Event Detection Services Using Data Service Middleware in Distributed Sensor Networks

Shuoqi Li, Ying Lin, Sang H. Son, John A. Stankovic, and Yuan Wei

Department of Computer Science, University of Virginia, USA
{sl7q, yl9f, son, stankovic, yw3f}@cs.virginia.edu

Abstract. This paper presents the Real-Time Event Detection Service using Data Service Middleware (DSWare). DSWare provides data-centric and group-based services for sensor networks. The real-time event service handles unreliability of individual sensor reports, correlation among different sensor observations, and inherent real-time characteristics of events. The event service supports confidence functions which are designed based on data semantics, including relative importance of sub-events and historical patterns. When the failure rate is high, the event service enables partial detection of critical events to be reported in a timely manner. It can also be applied to differentiate between the occurrences of events and false alarms.

1 Introduction

Sensor networks are large-scale wireless networks that consist of numerous sensor and actuator nodes used to monitor and interact with physical environments [11][14]. From one perspective sensor networks are similar to distributed database systems. They store environmental data on distributed nodes and respond to aperiodic and long-lived periodic queries [7][15][20]. Data interest can be pre-registered to the sensor network so that the corresponding data is collected and transmitted only when needed. These specified interests are similar to views in traditional databases because they filter the data according to the application's data semantics and shield the overwhelming volume of raw data from applications [8][26].

Sensor networks also have inherent real-time properties. The environment that sensor networks interact with is usually dynamic and volatile. The sensor data usually has an absolute validity interval of time after which the data values may not be consistent with the real environment. Transmitting and processing "stale" data wastes communication resources and can result in wrong decisions based on the reported out-of-date data. Besides data freshness, often the data must also be sent to the destination by a deadline. To date, not much research has been performed on real-time data services in sensor networks.

Despite their similarity to conventional distributed real-time databases, sensor networks differ in the following important ways. First, individual sensors are small in size and have limited computing resources, while they also must operate for long periods of time in an unattended fashion. This makes power conservation an important concern in prolonging the lifetime of the system. In current sensor networks, the major source of power consumption is communication. To reduce unnecessary data transmission from each node, data collection and transmission in sensor networks are always initiated by subscriptions or queries. Second, any individual sensor is not reliable. Sensors can be damaged or die after consuming the battery. The wireless communication medium is also unreliable. Packets can collide or be lost. Because of these issues we must build trust on a group of sensor nodes instead of any single node. Previous research emphasizes reliable transmission of important data or control packets at the lower levels, but less emphasis is on the reliability on data semantics at the higher level [23]. Third, the large amount of sensed data produced in sensor networks necessitates in-network processing. If all raw data is sent to base stations for further processing, the volume and burstiness of the traffic may cause many collisions and contribute to significant power loss. To minimize unnecessary data transmission, intermediate nodes or nearby nodes work together to filter and aggregate data before the data arrives at the destination. Fourth, sensor networks can interact with the environment by both sensing and actuating. When certain conditions are met, actuators can initiate an action on the environment. Since such actions are difficult to undo, reducing false alarms is crucial in certain applications.

The remainder of this paper is organized as follows: In section 2, we present related work. In section 3, we present the design of Data Service Middleware (DSWare) and some major components of DSWare. DSWare is a specialized layer that integrates various real-time data services for sensor networks and provides a database-like abstraction to applications. In section 4 we present a detailed description of the event detection mechanism. Event detection is one of the most important data services in sensor networks because it is a way to "dig" meaningful information out of the huge volume of data produced. It aims to find the "right data" at the "right place" and ensure the data is sent at the "right time". Event Detection Services in DSWare associate a confidence value with each decision it makes based on a pre-specified confidence function. It incorporates the unreliability of sensor behavior, the correlation among different factors, and reduces false alarms by utilizing data semantics. Section 5 presents the performance evaluation of the event detection mechanism. We conclude the paper in Section 6.

2 Related Work

There are many ongoing middleware research projects in the area of sensor networks, such as Cougar, Rutgers Dataman, SINA, SCADDS, Smart-msgs, and some virtual-machine-like designs [1][2][3][4][8][12][17][26]. COUGER and SINA are two typical data-centric middleware designs which have goals that are similar to our design goal of providing data services. In COUGER, sensor data is viewed as tables and query execution plans are developed and possibly optimized in this middleware. Our work on DSWare is more tailored to sensor networks, including supporting group-based decision, reliable data-centric storage, and implementing other approaches to improve the performance of real-time execution, reliability of aggregated results and reduction of communication. SINA is a cluster-based middleware design which focuses on the cooperation among sensors to conduct a task. Its extensive SQL-like primitives can be used to issue queries in sensor networks. However, it does not provide schemes to hide the faulty nature of both sensor operations and wireless communication. In SINA it is the application layer that must provide robustness and reliability for data services. In DSWare, the real-time scheduling component and built-in real-time features of other service components make DSWare more suitable than SINA for real-time applications in ad hoc wireless sensor networks.

Multisensor data fusion research focuses on solutions that fuse data from multiple sensors to provide more accurate estimation of the environment [16][22]. In mobile-agent-based data fusion approaches, software that aggregates sensor information are packed and dispatched as *mobile agents* to "hot" areas (e.g., the area where an event occurred) and work independently there. The software migrates among sensors in a cluster, collects observations, then infers the real situation [22]. This approach and our group-based approach both make use of consensus among a number of nearby sensors of the same type to increase the reliability of a single observation. The mobile-agent-based approach, however, leverages on the migration traffic of mobile agents and their appropriate processing at each sensor node in its routes. For instance, if a node in the route inserts wrong data or refuses to forward the mobile agents, the aggregation and subsequent analysis are untrustful. Our approach does not have such limitations: malfunctioning of individual nodes does not infect the entire group.

A fuzzy modelling approach is sometimes used for data fusion in sensor networks. It is used to model the uncertainty in sensor failures and faulty observations [25]. This approach is useful in modelling the sensor error rates due to equipment wear and aggregating local decisions from multiple sensors that measure the same type of data. Some optimal decision schemes focus on the fusion of asynchronously arriving decisions [10][24]. E. Bosse et. al. presented a modelling and simulation approach for a real-time algorithm in multi-source data fusion systems [9]. These data fusion schemes are suitable for increasing the accuracy of decisions, but require extensive computing resources. In our approach to event detection, the computation in fusion nodes is small.

Dempster-Shafer evidential theory is also applied to incorporate uncertainty into decisions in some sensor fusion research [21]. This scheme uses *Belief* and *Plausibility* functions to describe the reliability feature of each source and uses a normalized Dempster's combination rule to integrate decisions from different sources. Our *confidence function* is similar to Dempster-Shafer method except that we place the evidence in both temporal and spatial spectrums to take data real-time validity intervals and possible contexts into consideration.