Poster Abstract: Routing and Processing Multiple Aggregate Queries in Sensor Networks

Niki Trigoni Alexandre Guitton Antonios Skordylis
Birkbeck College, University of London, London WC1E 7HX, United Kingdom
{niki,alexandre,adonis}@dcs.bbk.ac.uk

Abstract

We present a novel approach to processing continuous aggregate queries in sensor networks, which lifts the assumption of tree-based routing. Given a query workload and a special-purpose gateway node where results are expected, the query optimizer exploits query correlations in order to generate an energy-efficient distributed evaluation plan. The proposed algorithms, named STG and STS, identify common query sub-aggregates, and propose common routing structures to share the sub-aggregates at an early stage. Moreover, they avoid routing sub-aggregates of the same query through long-disjoint paths, thus further reducing the communication cost of result propagation. In this poster, we provide examples to illustrate the functionality and the communication savings of STG and STS compared to the existing tree-based approach.

Categories and Subject Descriptors

C.2.4 [**Computer-Communication Networks**]: Distributed Systems—*distributed databases, distributed applications*.

General terms

Algorithms, Performance.

Keywords

Sensor networks, Aggregate queries.

1 Introduction

Sensors are nodes with communication, computation, storage and sensing capabilities that can be deployed in large areas to monitor the ambient environment. They communicate their readings to one or more basestations (referred to as *gateways*) in a wireless multihop manner. A typical way of extracting information from a sensor network is to disseminate declarative aggregate queries into the network, asking nodes to periodically monitor the environment, and return aggregate results to the gateway in regular rounds. An example of such long-running queries is *"select avg(temperature) from Sensors where loc in Region every 10 min"*.

SenSys'06, November 1–3, 2006, Boulder, Colorado, USA.

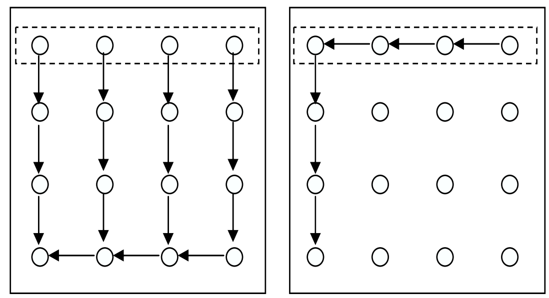

Figure 1. Example with one query.

Since nodes are battery-powered, energy preservation is a major consideration in system design, as it directly impacts the lifetime of the network. Recent studies have shown that radio communication is significantly more expensive than computation or sensing in most existing sensor node platforms. Hence, the main consideration in designing query processing algorithms is to minimize the communication overhead of forwarding query results from the sources to the gateway node. We assume that the cost of disseminating query information into the network can be ignored in certain scenarios. For example, in the case of long-running queries, query dissemination occurs once, whereas result propagation occurs repeatedly at regular rounds. Moreover, many monitoring scenarios apply a pure push model, in which nodes are programmed to proactively send specific information to the gateway. The energy expenditure in these cases depends primarily on the volume of propagated query results.

Tree-based routing has been proposed as an energy-efficient mechanism for processing aggregate queries in sensor networks [1, 2]. Tree construction is performed using simple flooding algorithms [2], data-centric reinforcement strategies [1] or energy-aware route selection schemes [4, 5]. After a tree is constructed, sensor nodes forward their readings along the paths of the tree, evaluating partial query results at intermediate nodes. The aforementioned research focused on processing a *single aggregate query given a routing tree*; the tree is generated using a tree selection scheme and is thereafter used for result propagation. More recent research has focused on *optimizing multiple aggregate queries given a routing tree* [3]. Query commonalities are taken into account to reduce the communication cost of result propagation, but without making any attempt to select suitable tree routes [3].

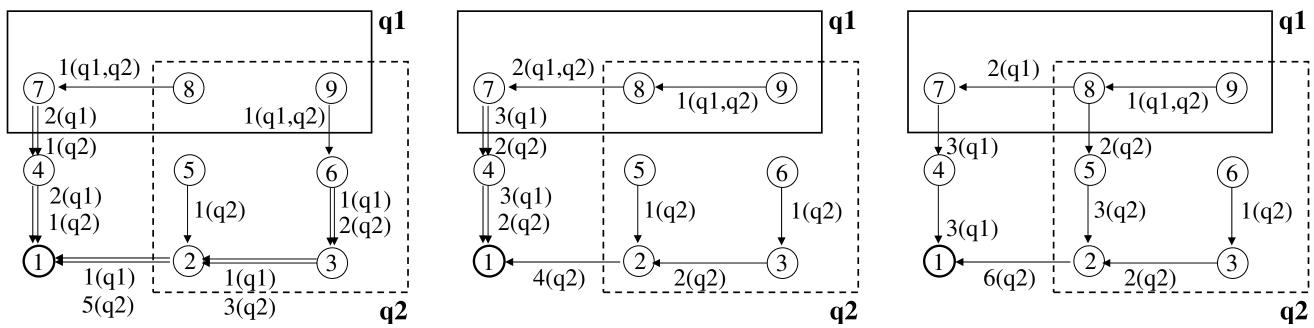

Figure 2. Example with two queries: (i) the left plan is based on a randomly selected tree, (ii) the middle plan is the output of STG, and (iii) the right plan is the output of STS.

Unlike previous approaches, this work considers the more general problem of multi-query optimization lifting the assumption of an existing aggregation tree. The objective is to find efficient routes that minimize the communication cost of executing multiple aggregate queries, by studying the interplay between the processing and routing aspects of query evaluation. Unlike previous work, there is no limitation for the selected routes to form a tree structure. The only requirement is that the optimizer must operate in a distributed manner, and should scale gracefully with the network size. In the next section, we show the importance of carefully selecting routes to process one or more aggregate queries.

2 Illustrative examples

The potential advantages of carefully selecting a routing and processing plan for executing aggregate queries are shown in the following examples. Figure 1 shows an example of processing a single aggregate query, which asks for the sum of all readings in the dotted rectangular area. Notice that a total number of 15 messages are sent along the left minimum-hop tree of Figure 1, whereas only 6 messages are forwarded along the carefully selected right tree of the same figure. The right routing tree is better not only in terms of total communication cost, but also in terms of communication cost in the critical area around the gateway. Informally, the benefit of the second plan is that it aggregates all readings of a query early and avoids sending different subaggregates through disjoint paths.

Figure 2 illustrates the benefits of building a suitable execution plan in the case of processing multiple *count* queries. For ease of understanding the graphs also include node ids and messages forwarded through network links. Messages have the format $v(q_1, \ldots, q_n)$, which denotes that value v contributes to queries $q_1, \ldots, q_n$. The left plan does not exploit query commonalities, and therefore fails to aggregate together readings (of nodes 8 and 9) within the intersection area. The middle plan incurs smaller communication cost, because it exploits query commonalities, but still forwards the subaggregate of the intersection area separately all the way to the gateway. This behavior is similar to our first heuristic called SegmentToGateway (STG). The right plan has optimal behavior because it exploits query commonalities and it avoids sending partial aggregates through long disjoint paths. Notice that the optimal plan does not follow a tree structure, as node 8 sends the partial aggregate of the intersection area to two parents. The intersection partial aggregate is thus merged immediately with the other two query subaggregates and, eventually, only two partial results are sent to the gateway. This would be the plan identified by the second proposed algorithm, called SegmentToSegment (STS).

3 Conclusions

The examples above illustrate the interplay of routing and processing in evaluating aggregate queries in sensor networks. STG exploits the fact that the intersecting query areas naturally divide the network into smaller segments, and partially aggregates all values within the same segment before forwarding the partial aggregate to the gateway. In addition, STS ensures that segment aggregates are not propagated to the gateway through disjoint paths. Unlike competing algorithms (including STG), which focus on merging data to reduce their size, STS's novelty lies in identifying cases where data splitting provides a greater potential for later merging.

4 References

[1] C. Intanagonwiwat, R. Govindan, and D. Estrin. Directed diffusion: a scalable and robust communication paradigm for sensor networks. In *Mobicom*, 2000.

[2] S. Madden, M. Franklin, J. Hellerstein, and W. Hong. TAG: a tiny aggregation service for ad-hoc sensor networks. In *OSDI*, volume 1, pages 131–146, 2002.

[3] N. Trigoni, Y. Yao, A. Demers, J. Gehrke, and R. Rajaraman. Multi-query optimization for sensor networks. In *DCOSS*, pages 307–321, 2005.

[4] M. Woo, S. Singh, and C. S. Raghavendra. Power-aware routing in mobile ad hoc networks. In *Mobicom*, pages 181–190, 1998.

[5] C. Yu, R. Govindan, and D. Estrin. Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks. TR UCLA/CSD-TR-01-0023, Univ. of Southern California, 2001.