J. Parallel Distrib. Comput. 68 (2008) 1207–1221

Contents lists available at ScienceDirect

J. Parallel Distrib. Comput.

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

Consistency-driven data quality management of networked sensor systems

Kewei Sha, Weisong Shi *

Department of Computer Science, Wayne State University, 420 State Hall, 5143 Cass Ave, Detroit, MI 48202, USA

ARTICLE INFO

Article history:
Received 19 April 2007
Received in revised form
26 August 2007
Accepted 3 June 2008
Available online 14 June 2008

Keywords:
Data quality
Consistency models
Wireless sensor networks
Energy efficiency
Adaptation

ABSTRACT

With more and more real deployments of wireless sensor network applications, we envision that their success is nonetheless determined by whether the sensor networks can provide *a high quality stream of data* over a long period. In this paper, we propose a consistency-driven data quality management framework called *Orchis* that integrates the quality of data into an energy efficient sensor system design. Orchis consists of four components, *data consistency models*, *adaptive data sampling and process protocols*, *consistency-driven cross-layer protocols* and *flexible APIs to manage the data quality*, to support the goals of high data quality and energy efficiency. We first formally define a consistency model, which not only includes *temporal consistency* and *numerical consistency*, but also considers *the application-specific requirements of data* and *data dynamics in the sensing field*. Next, we propose an adaptive lazy energy efficient data collection protocol, which adapts the data sampling rate to the data dynamics in the sensing field and keeps lazy when the data consistency is maintained. Finally, we conduct a comprehensive evaluation to the proposed protocol based on both a TOSSIM-based simulation and a real prototype implementation using MICA2 motes. The results from both simulation and prototype show that our protocol reduces the number of delivered messages, improves the quality of collected data, and in turn extends the lifetime of the whole network. Our analysis also implies that a tradeoff should be carefully set between data consistency requirements and energy saving based on the specific requirements of different applications.

1. Introduction

As new fabrication and integration technologies reduce the cost and size of wireless micro-sensors, we are witnessing another revolution that facilitates the observation and control of our physical world [1,6,5,22], just as networking technologies have changed the way individuals and organizations exchange information. Microsensors such as Motes from Intel and Crossbow [4] have been developed to make WSN applications possible; TinyOS [10,12] has been designed to provide adequate system support to facilitate sensor node programming; Several applications, such as habitat monitoring [35], ZebraNet [25], Counter-sniper system [36], environment sampling [2], target tracking [34], and structure monitoring [42], have been launched, showing the promising future of wide range of applications of wireless sensor networks (WSNs).

With the main function of collecting interesting and meaningful data, the success of WSN applications is nonetheless determined by whether they can provide a *high quality stream of data over a long period*. The inherent feature of unattended and untethered deployment of WSN in a malicious environment, however, imposes challenges to the underlying systems. These challenges are further complicated by the fact that WSNs are usually seriously energy and storage constrained. However, most previous efforts focus on devising techniques to save the sensor node energy and thus extend the lifetime of the whole WSN. We envision that data quality management has been becoming a more and more important issue in the design of WSNs.

In principle, the quality of data should reflect the timeliness and accuracy of collected data that are presented to interested recipients who make the final decision based on these data. Complementing to the work on the sensor design that improves the accuracy of sensing, in this paper, we intend to study the relationship between data quality and energy-efficient design of WSNs. To integrate and manage data quality in WSNs, we propose a framework named *Orchis*, which includes a set of data consistency models customized to WSNs, a set of APIs to management the quality of collected data, an adaptive protocol for data sampling, a set of consistency-driven cross layer protocols to support achieving the goals of data consistency and energy efficiency. The novelty of this work is that we propose to use consistency models, including temporal, numerical, and frequency three perspectives, as metrics to measure the quality of the collected data in wireless sensor networks, and based on these models, we propose a framework to manage data quality of WSNs. To the best of our knowledge, we are the first to propose consistency models in wireless sensor networks and to try to manage the data quality

* Corresponding author.
E-mail addresses: kewei@wayne.edu (K. Sha), weisong@wayne.edu (W. Shi).

0743-7315/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jpdc.2008.06.004

from the viewpoint of data consistency. Among these components, we address two of them, consistency models and the Alep protocol in detail in this paper. First, we formally define a new metric, data consistency model, to evaluate the data quality. Intuitively, most people think that the higher requirements of data quality, the more energy will be consumed. However, we find that this intuition is not necessarily held, and that the energy can be saved if we consider data consistency and data dynamics together. This fact in turn inspires us to attack the problem from the perspective of data consistency and data dynamics, and exploit the data consistency in system protocol design. Thus, an adaptive, lazy, and energy-efficiency data collection protocol called *Alep* is proposed. Finally, our comprehensive performance evaluation based on both simulation and prototype implementation shows that Alep improves the quality of data, saves energy, and extends the lifetime of WSNs.

The contributions of this paper are four-fold. First, to the best of knowledge, we are the first to propose a general framework that integrates data quality management in the design of WSNs. Second, we formally define data consistency models as metrics to evaluate data quality. Third, we propose an adaptive, lazy, energy-efficiency protocol to improve data quality and save energy. Finally, a comprehensive performance evaluation has been undertaken based on both TOSSIM [12] and a prototype implementation using 13 MICA2 Motes. The rest of the paper is organized as follows. We first analyze the importance of data quality to WSN applications, then, abstract the specific consistency-related features of WSNs and their applications in Section 2. Section 3 depicts a consistency-driven data management framework. In Section 4, we present the formal definition of data consistency and data dynamics. An adaptive lazy energy-efficient protocol for data collection is described in Section 5. And Sections 6 and 7 report a comprehensive performance evaluation based on TOSSIM simulator and a prototype implementation of 13 MICA2 motes respectively. Finally, related work and conclusion remarks are discussed in Sections 8 and 9 respectively.

2. Data consistency analyses

Data consistency plays a very important role in the success of WSN applications. For example, researchers in sleeping start to consider leveraging wireless sensors to collect the environment information of patients. In order to convince the sleeping research community that WSNs are indeed a viable approach for their traditional self-report survey based approach, researchers usually collect two types of data. One is reported records by the patient, the other is data collected from sensors. We learned that these two sets of data do not match very well. As a result, they do not know which set of data is more appropriate for their research.[1] Thus, we argue that it is vital to have a mechanism to manage and control the quality of data collected by WSNs. This in turn will guarantee the fidelity brought by this promising new approach, and pave the way of wide acceptance in different applications. We conjecture that data consistency is a good metric that can be used to evaluate and control the data qualify. Unlike data consistency in traditional distributed systems, data consistency in WSNs, an application specific concept, has to consider specific features of WSNs and specific requirements of applications. In this section, we first analyze the special consistency related WSN features and then abstract the consistency requirements from the perspective of applications.

[1] Based on oral communication with our colleagues at College of Nursing, Wayne State University. This also inspires us to undertake this research.

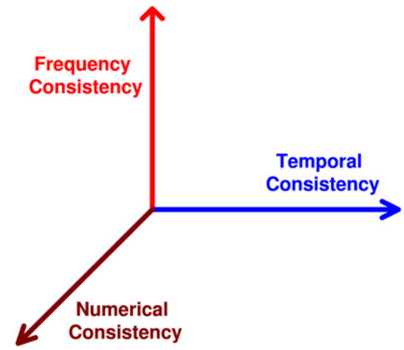

Fig. 1. A three-dimension view of consistency requirements.

2.1. Consistency related WSN features

Although a WSN is an instance of a distributed system, there are several significant differences between them. First, WSNs are a resource-constrained system. Due to the constraints of the memory size and the large amount of sampled data, data is usually not stored in sensors for a long period, but it will form data streams to be delivered to the sink(s) or base station(s). As a result, data consistency in WSNs will not focus on the read/write consistency among multiple data replicas as in traditional distributed systems. Instead, data consistency in WSNs is more interested in the spatial and temporal consistency of the same data, i.e. the consistency among several appearances of the data at different locations and at different time. Second, WSN applications may have more interests in a set of data which can depict the trend of the parameter being monitored or report an event. Thus, consistency models for data streams are more important than those for individual data. Third, compared with traditional distributed systems, the unreliable wireless communication is common, rather than exceptional, in WSNs. Thus, in consistency models, the data loss resulting from unreliable wireless communication should also be considered. Furthermore, in previous definition of data consistency [37], the effect of channel noises and attacks are neglected. We argue that attacks are normal nowadays and the security measures should be integrated in the system design from the initial stage.

2.2. Consistency requirements and data dynamics

WSNs are mostly application-specific systems that are widely used in variant applications, which have different data consistency requirements. Besides, WSNs are also data-centric systems, so that data consistency is closely related with data dynamics in the data field.

Considering both individual data and data streams, we argue that the quality of the data should be examined from three perspectives: *the numerical consistency*, *the temporal consistency*, and *the frequency consistency*, as shown in Fig. 1. The *numerical consistency* requires that the collected data should be accurate. Here we have two kinds of concerns on numerical errors: *absolute* and *relative*. Absolute numerical error happens when the sensor reading is out of normal reading range, which can be pre-set by applications. In the case of absolute numerical error, we can remove it and estimate a reasonable value for it. Relative numerical error depicts the error between the real field reading and the corresponding data at the sink. To trade off the resource usage and data accuracy, we can leverage estimation technologies to estimate readings at the sink while still providing the data with the same level of accuracy. As a result, some sensor readings can be dropped to save resource usage. Subsequently, there are relative numerical errors between the real ground truth and the collected data at the sink. The *temporal consistency* means that the data should be