International Workshop on Wireless Networks and Energy Saving Techniques (WINTEST-2014)

Information Gathering Schemes For Collaborative Sensor Devices

Salah A. Aly, Turki A. AlGhamdi, Mohamed Salim, Hesham H. Amin, Adnan A. Gutub

Smart Networks Lab, College of Computer and Information Systems (CIS), Umm Al-Qura University, Makkah, KSA

Abstract

We develop a novel information gathering scheme for Wireless Sensor Networks (WSNs) in which we consider \( n \) sensor nodes distributed randomly in a certain field to measure a physical phenomena. We desire to disseminate nodes’ data throughout the network such that a base station will be able to collect the sensed data by querying a small number of nodes. We propose a data dissemination and collection scheme to solve this problem, and we also present simulation results and derived bounds of this scheme.

Keywords: Data Collection Algorithms, Data Dissemination, WSNS

1. Introduction

Wireless Sensor Networks (WSNs) are expanding rapidly due to various applications and ease of development. However, WSNs encounter several challenges to be deployed efficiently in a given environment. Such challenges are limited source energy, limited transmission bandwidth, shortage coverage range, data dissemination, data persistence, redundancy of defective nodes and data security. A typical wireless sensor network (WSN) can be used in many applications such as monitoring a physical phenomena from the surrounding environment like temperature, gases, humidity, volcanoes and tornados.

Many techniques are used in data dissemination [1], [9], [11] and cluster head election [4], [5], [7], [12]. Fountain codes and random walks have been used to disseminate data from \( \kappa \) sources to a set of storage nodes \( \tau \), see [2], [6], [10], [11]. LEACH algorithm [3], [13] is the most popular clustering algorithm. Several cluster head selection algorithms are based on LEACH architecture. The main drawback of the mentioned techniques is the requirement that positions of all sensors must be known.

We consider a model for large-scale wireless sensor networks with \( n \) identical sensing nodes distributed randomly and uniformly in a certain field. The nodes do not know the locations of the neighboring nodes as required in [5], [8] and they don’t maintain routing tables. In this work, we propose two algorithms for data dissemination and...
data collection in wireless sensor networks. The first algorithm is Pre-known Head selection data Dissemination and Collection Algorithm (PHDCA). The second algorithm is Random Head selection data Dissemination and Collection Algorithm (RHDCA). We aim to develop an efficient method to randomly distribute and collect information from $n$ sensors by querying $10\% - 20\%$ of nodes for retrieving information about all network nodes with a high probability. This work is organized as follows. In Section 2 we introduce the network model. In Sections 3 and 4, we describe and demonstrate some analysis for the DCA’s algorithm. In Section 5 we present simulation studies of the proposed algorithm.

2. Network Model

Consider a set of $n$ identical sensing nodes distributed randomly in a field $F$ of dimensions $A = L \times W$, where $L$ and $W$ are the length and the width of $F$, respectively. We assume that each node has at least one neighboring node, meaning that with probability $P = 1$ there are no isolated nodes.

Definition 1 (Cluster head). The cluster head node (HN) is an arbitrary node among all network nodes $\mathcal{N}$ which exchanges its neighbors data with the other neighboring cluster head nodes.

Definition 2 (Node degree). The node degree $d_{s_i}$ is the number of neighboring nodes to the node $s_i$ within its connectivity range. The average mean degree of all nodes in $\mathcal{N}$ is given by

$$\mu = \frac{1}{n} \sum_{i=1}^{n} d_{s_i}. \quad (1)$$

The total period $(T)$ is the period after which the sensed data has been disseminated in the network $\mathcal{N}$ and is divided into $\epsilon$ equal time slots, $T = \epsilon \times t$, for some integer number $\epsilon$. The algorithms performance and simulation results confirm our theoretic bounds. The head nodes consume more energy than other nodes due to excess transmissions needed for data dissemination and data collection. So, such nodes are dynamically selected to apply fairness in energy consumption on all nodes. Also, the dynamical selection improves the performance of data dissemination in the network. The head nodes will be changed every time slot $t$. The number of head nodes in the network is $k$ (where $k/n \approx 10\%$). The selection of $T$ depends on the intended application (i.e. $T$ is small for high data rate applications and large for low data rate applications).

Assumptions:

- Let $S = \{s_1, ..., s_n\}$ be a set of $n$ identical sensing nodes distributed randomly in a field $F$ of dimensions $A = L \times W$, where $L$ and $W$ are the length and the width of $F$, respectively.
- Let $H = \{h_1, ..., h_k\}$ be a set of $k$ head nodes selected from the $n$ sensing nodes to disseminate the data in the network and they will be changed at each time slot $t$.
- Let $T$ be the period after which the sensed data has been disseminated in the network and it is divided into $\epsilon$ equal slots $t = \{t_1, ..., t_\epsilon\}$.
- The nodes use flooding to know their neighbors, as each node will send a message containing its $ID_{s_i}$ to all neighboring nodes. Each node receives an incoming $ID_{s_i}$ from any node $s_i$ will consider the node of the incoming $ID_{s_i}$ as its neighbor.
- Each node in the network generates a packet $P_{s_i}$ as follows: $P_{s_i} = (ID_{s_i}, x_{s_i}, flag)$, where, $ID_{s_i}$ is the ID of node $s_i$, $x_{s_i}$ is the sensed data of node $s_i$ and $flag$ is a variable set to 0 in flooding process or to 1 otherwise.
- Each node has a radio range coverage $r_i$. The node $s_i$ will be considered as a neighbor of $s_j$ if and only if $d_{s_i, s_j} \leq r_j$, where $d_{s_i, s_j}$ is the distance between nodes $s_i$ and $s_j$.

3. DCA’S ALGORITHM

3.1. PHDCA ALGORITHM

In PHDCA algorithm we dynamically select the $k$ cluster head nodes that disseminate the data in the network according to a pre-known manner. The algorithm can be classified into four phases as follows:
Input: A sensor network with $S = \{s_1, \ldots, s_n\}$ source nodes, $n$ source packets $x_{s_1}, \ldots, x_{s_n}$.

Output: storage buffers $y_1, y_2, \ldots, y_n$ for all sensors $S$.

$t = 1$; //initiate the value that represents the number of time slot in the period $T$.

foreach node $u = 1 : n$ do
    if $u \leq 0.1n$ then
        $u$ is a head node;
    end
end

foreach node $u = 1 : n$ do
    Generate a packet containing $ID_u$, $flag = 0$ and broadcast this message to its set of neighbors;
    $P_u = (ID_u, x_u, flag)$;
end

while still remains surviving nodes do
    foreach node $u = 1 : n$ do
        if $u$ sensed new data then
            $u$ will send this data to some of its neighbors randomly;
        end
    end

    foreach head node $h = 1: k$ do
        $h$ and its neighboring head nodes exchange their neighbors data with each others;
    end

    if $t$ expired then
        Generate new $k$ head nodes as follows:
        $t++$;
        foreach node $u = 1 : n$ do
            if $0.1n(t - 1) < u \leq 0.1nt$ then
                $u$ is a head node;
            end
        end
        if $t == \epsilon$ then
            $t=0$;
            $n = n_{received}$; //updates $n$ by the received estimated node number from base station.
        end
    end
end

Algorithm 1: PHDCA algorithm

- **Initialization phase:** In this phase, the head nodes are initially selected from $ID_{s_1} = 1 : 0.1n$ at the first time slot $t_1$.
- **Flooding phase:** In this phase, each sensor will broadcast a message containing its $ID_{s_i}$ to be able to discover its neighbors to store them in its data base. If any node receives any incoming $ID_{s_i}$, it will consider the node of the incoming $ID_{s_i}$ as its neighbor. Also, the broadcasting message containing a flag equal zero to indicate the flooding phase.
- **Sensing and data dissemination phase:** In this phase, such sensor reads a new data, it will send this data to some of its neighboring nodes. The neighboring head nodes will disseminate the data in the network by exchanging their neighbors data among them. The head nodes will be changed at each time slot and repeated each period $T$.
- **Data collection phase:** In this phase, the base station can query small number of any nodes to retrieve the data sensed by the $n$ sensing nodes and make an estimation for $n$ to send it to the first survived node.
3.2. RHDCA ALGORITHM

In PHDCA algorithm, we assumed that the selection of head nodes is pre-known at each time slot $t$ and the head nodes are repeated each period $T$. We extended PHDCA to obtain RHDCA that randomly selects $k$ head nodes at each time slot $t$. The performance of RHDCA is topology independent due to randomly selection of head nodes. The difference between the two algorithms is the sensing and data dissemination phase as follows:

**Sensing and data dissemination phase:** In this phase $k$ head nodes are selected randomly at each time slot $t$. The $k$ head nodes may be not repeated each period $T$. Also, each sensor reads a new data, it will send this data to some of its neighboring nodes. The neighboring head nodes will disseminate the data in the network by exchanging their neighbors data among them.

4. ANALYSIS

In this section we analyze the proposed DCA’s algorithm.

**Lemma 3.** The probability that a set $M$ of sensors has at least one cluster head node is given by

$$
\Pr(M \cap H) = 1 - \prod_{i=1}^{m} \left(1 - \frac{k}{n-i+1}\right),
$$

(2)

where, $m = |M|$ is the number of nodes in $M$.

**Proof.** Number of ways in which the $m$ nodes can be drawn from the total number of nodes $n$ is \( \binom{n}{m} = \frac{n!}{m!(n-m)!} \). Number of ways so that no head nodes exist in the set $M$ is \( \binom{n-k}{m} \). So, the probability that the set $M$ has no cluster head nodes is \( \frac{(n-k)}{\binom{n}{m}} \). Hence, the probability that the set $M$ has at least one head node is \( 1 - \frac{(n-k)}{\binom{n}{m}} = 1 - \prod_{i=1}^{m} \left(1 - \frac{k}{n-i+1}\right) \).

**Lemma 4.** The probability that a set $M$ of sensors has a set $Z$ of cluster head nodes is given by

$$
\Pr(Z) = \frac{\binom{n-k}{m-z}}{\binom{n}{m}} \binom{k}{z},
$$

(3)

where, $z = |Z|$ is the number of nodes in $Z$.

**Proof.** Number of ways in which the $m$ nodes can be drawn from the $n$ sensing nodes is \( \binom{n}{m} \). From the Fundamental Counting Theorem, the total number of ways in which $z$ head nodes and $m-z$ non head nodes can be drawn from the $n$ sensing nodes is \( \binom{n-k}{m-z} \binom{k}{z} \). So, The probability that a set of $n$ sensor has $z$ head nodes is \( \frac{\binom{n-k}{m-z} \binom{k}{z}}{\binom{n}{m}} \).

**Definition 5 (Head energy consumption ($E_{th}$)).** is the energy consumption at all nodes due to data dissemination in the network $N$ when all nodes have the same coverage range and packet size.

**Lemma 6.** The total energy consumption at the sensing nodes due to sending the sensed data to their neighbors is given by

$$
E_s = n(p_t + \mu p_r),
$$

(4)

where all nodes have the same coverage range and packet size.

**Proof.** The energy consumption at nodes $n$ due to sending its sensed data is $np_t$. The energy consumption at nodes $n$ due to all received packets is $\sum_{i=1}^{n} p_r \times d_{s_i}$. Hence, assuming that each node updates its data one time at each period $T$, the energy consumption at the $n$ sensing nodes is

$$
E_s = \sum_{i=1}^{n} (p_t + d_{s_i} p_r) = n(p_t + \mu p_r).
$$

(5)
5. Simulation and Performance Evaluations

In this section we show some simulation results to illustrate the performance of the proposed algorithm.

**Definition 7.** Decoding Ratio ($\eta$) is the ratio between the number of queried nodes $\hat{n}$ and the total number of sources $n$, 
$$\eta = \frac{\hat{n}}{n}$$

**Definition 8.** Successful Decoding Probability ($P_s$) is the probability that the $n$ source packets are all recovered from the $\hat{n}$ querying nodes.

Fig. 1 shows the relation between the successful decoding probability and the decoding ratio for different values of sensing nodes $n$ in PHDCA and RHDCA algorithms. Fig. 1(a) and Fig. 1(b) show that increasing the number of network nodes $n$ and fixing the covering radius $r$ of all nodes will result in an improvement in the successful decoding probability as well. We can notice that as the number of nodes increasing, the number of queried sensors can be decreased to recover the data with a reasonable successful probability. Particularly, for $n > 500$, we see that querying up to 10% will reveal about 85% of network data in PHDCA and about 92% of network data in RHDCA. Fig. 2 shows the amount of energy consumption at each node after the dissemination of data in the network $\mathcal{N}$ in PHDCA and RHDCA algorithms. From this figure we can notice that the energy consumption in PHDCA algorithm is better than the obtained result in RHDCA algorithm. We assumed that the energy consumption at the sensing node due to sensing the data it self is neglected and each sensor node is assumed to be of initial battery charge 5 Joule. We calculated energy consumption according to [13], they assumed that the energy lost at a sensor node $s_i$ due to transmission of one packet is given by
$$p_t = (50 \times 10^{-9} + 100 \times 10^{-12} \times r_{si}^2) \times \psi_{si},$$  \hspace{1cm} (6)

and the energy lost at a sensor node $s_i$ due to receiving of one packet is given by
$$p_r = 50 \times 10^{-9} \times \psi_{si},$$ \hspace{1cm} (7)

where $\psi_{si}$ is the packet size of node $s_i$.

Our future work will include accurate practical algorithms to optimize energy consumptions in the wireless sensor network. In addition, we will analyze the proposed algorithms and present several performance studies.

Acknowledgements

This work is funded by a grant from the Long-Term National Plan for Science, Technology and Innovation (LT-NPSTI), No. 11-INF1702-10, the King Abdulaziz City for Science and Technology (KACST), Kingdom of Saudi Arabia. We thank the Science and Technology Unit at Umm Al-Qura University for their continued logistics support.
Fig. 2. The energy consumption at each sensing node in network $N$ when $A=100^*100$, $n=300$, $\epsilon=10$, buffer size=40, packet size=2Kbits, node energy=5J and $r=5$.

References

1. S. A. Aly, A. Ali-Eldin, and H. V. Poor. A distributed data collection algorithm for wireless sensor networks with persistent storage nodes. 4th IFIP International Conference on New Technologies, Mobility and Security, February 2011.
2. S. A. Aly, H. Darwish, M. Youssef, and M. Zidan. Distributed flooding-based storage algorithms for large-scale wireless sensor networks. IEEE International Conference on Communications, June 2009.
3. S. Bandyopadhyay and E. J. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, San Francisco, CA, USA, March 2005.
4. L. Buttyan and T. Holczer. Private cluster head election in wireless sensor networks. Proceedings of the Fifth IEEE International Workshop on Wireless and Sensor Networks Security (WSNS’09), IEEE, IEEE, Macau SAR, PRC., October 2009.
5. A. G. Dimakis, V. Prabhakaran, and K. Ramchandran. Distributed fountain codes for networked storage. Speech and Signal Processing, ICASSP, May 2006.
6. J. Feldman, A. Kamra, V. Misra, and D. Rubenstein. Growth codes: maximizing sensor network data persistence. InProc. 2006 conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, 2006.
7. M. J. Handy, M. Haase, and D. Timmermann. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. The 4th International Workshop on Mobile and Wireless Communications Network, 2002.
8. W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. An application specific protocol architecture for wireless microsensor network. IEEE Transaction on Wireless Communications, 2002.
9. N. Imran, S. Khan, and I. Rao. A trust worthy and well-organized data disseminating scheme for ad-hoc wsn. International Journal of Computer Networks & Communications, May 2010.
10. S. Kokalj-Filipovic, P. Spasojevic, R. Yates, and E. Soljanin. Decentralized fountain codes for minimum-delay data collection. In CISS, March 2008.
11. Z. Kong, S. A. Aly, and E. Soljanin. Decentralized coding algorithms for distributed storage in wireless sensor networks. IEEE Journal on Selected Areas in Communications, February 2010.
12. C. M. Liu, C. H Lee, and L. C. Wang. Distributed clustering algorithms for data-gathering in wireless mobile sensor networks. J. Parallel Distrib. Comput., 2007.
13. O. Younis and S. Fahmy. Heed: A hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks. IEEE Transactions on Mobile Computing, 2004.