Computers and Electrical Engineering 000 (2016) 1–17

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Cluster-based data aggregation for pest identification in coffee plantations using wireless sensor networks

Roshan Zameer Ahmed[a,*], Rajashekhar C. Biradar[b], Shilpa Shashikant Chaudhari[a]

[a] Department of Electronics and Communication Engineering, Reva ITM, Bangalore, India
[b] School of Electronics and Communication Engineering, Reva University, Bangalore, India

ARTICLE INFO

Article history:
Received 23 July 2015
Revised 2 May 2016
Accepted 2 May 2016
Available online xxx

Keywords:
WSNs
Data aggregation
Redundancy
Clustering

ABSTRACT

This paper proposes a Cluster-Based Data Aggregation (CBDA) method for identifying pests in Arabica Coffee plantation using Wireless Sensor Networks (WSNs). Acoustic signals that are generated with biting sound by the pests inside stem are captured by WSN. Information regarding existence of pests is aggregated at Cluster-Head (CH) and is conveyed to base station. CH is selected using five states of each node: i-band, o-band, cluster-head request, idle and cluster-head. CH performs data aggregation with residual energy, time stamp using Kolmogorov's zero-one law to eliminate redundancy. Simulation analysis of CBDA is compared with fast local clustering, energy-efficient reliable data aggregation technique and energy-efficient data aggregation transfer in terms of aggregation ratio, message overhead, control overhead, packet delivery ratio, algorithmic complexity, delay, energy consumption, time-out period and clustering time. The CBDA simulation results outperform compared to the corresponding techniques.

1. Introduction

Coffea Arabica and Coffea Robusta production are considered as the economical foundation in many developing countries. It is the most beneficial agricultural product throughout the world [1]. Coffee White Stem Borer (CWSB) pest poses a serious threat to Coffea Arabica since it consumes the food and water resources utilized by the plant and kills it internally. The process begins with the female borer depositing around 50-100 eggs on the stem of the Coffea Arabica that are invisible to the naked eye. These eggs incubate in 10-12 days constituting the grub. The grub forms galleries by boring in different directions without crossing each other in the main stem and primary branches, gaining route to food and water resources of the stem. On proceeding further they close the entry points by covering the outer layer of stem with the residual material of boring. The boring activity is carried out for a lifespan of one year [2]. The pupil stage lasts for 3-4 weeks, then metamorphose into a grown beetle which resides in the tunnel for 3-7 days and comes out of the stem by biting an exit hole [3]. Cavity formed during the boring activity caused by grub in the stem is between 4 mm and 1 inch in size. The proliferation (multiplication) of the pest can be reduced by capturing it and blocking its exit. The affected plant shrinks, exhibiting symptoms of fading by turning yellow and forming clear ridges around the stem.

* Corresponding author.
E-mail addresses: roshanahmed31@revainstitution.org (R.Z. Ahmed), raj.biradar@revainstitution.org (R.C. Biradar), shilpachaudhari@revainstitution.org (S.S. Chaudhari).

http://dx.doi.org/10.1016/j.compeleceng.2016.05.001

The CWSB pest identification is a challenge that can be solved by adopting sensor technology over the coffee field. Usage of sensor technology requires design of WSN which consists of one or multiple base-stations, and a couple of sensor nodes deployed over a location called sensing field like a coffee field. It is designed to gather data from the sensor nodes to the Base Station (BS) using hop-by-hop communication [4]. The challenges involved with WSN design include limited energy, identification of efficient interference avoidance mechanism, design of proper security management services and large-scale node deployment. Developing optimized hardware, embedded software, data gathering and routing algorithms reduce the consumption of energy in WSN.

Radio communications used for data gathering and routing are the major sources of energy consumption [5]. One of the solutions to optimize the energy consumption during data transmission is by bringing down the communication overhead through data aggregation, which reduces redundancy and unnecessary data forwarding, from sensor nodes to the BS [6]. Instead of sending the received data from the sensing field by each node to the BS individually, data can be collected at one particular node and then combined to reduce the number of data packets. This process of combining sensed data into a reduced number of packets is called data aggregation. This helps to improve the lifetime of the WSN by reducing the traffic load, and the corresponding energy required for individual redundant data transmission without affecting the data accuracy. The existing architectures used in data aggregation techniques are either cluster-based or tree-based. In the cluster-based data aggregation, the network is split into clusters with a leader called the CH, which collects the data locally and then sends it to the BS. The cluster structure is very promising in cutting down the energy consumption in WSN. In tree-based data aggregation, sensor nodes are organized in a form similar to a tree structure. As the root of the aggregation tree is near to the BS, it involves more communications to BS, which results into more energy consumption. The CH selections in the existing cluster-based data aggregation are based on energy availability, network lifetime, positioning, deterministic selection based on node-id/node-degree, probabilistic adaptive selection based upon the resource availability [7–9]. These techniques need improvement in selection of the CH.

1.1. Related works

Some of the related works are as follows. The authors in [10] propose an assessment on WSN deployment in urban areas, with the amalgamation of its applications examined through strenuous research with fine links to distinguish between them. The authors in [11] review on potential renewable-energy resources using their characteristics and applications in WSN. The authors in [12] present the most widely used OS for WSNs, TinyOS. Cluster-based protocol for dynamic data aggregation in [13] created a pattern of CHs at the run time by making use of the field programmable gate arrays minimizing the query processing time and power consumption of the network. Aggregation tree model by the authors in [14], describes the route to capture the data after aggregation from the leaf cells to the root of the tree. In [15] authors summarize techniques for protecting the data confidentiality, integrity and security in data aggregation. The authors in [16] consider structures for aggregation and routing in scheduling data through nodes in the network for the purpose of reducing the energy cost. In [9] authors proposed Low Energy Adaptive Clustering Hierarchy (LEACH) which is a probabilistic protocol which uses one hop routing. The node under the rotation trial is considered as the CH, given that the random number is below the threshold. The authors in [17], describe data aggregation in terms of sleep scheduling algorithms by assigning consecutive time slot for data collection limiting information on the data buffer overflow mechanism. The authors in [18] describe a fault tolerance data aggregation process that removes the deceptive data sent by the compromised nodes by making use of locality sensitive hashing scheme. In [19] fuzzy petri nets are used for the CH selection, global reliability and energy load balancing at each node. Adaptive cluster selection strategy in [20,21], selects CHs assuming the distance from the node to the BS along with residual energy of the node. The authors in [22] proposed Fast Local Clustering (FLOC), that partitions a multi-hop wireless network into non-overlapping and nearly equal-sized clusters. It accomplishes the fault-local self-stabilization with locality of clustering. In EERDAT [23], each cluster selects a coordinator node randomly in the network that monitors the operation of the sensor nodes and commands them for specific operations. The CH is selected based on the cost value containing the residual energy factor and distance with respect to the coordinator node. In EEDAT [8], data aggregation is described as a two phase operation: aggregation phase and adaptation phase. The aggregation phase eliminates redundancy from raw data by finding similarities between the data during a certain period. The adaptation phase utilizes "sets-similarity join" function during successive periods to identify duplicate data-sets sent to the CH.

Proposed CBDA is compared with FLOC, EERDAT and EEDAT since these works are closely matching CBDA in following aspects. The clustering architecture considered in FLOC is dynamic that results in sudden change of CH and the topology of the network changes resulting in a possibility that few nodes may exist without any cluster whereas clustering approach in CBDA solves this problem by assigning a CH to each sensor node in a fixed environment of WSN. Even though, each node belongs to a cluster, the same node cannot reside in multiple clusters. CBDA maintains states of the nodes at CH. Proper handshaking signals for clustering have been defined in CBDA, when compared to FLOC. For these reasons the clustering mechanism in CBDA cannot be replaced by the mechanism used in FLOC. EERDAT maintains a queue mechanism to identify redundant packet in the network. Aggregated data at the CH is then forwarded to the coordinator node that checks for the loss ratio and further performs aggregation on the received data before it forwards it to the sink node. This results in excess network processing, delay and energy consumption. Based on the loss ratio identified by the coordinator node the cluster size is made to shrink. Along with the maintenance of queue mechanism for identifying the redundant packets, signal strength evaluation and sequence matching is performed in EEDAT whereas CBDA uses residual energy for data aggregation.