Exposé zur Studienarbeit
Distributed Calculation of Local Averages in Sensor Networks

betreut von: Timo Gläßer, Prof. Ulf Leser
bearbeitet von: Björn Schümann

April-Juni 2008

1 Motivation

In recent years Wireless Sensor Networks have become an important and useful tool used for a wide area of applications. WSN consist of a bulk of sensor nodes, i.e., small battery powered devices equipped with a set of sensors and a wireless communications interface. WSNs have been used in applications like monitoring the habitats of birds [SPMC04], the structural health of buildings [KPC+07] or the areas near eruptive volcanoes [WAJR+05].

Since an increasing number of sensor nodes provides a finer resolution of data points the networks are getting bigger. Sharing workload between nodes provides opportunities to extend the lifetime of the network and reduces the risk of data loss in case of failing nodes. Previous solutions tried to execute most of the processing workload at the gateway node. This often leads to a congestion of workload on some nodes, resulting in a high energy consumption on these nodes. This becomes unfeasible and inefficient in growing networks. As a consequence the processing of the raw data in the nodes becomes more and more important. At the same time the nodes are build on hardware platforms that have more processing power and memory, making query processing in the nodes feasible.

Large-scale networks have to be flexible enough to handle changes in the network topology, caused for example by failing and moving nodes and require the ability to run multiple queries from different users distributed over different gateway nodes. There is still a need for better tools providing a simple way of programming networks that fulfil these requirements for data gathering and monitoring tasks.

Finding an efficient way of answering queries over data in a specific subarea in these networks is one of the challenges. With an increasing number and density of nodes the single sample of a single node gets less and less important. Integrating the samples of a group of nodes often improves the data quality. Consequently aggregating and averaging over the data of the nodes surrounding a specific location is a common and important but expensive task. The query needs to be routed on a direct path to the area in question. The nodes located there have to work together to process the data, before the answer can be send back. There are a lot of ways to route, spread and process the query and finding a efficient and flexible solution is an interesting and demanding problem.

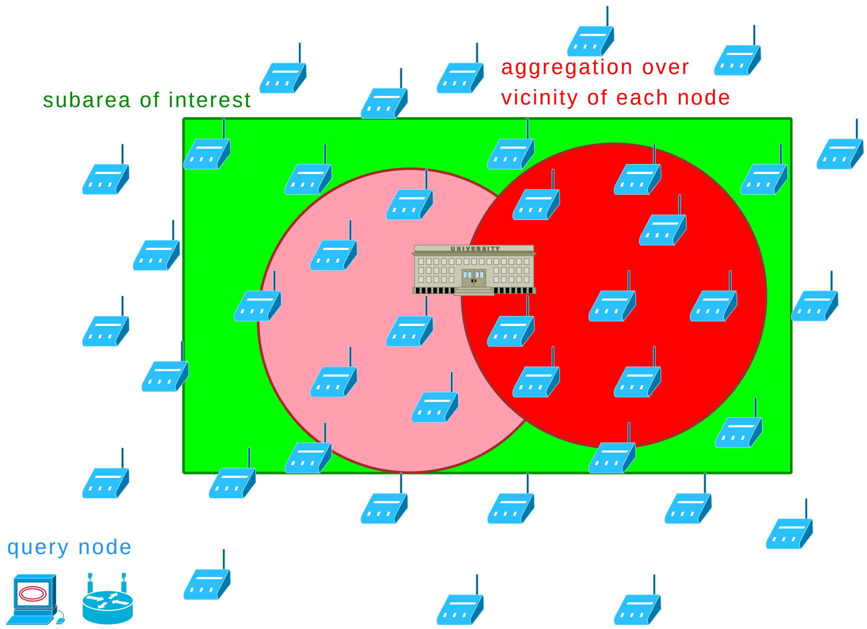

Figure: Aggregation over the vicinity of each node to improved data quality

2 Background

Various architectures have been proposed as a Middleware for Sensor Networks. They try to close the gap between the users questions and applications and the low level programming of the sensor devices. The ideas range from special languages [MWM06] and compilers to virtual machines running a specific byte code interpreter[LC02].

Cougar [YG02] and later TinyDB [MFHH03], introduced a new dimension into Middleware research by proposing a database approach. Instead of writing embedded C Code the user can use a SQL-like interface to extract the data from the network. Senor readings are represented in a virtual database table. A gateway node parses SQL-like queries and distributes an optimized query execution plan to all nodes. Each sensor node has its own query processor and processes the sensor data before sending his part of the answer to the parent node. TinyDB introduced many important ideas for query optimization in sensor networks, but the project was finished in 2003. The project is only weakly documented, very difficult to port to new sensor platforms and still based on the existing code TinyOS1.x, the first release of an operating system designed for wireless embedded sensor networks [MA05] The central approach of parsing and optimization in the gateway node restricts the optimization to single queries. Multiple queries are processed independently including duplicate sampling of sensors and transmitting of results.

Other approaches base the communication in sensor networks on events[JSS05]. In most applications data transmission is either triggered by an event in a sensor node or by a query in a gateway. The publish/subscribe paradigm seems to be the right approach to handle this kind of data dissemination. A publisher provides information to subscribe in an asynchronous fashion.

3 Objective

A publish/subscribe query processing system should provide the basis to run distributed aggregation queries in a specific subarea of the network and return the answer to all nodes subscribed to this query.

The objective of this paper is to develop and compare different execution plans for queries testing the local average around each node in a specific area of the network. Testing an aggregation result would be a **HAVING**-clause in an SQL-Query. One example of such a query would be the test whether there exists a node with an high local average temperatur in its vicinity.

```
SELECT A.LOC, AVG(B.TEMP) FROM Nodes A, Nodes B
WHERE A.LOC IN (1000,1000,2000,2000)
  AND B.LOC IN (1000,1000,2000,2000)
  AND DIST(A.LOC, B.LOC) < 100
GROUP_BY A.LOC
HAVING AVG(B.TEMP) > 50
```

Example query: Testing the average temperature 100m around each node in a specifc subarea of the network

Testing local averages or other aggregation directly in the subarea of interest could reduce the transmissions to the gateway node, but computing an aggregated average from different nodes demands a lot of transmissions between the nodes in the area. Finding an way to get the results for each node with less transmissions, saves energy and therewith extends the lifetime of each node.

4 Procedure

The first task will be a theoretical evaluation of possible execution plans. Their performance and the expected number of transmission should be analysed to under different scenarios. The goal is to get a feel for the behaviour one should expect from the solutions under various circumstances and parameters. (4 weeks)

After that the different approaches will be implemented in a prototype using JiST/SWANS to simulate a sensor network executing such queries. JiST/SWANS is a scalable wireless network simulator built atop of a discrete event simulation engine.[BHvR05] (4 weeks)

The simulation the different approaches in various scenarios will verify the theoretical results of part one. A discussion of the results should help optimize those queries and provide the information needed to improve automatic query planing. (2 weeks)

Writing a report about the ideas, procedures and results. (2 weeks)

References

[BHvR05] Rimon Barr, Zygmunt J. Haas, and Robbert van Renesse. Scalable wireless ad hoc network simulation. In Jie Wu, editor, *Handbook on Theoretical and Algorithmic Aspects of Sensor, Ad hoc Wireless, and Peer-to-Peer Networks*, chapter 19, pages 297–311. CRC Press, 2005.

[JSS05] Binjia Jiao, Sang H. Son, and John A. Stankovic. Gem: Generic event service middleware for wireless sensor networks. In *Proceedings of the Second International Workshop on Networked Sensing Systems (INSS05)*, June 2005.

[KPC+07] Sukun Kim, Shamim Pakzad, David Culler, James Demmel, Gregory Fenves, Steven Glaser, and Martin Turon. Health monitoring of civil infrastructures using wireless sensor networks. In *IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks*, pages 254–263, New York, NY, USA, 2007. ACM Press.

[LC02] Philip Levis and David Culler. Mat: a tiny virtual machine for sensor networks. In *ASPLOS-X: Proceedings of the 10th international conference on Architectural support for programming languages and operating systems*, pages 85–95, New York, NY, USA, 2002. ACM Press.

[MA05] Rene Müller and Gustavo Alonso. A query processing middleware for sensor networks. Master's thesis, ETH, Zürich, CH, 2005.

[MFHH03] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, and Wei Hong. The design of an acquisitional query processor for sensor networks. In *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 491–502, New York, NY, USA, 2003. ACM.

[MWM06] Geoffrey Mainland, Matt Welsh, and Greg Morrisett. Flask: A language for data-driven sensor network programs. Technical Report TR-13-06, Harvard Univ., Boston, MA, USA, 2006. Harvard Tech Report TR-13-06.

[SPMC04] Robert Szewczyk, Joseph Polastre, Alan Mainwaring, and David Culler. Lessons from a sensor network expedition. In *Proceedings of the First European Workshop on Sensor Networks (EWSN)*, pages 307–322, jan 2004.

[WAJR+05] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh. Monitoring volcanic eruptions with a wireless sensor network. In *European Workshop on Sensor Networks*, pages 108–120, 2005.

[YG02] Yong Yao and Johannes Gehrke. The cougar approach to in-network query processing in sensor networks. *SIGMOD Rec.*, 31(3):9–18, 2002.