emergence of cooperation and proposed two schemes called OT and GT for milder conditions. They are theoretically effective, but practically unstable.

Data replication is very effective for preventing deterioration of data accessibility due to network division in wireless networks. Ref. [10] proposes data replication schemes in ad hoc networks. These schemes are based on the intuition that to improve data accessibility, replicating the same data near neighboring nodes should be avoided. In SAF, the access frequency for certain data is the major concern to decide which node should get the replica. DAFN pays extra attention to its connectivity with the neighbors, and DCG sets nodes into groups for later discussion. A later study [11] extends the above methods by considering a more real environment with periodic date update. In Ref. [12], each node belongs to certain cluster in which the probability of path availability can be bounded. Nodes exchange information with stable neighbors. Ref. [13] proposes some new schemes for data replication: Greedy-S considers both the size and access frequency of data; OTOO adopts a metric of combined access frequency which is related to the node and a neighbor of its own; RG sets groups for nodes that can share replica. Unlike the study of Hara, link failure probability and query delay are taken into consideration. Ref. [14] selects nodes as data replicas holders taking into account link bandwidth and remaining amount of batteries. Various parameters demand specific future study though.

3 Our Proposed Game-theory Based Approach

3.1 Energy model

We use the same energy model in [15]. Based on the distance between transmitter and receiver, a free space (d^2 power loss) or multi-path fading (d^4 power loss) channel models are used.

Each sensor node will consume the following E_{Tx} amount of energy to transmit a *l*-bits packet over distance *d*, where the E_{elec} is the energy dissipated per bit to run the transmitter or receiver circuit, ε_{fs} and ε_{mp} represent the transmitter amplifier's efficiency and channel conditions:

$$E_{Tx}(l,d) = \begin{cases} lE_{elec}+l\varepsilon_{fs}d^2, d<d_o \\ lE_{elec}+l\varepsilon_{mp}d^4, d\geq d_o \end{cases} \tag{1}$$

To receive a packet, radio consumes energy:

$$E_{Rx}(l) = lE_{elec} \tag{2}$$

3.2 Selection of Cluster Heads

We assume that the network is composed of sensor nodes. They are uniformly dispersed within a circle field and continuously monitor their surrounding environment. In our study, the entire network is divided into K equal clusters, as is shown in Fig.1 where $K = 5$. Each cluster has one cluster head for data aggregation. Instead of direct communication with the sink, each member node in one cluster sends data to its CH. Each CH receives the delivered data, makes aggregation and finally sends data to the sink far away. Such clustering method reduces the traffic load. Moreover, CHs locate in a more uniform way than the probabilistic deployed situation in LEACH. It prolongs the network lifetime and reduces the energy hole problem.

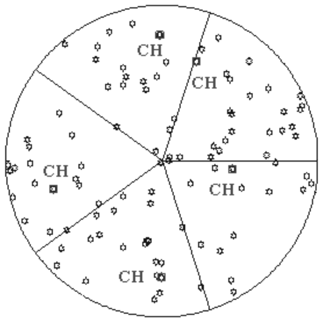

Fig.1. An example of CH selection

The selection of CHs is important. If a CH runs out of energy, all collected data in its cluster get lost and can no longer reach the sink. Therefore, we have the residual energy of a node stand out as a metric. Any node with the maximum residual energy in a cluster is chosen as a CH. Thus with CH roles change periodically, the network can survive for long time. However, selfish nodes may tell lies about the value of its residual energy to avoid being selected. Hence the CH selection would fail to work. To solve the problem, we regard the CH declaration as a game and adopt a game-theoretic model to promote cooperation of selfish nodes.

We formally define the game as $G = < N, S, U >$, where $N \ (|N| = n)$ is the set of players, representing sensor nodes in the network; $S = \{S_i\}$ is the set of available strategies; $U = \{U_i\}$ is the set of utility functions.

Players can either declare itself as a CH or stay selfish by refusing to be the CH. Letting D be the strategy "declare myself as CH" and RD be the strategy "refuse to declare as CH", the strategy space is $S = \{Declare,\ Refuse\ to\ Declare\} = \{D, RD\}$.

We define c_D and c_{RD} respectively represent the cost of the node when it declares itself as CH and the refusing situation, shown in Equation (3) and (4) respectively. Here, n_{CH_i} stands for the number of nodes in cluster CH_i.

$$c_D = n_{CH_i} E_{Rx} + E_{aggr} + E_{Tx(CH_i, \sin k)}$$

(3)

$$c_{RD} = E_{Tx(s_i, CH_i)} \tag{4}$$

Role of cluster heads changes periodically. In the long term, c_D and c_{RD} can be regarded as constants for simplicity. Moreover, as the sink locates far away from the sensing region, the cost for delivering data to the sink is much larger than that to its CH according to the energy model, namely $c_D > c_{RD}$. In case nodes are reluctant to declare as CH, a payoff v is provided. From the perspective of an arbitrary node, as the equation (5) shows, if one node i declares, the utility is $v - c_D$; if node i refuses to be CH and luckily one other nodes in its cluster takes the responsibility by declaring itself as CH, the utility of i becomes $v - c_{RD}$; however, the worst condition is that neither the node itself nor any other node declares as CH, therefore the player will be unable to send data towards the sink which leads to zero payoff in result.

$$U_i(\mathrm{S}) = \begin{cases} v - c_D & if \;\; S_i = D \\ v - c_{RD} & if \;\; S_i = RD \;\; and \;\; \exists j \in N, \;\; s.t. \;\; S_j \in D \\ 0 & if \;\; S_j = RD, \;\; \forall j \in N \end{cases} \tag{5}$$

According to the assumption, all nodes are homogeneous. So it is impossible for each node to find a best response to the strategy choices of its opponents. Namely no pure-strategy Nash Equilibrium exists in our game. However, if we assume that each player is allowed to choose its strategy randomly following a probability distribution, a mixed-strategy Nash Equilibrium can be found.

For each node, the possibility of declaring itself as CH (i.e. playing D) is set as p, and the probability of refusing to declare (i.e. playing RD) is $1 - p$. The expected utility function of playing D is obtained as $U_D = v - c_D$. The expected utility function of playing RD is obtained as $U_{RD} = (v - c_{RD}) \cdot (1 - (1 - p)^{N-1})$, which implies at least one other node plays D.

At the equilibrium, we have $U_D = U_{RD}$. By solving the expression, we have

$$p = 1 - \left(\frac{c_D - c_{RD}}{v - c_{RD}}\right)^{\frac{1}{N-1}} \tag{6}$$

3.3 Selection of the Candidate CHs

Cluster heads play a significant role during data collection approach. Therefore, once part of the network becomes unstable and by chance it causes a disconnection of the link between certain CH and the sink, all data in the cluster would be lost. To solve this problem, we adopt data replication and set another sensor node despite the current CH as a candidate CH. This candidate node replicates CH's data. Thus when the sink fails to communicate with the CH at one point, it can still get a replica of data from another node. Robustness of the network is improved due to such data replication. To encourage all member nodes to complete for the role of a candidate CH, an extra payoff is provided if it has direct communication to the sink.

In one cluster, we adopt a game-theoretic method to select a candidate CH. All member nodes despite the CH are players. They have the desire of turning into a candidate CH in order to win the possible payoff. We assume that every player has its own valuation, and they bid against each other to win the game. In a wireless sensor network, we assume that all data collected by sensor nodes is periodically updated. Accordingly, their corresponding CHs update data.

Let R_k^i represent the cost for i to replicate data in its CH. It is related to both its data access frequency r_k^i and a transmission cost to its CH that is denoted as c_k^i. We have

$$R_k^i = r_k^i \cdot \mathrm{c}_k^i \tag{7}$$

The focal point of data replication is to minimize total cost in the network. It not only includes its cost for data replication. In fact, once a member node i is selected as the candidate CH, it now has the possibility of communicating with the sink. Such communication cost $c_{\sin k}^i$ should also be taken into consideration. Moreover, due to extra burden of communication cost, there is chance that it consumes all its energy, becomes invalid, and in result causes the energy hole problem. To alleviate the problem, its residual energy $E_{residual}{}^i$ becomes an essential factor. We aim to find the proper candidate with not only less cost for data replication, but also larger residual energy and less communication cost with the sink. Players offer bids. The one with the highest value is elected as the winner and get selected as the candidate CH. For node i, we have its bid B_k^i defined as follows:

$$B_k^i = \frac{{E_{residual}}^i}{R_k^i \cdot c_{\sin k}^i} = \frac{{E_{residual}}^i}{r_k^i \cdot c_k^i \cdot c_{\sin k}^i} \tag{8}$$

For any player in the game of data replication, its bit remains the private information and cannot be known by any other players. As every player has its own valuation and whether or not one wants to be the winner depends only on the price he will have to pay, namely the bid in normal auction. Therefore, instead of submitting its real valuation, nodes may have a tendency to perform speculation by offering a higher value. Thus they may win the game with actually less payment.

Such situation is suitable for the adoption of the second price sealed auction, which is also known as “the Vickrey auction”. It was proposed by William Vickrey in 1961[10]. Such auction can suppress the potential speculation of any player and in the case of asymmetric information the outcome of the game can reach Pareto Optimality [2]. It stands for an ideal state of the resource allocation where no other outcome can make at least one player strictly better off on the premise that other players maintain well off. That is, a Pareto Optimal outcome cannot be improved upon without hurting at least one player.

4. Performance Evaluation

Simulate a wireless sensor network using NS2. We have 100 sensor nodes uniformly distributed in a 500×500 square region and the sink located far away, as it is shown in Fig.2.

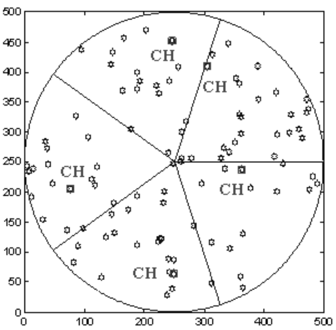

Fig. 2. Network for simulation

Here, we study the throughput for the sink that receives the collected data from the chosen cluster. If the connection between its CH and the sink breaks up, as it is shown in Fig.3 that a link disconnection occurs at 30^{th} second, the throughput of the sink drops to zero. However, with data replication, the sink is able to get the cluster's data from the replica in the candidate CH. Fig.4 shows that the throughput of the sink for the cluster's data is barely changed.

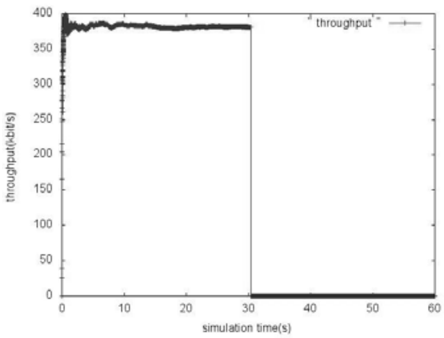

Fig. 3. Throughput of a sink related to a cluster's data (no data replication used)

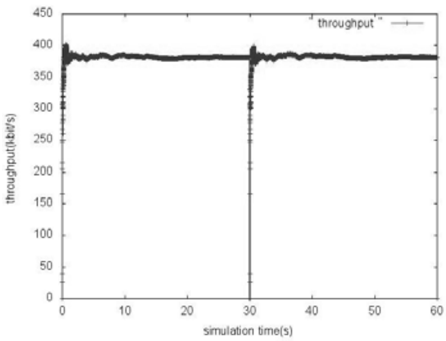

Fig. 4. Throughput of a sink related to a cluster's data (data replication used)

5. Conclusions

Clustering is an efficient method in wireless sensor networks. However, due to the limited resources, selfishness of nodes may affect the efficiency of clustering. In this paper, game theory is adopted to encourage nodes to serve as a CH. Moreover, set a candidate CH under the second price sealed auction and adopt data replication to ensure data transmission between a CH and the sink. Simulation results show that the throughput of the sink can still be guaranteed in case network divisions possibly occur.

Acknowledgement. This work was supported by Jiangsu Province Universities Natural Science Research Program (NO.11KJB510010) and Jiangsu Province Research and Innovation Project for College Graduates (NO.CXZZ12_0515).

References

1 I.Akyildiz, W.Su, Y.Sankarasubramaniam, E.Cayirci. "A survey on sensor networks". IEEE Commun. Mag., 2002, 40(8), pp.102-114.

2 R.B.Myerson. "Game Theory: Analysis of Conflict". Harvard University Press,1991,pp:vii-xi.

3 V.R.Q. Leithardt, C. Geyer, J. Sá Silva, R. Silva. "Use of Data Replication in WSNs". IEEE International Conference on Communications Workshops (ICC), Brazil, 2010, pp.1-5.

4 W.Heinzelman, A.Chandrakasan, H.Balakrishnan. "Energy-efficient routing protocols for wireless microsensor networks". in Proc.33rd Hawaii Int. Conf. System Sciences (HICSS), 2000.

5 S.Lindsey, C.S.Raghavendra. "PEGASIS: power-efficient gathering in sensor information systems". Proc. Int'l Conf. Comm. (ICC '01), 2001.

6 B.S.Lee, H.W.Lin, W.Tarng. "A Cluster Allocation and Routing Algorithm based on Node Density for Extending the Lifetime of Wireless Sensor Networks". 26th International Conference on Advanced Information Networking and Applications Workshops (WAINA), 2012, pp. 496-501.

7 R.Axelrod. "The Emergence of Cooperation among Egoists". The American Political Science Review,1981,75(2), pp. 306-318.

8 J.Wu, R.Axelrod. "How to Cope with Noise in the Iterated Prisoner's Dilemma". The Journal of Conflict Resolution, 1995,39(1), pp. 183-189.

9 F.Milan, J.J.Jaramillo, R.Srikant. "Achieving cooperation in multihop wireless networks of selfish nodes". Game Theory for Networks (GameNets 2006), Pisa, Italy, 2006.

10 T. Hara. "Effective replica allocation in ad hoc networks for improving data accessibility". 20th annual joint conference of the IEEE computer and communications societies, 2001, pp.1568-1576.

11 Shinohara M, Hara T, Nishio S. "Replica Allocation Considering Power Consumption in Mobile Ad Hoc Networks". 4th Annual IEEE International Conference on Pervasive Computing and Communications, 2006.

12 J.Zheng, X.C.Lu, Y.J.Wang. "A Clustering-based Data Replication Algorithm in Mobile Ad Hoc Networks for Improving Data Availability". Journal of Software, 2005,16(8), pp.1474−1483.

13 Y.Zhang, L.Yin, J.Zhao, G.Cao. "Balancing the Tradeoffs between Query Delay and Data Availability in Ad Hoc Networks", IEEE Transaction on Parallel and Distributed Systems, 2011(8).

14 W.Z.Lu, Y.C.Zhou, J.H.Li, B.P.Yan. "A Hierarchical Data Replication Method in Ad Hoc Networks ". 2010 2nd International Conference on Computer Engineering and Technology (ICCET), 2010, pp.23-27.

15 T.Rappaport. "Wireless communications: principles & practice". Englewood Cliffs, NJ: Prentice-Hall, 1996