Ad Hoc Networks 25 (2015) 38–53

Contents lists available at ScienceDirect

Ad Hoc Networks

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

A hybrid clustering technique using quantitative and qualitative data for wireless sensor networks

Jiang Zhu [a], Chung-Horng Lung [a,*], Vineet Srivastava [b]

[a] Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada
[b] Cistech Limited, Ottawa, Ontario, Canada

ARTICLE INFO

Article history:
Received 3 April 2014
Received in revised form 18 July 2014
Accepted 12 September 2014
Available online 28 September 2014

Keywords:
Wireless sensor networks
Clustering protocol
Hierarchical agglomerative clustering
LEACH

ABSTRACT

Clustering is an efficient method to organize sensor nodes in Wireless Sensor Networks (WSNs) for data transmissions and energy saving. To perform clustering, many methods require geographic location data for calculating the distance between sensor nodes. But location data may not always be available due to Global Positioning System (GPS) failures or may not be practical in consideration of all sensor nodes due to the high cost and energy consumption of GPS. Alternatively, Received Signal Strength (RSS) or RSS Indicator (RSSI) has been used to estimate the distance. But many studies have shown that RSS or RSSI is not reliable in practice. In order to mitigate these realistic problems, this paper proposes a hybrid clustering protocol – Hybrid Distributed Hierarchical Agglomerative Clustering (H-DHAC) – which uses both quantitative location data and binary qualitative connectivity data in clustering for WSNs. Our simulation results reveal that H-DHAC only has a slightly lower percentage of compromise in performance in terms of network life time and total transmitted data compared to similar approaches that use complete location data. However, H-DHAC still outperforms the well known clustering protocols, e.g., LEACH and LEACH-C. On the other hand, the cost of H-DHAC can be significantly lower in comparison to those approaches that use complete quantitative location data, as GPS is not required for all sensor nodes. In addition, H-DHAC still can be operational in the presence of GPS failures.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In Wireless Sensor Networks (WSNs), sensor nodes are usually randomly deployed and they operate in an unattended fashion, perhaps even in unpredictable or harsh environments. Since the energy of sensor nodes is limited in many or most applications, the major concern of WSNs has always been energy efficiency. Clustering is one of the efficient techniques to organize sensor nodes for energy-efficient data transmissions. Clustering in WSNs is a task that assigns sensor nodes into different groups. In this process, clustering coefficients are used for measuring similarities or dissimilarities between nodes. Sensor nodes with higher similarities, e.g., they are close in proximity, tend to be grouped together. Sensor nodes that belong to the same cluster can then collaborate and send aggregated data to the Base station (BS) instead of sending data separately to BS to reduce energy consumption.

A number of clustering protocols, see Section 2, have been proposed aiming to efficiently organize sensor nodes [1,2] and many of those clustering techniques relay on Global Positioning System (GPS) to obtain the location information for the sensor nodes or Received Signal Strength (RSS) to calculate the distance between two nodes. How-

* Corresponding author.
E-mail addresses: jiangzhu@sce.carleton.ca (J. Zhu), chlung@sce.carleton.ca (C.-H. Lung), vineet@cistech.ca (V. Srivastava).

http://dx.doi.org/10.1016/j.adhoc.2014.09.009
1570-8705/© 2014 Elsevier B.V. All rights reserved.

ever, in practice, many unexpected or uncontrollable factors may occur due to the complicated network environments, such as poor channel quality and sudden sensor node failures. Most of the existing clustering protocols are based on an ideal case and fail to take some key issues into consideration, such as the reliability of GPS or RSS, and the cost of GPS. By adopting the assumption that GPS or RSS is always reliable, or ignoring the practical implications of the reliability or cost issue, the solutions to WSNs become disputable. The following paragraphs explain more about the reliability issue of GPS and RSS.

In WSNs, clustering is the process of organizing sensor nodes that have high proximity in groups. Clustering coefficients can be differentiated in two categories: quantitative and qualitative. For example, the location of sensor nodes is one kind of quantitative data. The distance between two nodes can be obtained based on the well known Pythagorean Theorem or Euclidean distance. In this case, the distance is the dissimilarity quantitative coefficient. Normally, closer nodes are more likely to form a cluster. In WSNs, when two nodes are communicating with each other, less radio communication energy will be consumed if they are closer and they are transmitting and aggregating data within a cluster. Hence, it is usually desirable to use distance as the representation of dissimilarity.

Typically, quantitative location data are used for calculating the distance between two sensor nodes. Studies which use location data for clustering assume that every sensor node is aware of its location or has GPS equipped. Although GPS is the most accurate and direct localization identification method, it faces a certain possibility of failure and the cost of a sensor node generally becomes significantly higher with GPS equipped (see Section 2 for more detailed description).

RSS or RSSI is another kind of quantitative data to estimate the distance between sensor nodes. RSS is the radio signal power that is measured at the receiver's end. In the ideal case, if one takes the value of RSS as x axis and the value of real distance as y axis to draw a plan chart, the outcome is expected to show that they are linearly related. In other words, the larger signal power detected indicates closer distance. Some papers, such as LEACH [3] and LACBER [4], have adopted RSS as the distance estimator fully or partially in their clustering scheme. However, RSS and RSSI are not reliable distance estimators in practice which has been proven in many studies with experiments in real environments (also see Section 2 for more discussions).

In contrary to quantitative data, in which data is in the form of numeric values, qualitative data use binary values to describe particular information in clustering [5]. In the area of WSNs, qualitative data can represent the wireless radio connectivity information, which is a binary representation of whether two nodes can directly communicate with each other, e.g., a 1 value means two nodes are directly connected, a 0 value means otherwise. By having the complete connectivity information of nodes in a WSN, one can create up to an $N \times N$ (N is the number of nodes in the network) binary matrix to represent connectivity relationship of sensor nodes in the network. Afterwards, this binary matrix will be further converted to similarity matrix with qualitative coefficients which can then be used for clustering.

A similarity qualitative coefficient (can be easily converted to dissimilarity coefficient) which is calculated from binary qualitative data is a good measurement of difference between two sensor nodes. However, qualitative coefficients are not as accurate as quantitative coefficients for clustering. Although qualitative data may not be the most accurate one, the experimental results [6] do not reveal a huge difference. Therefore, qualitative connectivity data is an adequate substitute in the absence of quantitative location data.

The Hierarchical Agglomerative Clustering (HAC) [5] technique is a conceptually simple but effective centralized approach, which has been successfully applied in many areas. Distributed HAC (DHAC) [6] adapted HAC to WSNs as a distributed bottom-up hierarchical clustering protocol without having global knowledge or a centralized compute. To address the reliability and cost issues of GPS and RSS/RSSI, this paper proposes a Hybrid Distributed Hierarchical Clustering (H-DHAC) protocol which is based on the fundamental HAC model. Nevertheless, H-DHAC further modifies the DHAC approach by extending it to support *hybrid data* that uses both quantitative location data and qualitative connectivity data.

In other words, H-DHAC is proposed to overcome some of the aforementioned practical problems for GPS or RSS/RSSI failures, the high cost issue of GPS, and the lower lifetime of sensor nodes using only qualitative data. The main idea is that H-DHAC does not require every sensor node to have the location data and does not depend on RSS or RSSI to estimate distance. H-DHAC can operate even if some location information is missing. Therefore, H-DHAC not only provides a flexible design alternative for those who have a specific demand on cost or performance (e.g., sensor lifetime), but also supports reliable clustering by considering possible GPS failure(s) or missing GPS information. We assume that the GPS failure can be identified due to a loss of signal or response from the GPS component. A number of experiments have been performed to compare some representative or closely related approaches. The experimental results using NS-2 show that H-DHAC either outperforms, e.g., LEACH [3] and LEACH-C [3], or are comparable to some clustering techniques, e.g., DHAC, that make use of the location information.

The rest of this paper is organized as follows. Section 2 describes the background and challenges. Section 3 presents the proposed H-DHAC protocol. Section 4 presents the simulation and results, which is then followed by a comparison of cost using all GPS-capable nodes and H-DHAC. Finally, Section 5 concludes this paper.

2. Background and challenges of missing quantitative location data

One of the main design goals in WSNs is to prolong the network lifetime. A number of approaches have been reported in the literature in dealing with this crucial design issue [2,22,33]. Among those approaches, clustering is an important technique in WSNs, as clustering supports

Download English Version:

https://daneshyari.com/en/article/445390

Download Persian Version:

https://daneshyari.com/article/445390

Daneshyari.com