Information Fusion 25 (2015) 1–15

Contents lists available at ScienceDirect

Information Fusion

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

In-network wireless sensor network query processors: State of the art, challenges and future directions

Farhana Jabeen [a,*], Sarfraz Nawaz [b]

[a] COMSATS Institute of Information Technology, Islamabad, Pakistan
[b] Computer Laboratory, University of Cambridge, Cambridge, United Kingdom

ARTICLE INFO

Article history:
Received 4 August 2014
Received in revised form 28 January 2015
Accepted 29 January 2015
Available online 7 February 2015

Keywords:
Wireless sensor network
In-network processing
Distributed queries
Sensor network query processor (SNQP)
Distributed processing

ABSTRACT

In wireless sensor networks (WSNs), energy is valuable because it is scarce. This causes their life time to be determined by their ability to use the available energy in an effective and frugal manner. In most of the earlier sensor network applications, the main requirement consisted mainly of data collection but transmitting all of the raw data out of the network may be prohibitively expensive (in terms of communication) or impossible at given data collection rates.

In the last decade, the use of the database paradigm has emerged as a feasible solution to manage data in a WSN context. There are various sensor network query processors (SNQPs) (implementing in-network declarative query processing) that provide data reduction, aggregation, logging, and auditing facilities. These SNQPs view the wireless sensor network as a distributed database over which declarative query processor can be used to program a WSN application with much less effort. They allow users to pose declarative queries that provide an effective and efficient means to obtain data about the physical environment, as users would not need to be concerned with how sensors are to acquire the data, or how nodes transform and/or transmit the data.

This paper surveys novel approaches of handling query processing by the current SNQP literature, the expressiveness of their query language, the support provided by their compiler/optimizer to generate efficient query plans and the kind of queries supported. We introduce the challenges and opportunities of research in the field of in-network sensor network query processing as well as illustrate the current status of research and future research scopes in this field.

1. Introduction

Wireless sensor networks (WSNs) have made it possible to collect data at a higher spatio-temporal resolution than would have been possible using a traditional (manual) data collection method [1]. WSNs allow for interaction with the environment at very high spatial and temporal densities. Such networks potentially enable observation over a large, hazardous area in the physical world with difficult access to scientists and where cost constraints preclude expensive components or strategies [2,3].

As scientific understanding of physical phenomena presupposes a study of their manifestation in time and space, this makes WSNs well-suited for real-time monitoring, control, and analysis of transient physical phenomena (e.g., a moving band of rain, a shape-shifting region of low temperature). As an example of the usefulness of this kind of information, consider the following context. Efficient water management is a major concern for farmers of many crops. Imagine that a farmer has deployed sensor nodes [2,3], and is interested in part of a field where the soil moisture has dropped below a certain threshold, so that only those parts are irrigated, given the limited water supply. A number of recent deployments of sensors have been made for precision agriculture purposes: about 25 nodes were deployed in Camalie vineyards [3], 45 sensor nodes in Network Avanzato per il Vigneto (NAV) system [4] for precision viticulture, 150 nodes in LofarAgro project [5] to fight fungal-disease in the field, about 26 nodes in AgFrostNet project [6] for frost monitoring. In a different context, consider the Great Duck Island deployment [7] for monitoring the nests in the petrel colony established there. The nests are situated in underground burrows, distributed in discrete patches around the island. Environmental conditions vary widely from patch to patch. For example, the biologists involved were interested in determining which environmental conditions yield an optimal microclimate for breeding, incubation, and hatching. For this purpose, they

---

* Corresponding author.
E-mail addresses: farhanakhan@comsats.edu.pk (F. Jabeen), sarfraz.nawaz@cl.cam.ac.uk (S. Nawaz).

http://dx.doi.org/10.1016/j.inffus.2015.01.007
1566-2535/

wanted to monitor burrow occupancy and the environmental changes occurring inside the burrow and on the surface. In all these scenarios, nodes collect temperature, moisture, and other environmental properties and, using their short-range radio, communicate with each other to send the real-time information outside the network for analysis. WSNs, therefore, allow the farmer/scientists to get a real-time digital picture, in the form of sensed measurements, of the physical world. Raw data being collected enables boilogists/farmers to see what is going on and to adjust their management strategies. In most of these environmental monitoring applications, the WSN is, by and large, an isolated system with depletable resources. Replacement of the batteries of sensor nodes under such scenarios is not feasible because of the following reasons: (1) time consuming and expensive; and (2) minimal disturbance is crucial for avoiding distortion in results [7] (the birds under study might change their behavioral patterns or distributions).

A significant concern, and by far most commonly studied one in the sensor network literature, is the lifetime of the network. In WSNs, energy is valuable because it is scarce; sensor nodes only have finite energy reserves drawn from batteries. The underlying issue is that the cost of wireless communication is significantly greater than that of processing, and grows exponentially with the distance between nodes [8]. In WSN, one general approach to data analysis is the *warehousing approach* that focuses on little (or no) in-network processing, instead raw data is pushed out of network. Transmitting every node sensed value to some destination that is external to the WSN for storage and off-line analysis may be prohibitively expensive and sometimes not possible, given the typical data collection rates and network sizes. In this approach, apart from network longevity, scalability is an issue, as it will result in increased bandwidth requirements, raising the risks of packet loss due to collisions. In contrast to warehousing approach, the *in-network processing approach* aims to reduce the need to transmit large amount of data by injecting more complex processing into the node, thus supporting network scalability (in terms of nodes), and network longevity.

In this approach, WSNs are construed as fully-fledged distributed system platforms, since sensor nodes cooperate to execute message-passing distributed algorithms. One instance of this is sensor network query processing, which views the WSN as a distributed database where both the data sources and the computational resources are partitioned over a large area. This conception of sensor networks has led to the approach of retrieving data from a sensor network by viewing it as the computational environment upon which *structured query language* (SQL) like declarative queries are compiled and optimized to run [9]. Declarative queries allow users to specify what data they want from a sensor network without needing to know details such as how to contact the relevant sensing devices on sensor nodes, how to deploy application logic, how to manage its execution and how to transmit results back to the user. In this view, a WSN produces the required data as the outcome of evaluating the declarative queries that express the corresponding requirements.

Programming WSNs requires specialized knowledge and the scarcity of available resources on a WSN node puts a tight limit on code size and makes debugging a cumbersome process. Implementing a simple application-specific data collection application may require thousands of lines of code in an embedded programming language. Madden et al. [10] and Yao and Gehrke [11] propose that WSNs can be programmed with considerably less effort with the use of the database paradigm. This has given rise to *sensor network query processors* (SNQPs) implementing in-network declarative query processing over WSNs, examples of which include TinyDB [10], and Cougar [11]. This allows for low-cost programmability and low-cost repurposing, since rather than reprogramming the network; users only need to pose a different query to the SNQP. Declarative queries rely on a query optimizer to compile the query into *query execution plan (QEP)*, which specifies how the query is to be executed that satisfies the required needs. A SNQP, therefore, has characteristics in common with a *distributed query processor* (DQP), in that data sources and query evaluation takes place across several sites.

The main contributions of this paper are as follows:

- We present a brief background on classical, distributed and stream query processing, introducing the reader to the highly constrained distributed computing platform that sensor networks give rise to.
- We outline the challenges in conducting in-network query processing in WSN and compare SNQPs with other, more established, forms of query processing and describe limitations of state-of-the-art SNQPs.
- We highlight how query processing in WSNs is handled by the current literature on sensor network query processors, the expressiveness of their query language, the support provided by their compiler/optimizer to generate efficient query plans and the types of queries supported by these systems. While distributed techniques for stream processing have been reviewed elsewhere [12], most of these solutions are not integrated with a query processing engine and query language.
- We presented a SNQP evaluation framework and compare existing approaches using this framework.
- Finally, we propose future research directions for query processing in WSNs.

The remainder of the paper is organized as follows: Section 2 describes why query processing in sensor networks as a distributed database differs from previous work. It also elucidates that SNQP's are stream *query processors* and streams require different models, query languages and architectures for efficient query handling. Section 3 surveys the work related to supporting SNQP's for evaluating declarative queries. Section 4, presents SNQP evaluation framework and compares existing SNQP solutions. Section 5 presents SNQP evaluation framework and compares existing work. Section 6, discusses related work on query processing. Section 7, focuses on future research scope, open issues and finally Section 8, concludes the paper.

2. Query processing

2.1. Background: distributed query processing

Classical database systems allow the user to retrieve the data in which he/she may be interested by posting declarative queries in languages such as SQL or Object Query Language (*OQL*). In relational *database management system* (DBMS) the query evaluator comprises of a query optimiser and an execution engine. The components of *query processor* or query optimiser consist of following components: (i) *parser*, (ii) *logical optimizer*, and (iii) *physical optimizer*. Once the user poses the query to the *DBMS*, the parser parses it and type checking is performed. This results in some internal representation, usually an *abstract syntactic tree* that is passed to the query optimizer. The query optimizer aims to select the appropriate query evaluation plan among the various, semantically equivalent candidate plans that the optimization process generates each one with potentially different execution costs. A plan specifies precisely how the query is to be executed and is usually represented in the form of a tree in which nodes represent algebraic operators (e.g., ***Selections***, ***Projections***, and ***Joins***) and the edges represent the consumer-producer relationship between