Ad Hoc Networks 35 (2015) 149–160

Contents lists available at ScienceDirect

Ad Hoc Networks

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

Residual energy-based adaptive data collection approach for periodic sensor networks

Abdallah Makhoul [a,*], Hassan Harb [a,b], David Laiymani [a]

[a] FEMTO-ST Laboratory, DISC departement, University of Franche-Comté 17 AV du Maréchal Juin, Belfort 90016, France
[b] Departement of Computer Science, Lebanese University, Beirut, Lebanon

ARTICLE INFO

Article history:
Received 3 December 2014
Revised 30 July 2015
Accepted 20 August 2015
Available online 28 August 2015

Keywords:
Periodic sensor networks (PSNs)
Adaptive sampling models
Residual energy
Real data measurements

ABSTRACT

Due to its potential applications and the density of the deployed sensors, distributed wireless sensor networks are one of the highly anticipated key contributors of the big data in the future. Consequently, massive data collected by the sensors beside the limited battery power are the main limitations imposed by such networks. In this paper, we consider a periodic sensor networks (PSNs) where sensors transmit their data to the sink on a periodic basis. We propose an efficient adaptive model of data collection dedicated to PSN, in order to increase the network lifetime and to reduce the huge amount of the collected data. The main idea behind this approach is to allow each sensor node to adapt its sampling rate to the physical changing dynamics. In this way, the oversampling can be minimized and the power efficiency of the overall network system can be further improved. The proposed method is based on the dependence of measurements variance while taking into account the residual energy that varies over time. We study three well known statistical tests based on One-Way Anova model. Then, we propose a multiple levels activity model that uses behavior functions modeled by modified Bezier curves to define application classes and allow for sampling adaptive rate. Experiments on real sensors data show that our approach can be effectively used to minimize the amount of data retrieved by the network and conserve energy of the sensors, without loss of fidelity/accuracy.

1. Introduction

Wireless Sensor Networks (WSNs) have become a highly active research area due to their increasing potential impact on the quality of people's lives. They are used in many application domains for monitoring and tracking purposes [1,2]. In the majority of these applications, data generated across numerous sensors can produce a significant portion of the big data. For example, if we consider next generation driver assistance systems, Vehicle-2-Vehicle (V2V) or Vehicle-2-Roadside concepts, a large amount of sensor data is generated and needs to be fused and evaluated [3]. In WSNs, data collection can be categorized into three different models: query-driven, event-driven or time-driven [4]. In this paper, we focus on the third model of data collection, which is called Periodic Sensor Networks (PSNs), where each sensor monitors a given area and sends back data to the sink on a periodic basis. This periodic model is used for applications where certain conditions or processes need to be monitored constantly, such as the temperature, pressure, humidity, etc.

In PSN, there are two major challenges. First, the network should have a lifetime long enough to fulfill the application requirements. Second, massive and heterogeneous data collected from networks make data management more complex. Researchers' strategies are often targeted to minimize the amount of data retrieved/communicated by the network without considerable loss in fidelity/accuracy. The goal of this

* Corresponding author. Tel.: +33384587775.
E-mail addresses: abdallah.makhoul@univ-fcomte.fr (A. Makhoul), hassan.harb@univ-fcomte.fr (H. Harb), david.laiymani@univ-fcomte.fr (D. Laiymani).

http://dx.doi.org/10.1016/j.adhoc.2015.08.009
1570-8705/

reduction is first to increase the network lifetime, by optimizing energy consumption of the limited battery for each sensor node, and then to help in analyzing data and making decision.

Subsequently in periodic monitoring, the dynamics of the monitored condition or process can slow down or speed up; if the sensor node can adapt its sampling rates to the changing dynamics of the condition or process, over-sampling can be minimized and power efficiency of the overall network system can be further improved [5]. Therefore, in order to keep the network operating for long time, adaptive sampling approach to periodic data collection constitutes a fundamental mechanism for energy optimization and data reduction. In this paper, we propose an efficient adaptive sampling approach based on the dependence of conditional variance on measurements, e.g. one-way ANOVA model and statistical tests, that varies over time. We study three different statistical tests (Fisher, Tukey and Bartlett) based on the One Way Anova Model while taking into consideration the residual energy of each node. Indeed, ANOVA model provides a statistical test of whether or not the means of several independent groups are all equal. It has been proved as an effective method to classify objects (or data) into groups whereas statistical tests are used to indicate which groups are significantly different [6]. As a common method in statistical inference, ANOVA has many applications in agricultural, biological, and engineering sciences [7]. In parallel, we provide a multiple levels adaptive model that takes into account the application criticality. It defines dynamically multiple levels of sampling rate corresponding to how many samples are captured per unit of time (or period). It uses behavior functions modeled by modified Bezier curves to define application classes and allow for sampling adaptive rate. Simulation results are presented to validate the performance of the proposed approach.

The rest of this paper is organized as follows; Section 2 presents related work on data collection in sensor networks. Section 3 presents our model of adaptive sampling rate based on the variance study. Section 4 proposes a multi levels activity model that uses behavior functions to define application classes. Section 5 describes how to integrate the residual energy to allow each node to compute its sampling rate. Experimental results are exposed in Section 6. Finally, we conclude our paper and we provide our directions for future work in Section 7.

2. Related work

Although there has been a large number of recent works on sensor networks, only a fairly small number explicitly deals with adaptive sampling approach in WSN [5,8–13]. The main goal of an adaptive sampling approach is to make the rate of sensing dynamic and adaptable; if the sensor node can adapt its sampling rates to the changing dynamics of the condition or process, over-sampling can be minimized and the computational load at the sink will be more flexible. Big data and residual energy for sensor nodes are also studied in WSN, such [14–16] and [17–20], respectively, since they can affect directly the whole network lifetime.

The authors in [9] propose an energy-efficient adaptive sampling mechanism which uses spatio-temporal correlation among sensor nodes and their readings. The main idea is to carefully select a dynamically changing subset of sensor nodes to sample and transmit their data. In [10], a machine learning architecture for context awareness is proposed. It is designed to balance the sampling rates (and hence energy consumption) of individual sensors with the significance of the input from that sensor. In [11], the authors propose an Adaptive Sampling Approach to Data Collection (ASAP) which splits the network into clusters. A cluster formation phase is performed to elect cluster heads and to select which nodes belong to a given cluster. The metrics used to group nodes within the same cluster include the similarity of sensor readings and the hop count. Then, not all nodes in a cluster are required to sample the environment. [12] proposes a TA-PDC-MAC protocol, a traffic adaptive periodic data collection MAC which is designed in a TDMA fashion. This work is designed in the way that it assigns the time slots for nodes activity due to their sampling rates in a collision avoidance manner. The authors in [4,13] propose a two-level data aggregation technique to eliminate the inherent redundancy in raw data collected from periodic sensor networks.

Some works, such as [14–16], study the gathering and processing of big data generated by the sensors in WSN. In [14], the authors suggest to treat big data scenarios in WSNs as they pose similar questions and problems as traditional big data scenarios. Then, they propose an aggregation strategy tied to technological prerequisites which enables the efficient use of energy and the handling of large data volumes. In [15], a real WSN is studied which generates big data: Anna Creek Station ranch in Australia. It proposes a removal technique to delete the redundant data according to some rules and to efficiently process the data on the map phase. In [16], the authors propose a mobile sink routing and data gathering method through network clustering based on the modified Expectation-Maximization (EM) technique. On the other hand, extracting knowledge from sensor data has received a great deal of attention by the data mining community such as clustering approach [21], association rules [22], and frequent patterns [23,24].

Other works, such as [17–20], study the residual energy in WSN. In [17], a hierarchical approach is provided to construct a continuous energy map of a sensor network. The proposed method consists of a topology discovery and clustering phase, followed by an aggregation phase when energy information collected is abstracted and merged into energy contours in which nodes with similar energy level are grouped into the same region. The authors in [18] present a method for constructing multiple routing paths based on paths which have highest cumulative residual energy. An advantage of selecting the path with highest cumulative residual energy is that it minimizes the probability that nodes along the path will deplete their energy and thus minimizes path failures in the network. In [19], a distance-energy cluster structure algorithm is presented. It considers both the distance and residual energy of nodes to improve the process of cluster head election and the process of data transmission.

Adaptive sampling techniques are very promising, because of their efficiency to optimize energy consumption and the network overload. However, most of the previous proposed solutions are implemented in a centralized manner that requires huge computations and communications.