A FRAMEWORK FOR QUERY PROCESSING OVER HETEROGENEOUS LARGE SCALE SENSOR NETWORKS

Chung Ho Lee, Min Soo Kim, Yong Joon Lee

Electronic and Telecommunications Research Institute (ETRI), {leech, minsoo, yjl}@etri.re.kr

ABSTRACT: Efficient Query processing and optimization are critical for reducing network traffic and decreasing latency of query when accessing and manipulating sensor data of large-scale sensor networks. Currently it has been studied in sensor database projects. These works have mainly focused on in-network query processing for sensor networks and assumes homogeneous sensor networks, where each sensor network has same hardware and software configuration. In this paper, we present a framework for efficient query processing over heterogeneous sensor networks. Our proposed framework introduces query processing paradigm considering two heterogeneous characteristics of sensor networks: (1) data dissemination approach such as push, pull, and hybrid; (2) query processing capability of sensor networks if they may support in-network aggregation, spatial, periodic and conditional operators. Additionally, we propose multi-query optimization strategies supporting cross-translation between data acquisition query and data stream query to minimize total cost of multiple queries. It has been implemented in WSN middleware, COSMOS, developed by ETRI.

KEY WORDS: Query Processing, Query Optimization, WSN Middleware

1. INTRODUCTION

Due to advances in wireless sensor networks over the last few years, the number of areas of WSN applications is growing rapidly such as environmental monitoring, military applications, health monitoring, civil engineering, etc. And the scale of WSN applications is also getting gradually larger to cover the range of a city. Major challenges in large-scale WSN applications are how to integrate data gathering from heterogeneous sensor networks and how to minimize the cost of deployment. To ease the integration and deployment, a large amount of research has been conducted on WSN middleware which accommodates heterogeneous sensor network resources and provide an application programming interface for developing various applications. Among many WSN middlewares, TinyDB and Cougar adopt database approach that treats sensor readings as a relational database tables. And it provides SQL-like declarative query language for users and has to process multiple user queries concurrently.

Like in relational database systems, efficient query processing and optimization in the WSN middleware are critical for reducing network traffic and decreasing latency of query when accessing and manipulating sensor data on large-scale sensor networks. Until now, there has been a lot of research effort on query processing. These works, however, focuses on in-network query processing for sensor networks and assumes homogeneous sensor networks that have same hardware and software configuration.

In this paper, we present a framework for efficient query processing over heterogeneous sensor networks. Our framework introduces query processing paradigm considering two heterogeneous characteristics of sensor networks: (1) data dissemination approach such as push, pull, and hybrid approach; (2) query processing capability of sensor networks if they may support in-network aggregation, spatial, periodic, and conditional operators. Here, the acquisition query is fit for pull or hybrid based sensor network, on the other hand, the data stream query is suitable for push based sensor networks. Therefore, our framework provides both data acquisition query and stream query. So users of our framework don't need to worry about dissemination approach and capability of sensor networks when they want to issue a query. Additionally, this paper proposes multi-query optimization strategies supporting cross-translation between data acquisition query and stream query in order to minimize total cost of multiple queries. It has been implemented in WSN middleware, COSMOS, developed by ETRI.

The rest of this paper is organized as follows. Section 2 introduces the related work and background information on query processing and optimization in wireless sensor networks. In section 3, we describe basic assumption and presents query processing framework in detail. Section 5 presents a multiple query optimization strategy. In Section 6, we introduce our WSN Middleware, COSMOS, where our proposed framework is implemented. And section 6 concludes the paper.

2. RELATED WORK

In this section, we first describe the characteristics of heterogeneous sensor networks and then we introduce a few existing studies on query processing techniques over heterogeneous WSN.

2.1 Heterogeneousness of WSN

In heterogeneous sensor networks, each sensor network has different data dissemination approach and processing capability. At first, the sensor networks can be classified into three categories: pull, push, and hybrid approaches. As shown in figure 1, the pull approach enables the interested sensor nodes to deliver their data only when receiving an explicit request. In the push approach, individual sensor nodes proactively propagate their data without a request. The hybrid approaches combine the advantage of both push and pull. Secondly, the sensor networks can be classified by query processing capability. That is, some sensor networks may support conditional, in-network aggregate and spatial operators, other sensor networks may only execute the sensing function.

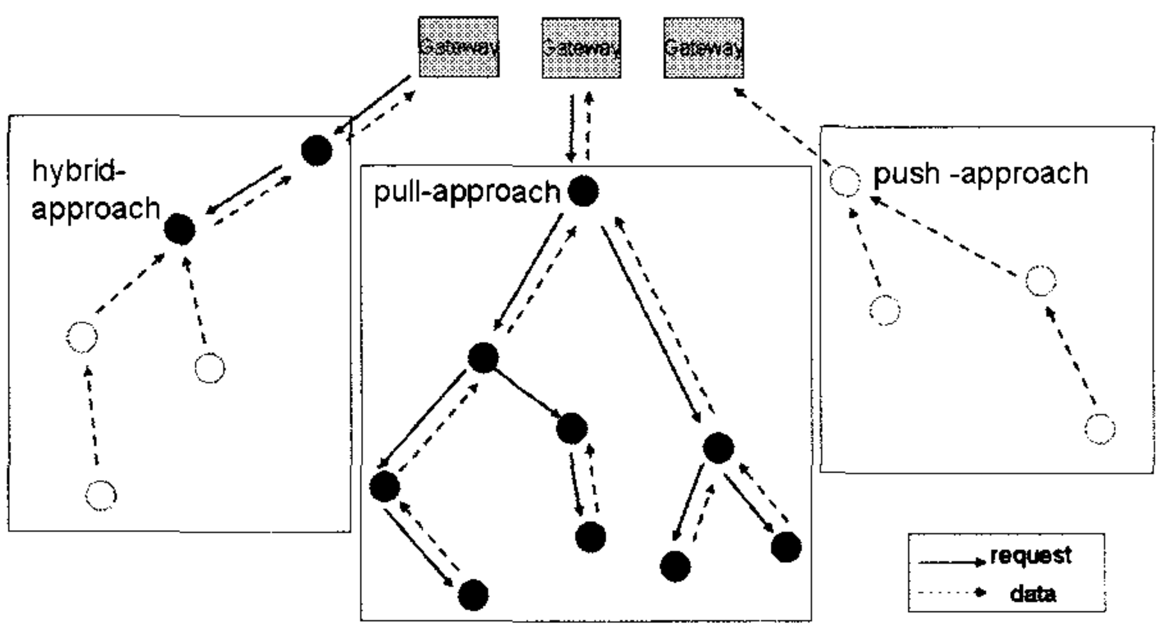


Figure 1. Three kinds of sensor network according to the data dissemination approach.

2.2 Query processing in WSN

The query processing problem in WSN has long been studied by the database community. The sensor network query-processing systems (SNQPs) such as TinyDB and Cougar established the foundation of query processing in WSN. And they also proposed a few query optimization algorithms. However these works mainly focused on processing acquisition queries over homogeneous sensor networks that have a characteristic of pull dissemination approach.

As the closest work to ours, there has been two works. The first one is an integration framework for sensor networks and data stream management system (DSMS) proposed by Abadi et al. Although they deal with processing of acquisition query and data streaming across heterogeneous components: SNQP and DSMS, the heterogeneous characteristics of WSN are not addressed. In other words, it assumes that all of sensor network are managed by SNQP and it supports integration of homogeneous sensor networks. Secondly, Aberer et al. have developed Global Sensor Networks middleware, called GSN, and proposed a query processing mechanism based on GSN. The GSN aims at efficient infrastructure for large-scale heterogeneous sensor networks consisting of a sensor network and a RFID reader, a few wireless network cameras and so on. However, query processing of GSN is focused on the data stream query processing in push based sensor networks.

2.3 Multiple Query Optimization in WSN

Many WSN applications often need to process multiple user queries simultaneously. Consequently, WSN middleware should support optimization of these multi-queries. There has been related works, where most work has focused on the optimization of a single long running query. And only a few works deal with this problem.

Trigoni et al. proposed multi-query optimization for aggregate queries on sensor networks. Silbersten et al. exploited the combination of multicast and in-network optimization to optimize many to many aggregations. Xiang et al. proposed two-tier multiple query optimization (TTMQO) scheme. Within this method, the first rewrites a set of queries into an optimized set that shared the commonality and eliminates the redundancy among the queries. And the second tier delivers query results by sharing the sensor readings among similar queries.

Besides these works, multi-query optimization has also been studied in the context of DSMS, where the application context is different from ours.

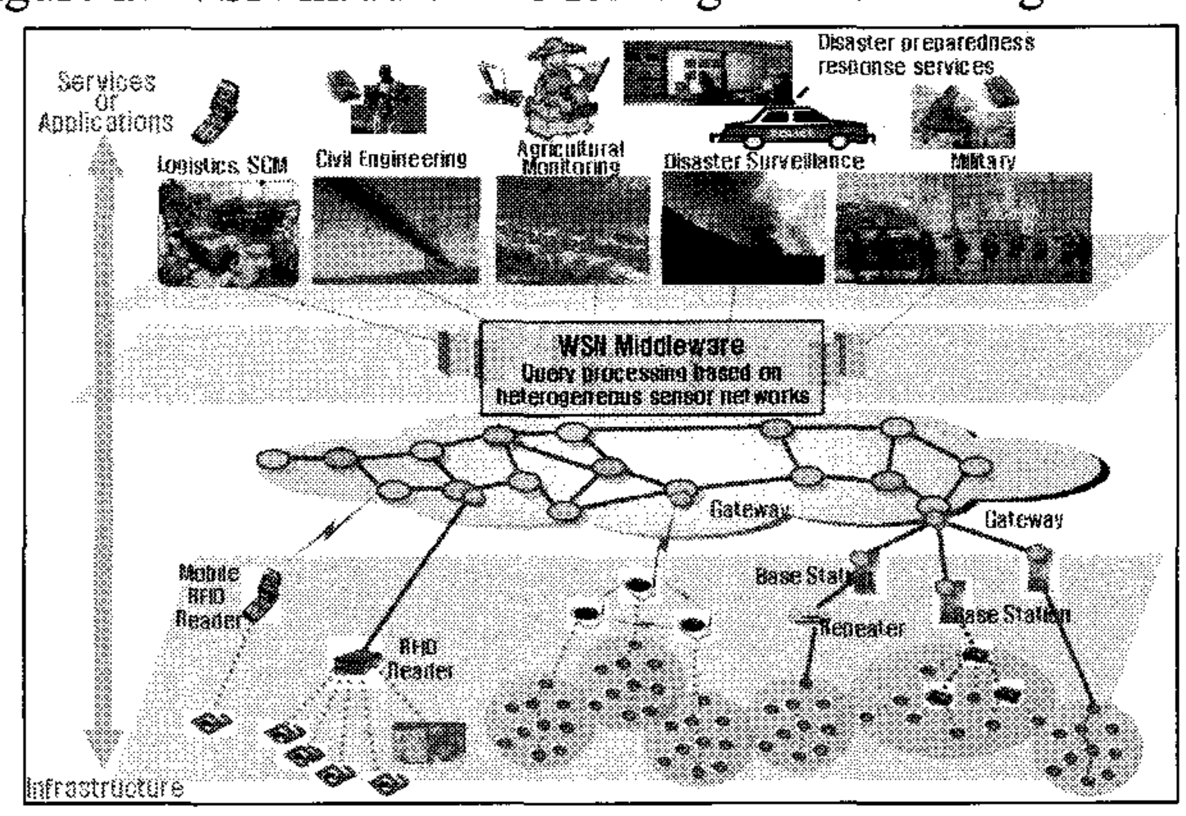
3. FRAMEWORK FOR QUERY PROCESSING

In this section, we define first some assumptions. Next, we present a framework for processing various queries.

3.1 Our model and Assumptions

To make a large-scale sensor network that can cover a target area such as city, it is necessary to integrate several heterogeneous sensor networks. Here, we assume heterogeneous sensor networks that have different characteristics in regard to both data dissemination approach and query processing capability described in section 2. Additionally, we assume that each sensor network is connected to WSN middleware through its gateway or adaptor. The role of WSN middleware is to provide development, maintenance, and execution of WSN applications. Especially, the WSN middleware should provide the mechanism for processing various queries issued from WSN applications as shown in figure 2.

Figure 2. WSN middleware for large-scale heterogeneous



sensor networks.

3.2 Framework for Query Processing

Given various queries from multiple applications, the problem is how to efficiently process the corresponding queries without having to worry about the characteristics of sensor networks. That is, users of applications hope that all kinds of queries such as conditional query, periodic query, event query, aggregation query, spatial query, data acquisition query, and data stream query can be successfully performed by WSN middleware regardless of data dissemination approach and capability of sensor networks.

To solve this problem, we propose a general framework that provides a query interface for processing both data acquisition query and data stream query without having to worry about characteristics of sensor networks. Within our framework, there are two main mechanisms for query processing.

First, our framework provides translation mechanism from the data acquisition query to the data stream query when the target sensor network only supports push dissemination approach. On the other hand, if the target sensor network support pull or hybrid dissemination approach, the translation is performed from data stream query to data acquisition query. Of cause, it's on the assumption that two queries are identical with each other. An example of such a query is "Get the sensor node's identifier and temperature from a sensor network once per second 10 seconds". This data acquisition query can be expressed by data stream query. As shown in figure 3, Q1 and Q2 are equivalent expressions and can be translated each other according to the data dissemination approach of sensor networks.

```
Q1: select id, temperature
    from sn1
    period 10s
```

```
Q2: select id, temperature
    from sn1[range 10s]
```

Figure 3. Two Equivalent Expressions for (a) Q1: Data Acquisition Query and (b) Q2: Data Stream Query.

Second, our framework also provides mechanism for processing various types of queries: conditional, periodic, event, spatial and aggregation query even if capability of the sensor network can't afford to execute corresponding query. It translates the original query into substitute query that can be executed and processed in the target sensor network. For example, if the periodic query can't be executed in a sensor network, the periodic query is translated into a snapshot query, where our framework performs execution of snapshot query as many iterations as are necessary to be identical with the periodic query. And it sleeps for 10 seconds to keep the given interval.

```
Q3: select id, temperature
    from sn1
    period 10s
```

```
Q4: do {
      select id, temperature
      from sn1
      wait (10s)
    } while (1)
```

Figure 4. Two Equivalent Expressions for (a) Q3: Periodic Query and (b) Q3: Snapshot Query.

4. MULTI-QUERY OPTIMIZATION

This section describes a multi-query optimization strategy suitable for our query processing framework.

Given multiple queries, the goal of optimization is to reduce the network traffic of sensor networks and efficiently conserve the energy consumption. Additionally, it is also important to decrease latency of query and finally improve performance. To achieve this goal, we adopt two strategies.

The first is to minimize the execution frequency of data acquisition queries by translating a part of them into data stream query as much as possible while guaranteeing the equivalence between data acquisition query and data stream query. Figure 5 shows the example of multi-query optimization applying the first strategy. There are two data acquisition queries like Q5and Q6. As there is redundancy between these queries, the Q6 can be replaced by Q6'. Consequently, we can cut down the total cost of two queries by half.

```
Q5: select id, temperature
    from sn1
    period 10s

Q6: select id, temperature        ⇨   Q6': select id, temperature
    from sn1                               from sn1 [range 20s]
    period 20s
```

Figure 5. An Example of the First Optimization Strategy.

The second is to make the most use of sensor network capability as much as possible while ensuring the correctness of all queries and improving system throughput. That is, there are some operators such as spatial, aggregation or conditional operator that can be moved from data stream query to data acquisition query using capability of sensor networks. Figure 6 shows the example of multi-query optimization applying the second strategy. In figure 6, there is a data stream query, Q7 having a conditional operator. If the target sensor network has capability to process the conditional operator, it is more efficient way that the sensor network is responsible for processing the corresponding operator by translating the data stream query into the data acquisition query like Q7'. However, if there is another data stream query having different condition operator such as "temperature < 5", the above translation can't be allowed.

```
Q7: select id, temperature          ⇨   Q7': select id, temperature
    from sn1[range 10s]                       from sn1
    where temperature > 10                    where temperature > 10
                                              period 10s
```

Figure 6. An Example of the Second Optimization Strategy.

5. IMPLEMENTATION

This section introduces the implementation of our framework called SIMC (Sensor Information Management Component) belong to the COSMOS system developed by ETRI. Figure 7 shows the architecture of the SIMC composed of four modes.

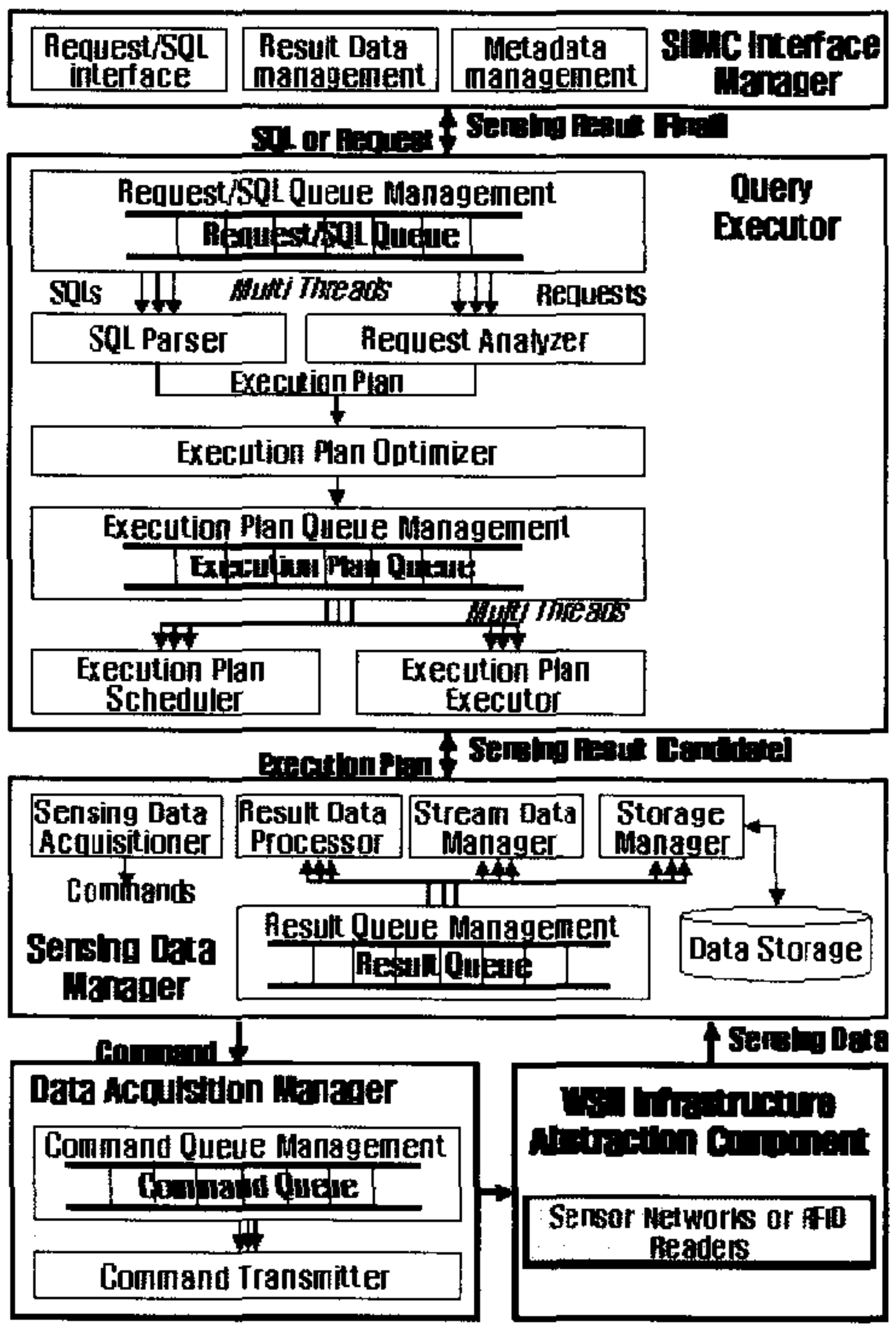


Figure 7. System Architecture of the SIMC.

The SIMC interface manger provides both API and SQL interface to WSN applications. The query executor manages synchronized queues for the queries issued from applications and uses multi-threads pool to maximize the performance. And it also generates optimized execution plans for the given queries, where our query optimization strategies are applied. The execution plan scheduler controls and executes the queries according to the capability of sensor network. The sensing data manager controls all data sources that may be used for running execution plan of the query executor. Here, data source is not only pull-based sensor network, but also stream data from push-based sensor network. The result queue manager asynchronously receives sensor data for the query executor. The data acquisition manager translates the execution plan of SIMC into the standardized command packets that the real sensor network can recognize.

6. CONCLUSION

In this paper, we have presented a framework for efficient query processing over heterogeneous large-scale sensor networks. Our proposed framework introduced query processing paradigm considering two heterogeneous characteristics of sensor networks: data dissemination approach and query processing capability of sensor networks. Our framework provided an interface for both data acquisition query and stream query. Users don't need to worry about dissemination approach and capability of sensor networks when they want to issue a query. Additionally, we proposed multi-query optimization strategies supporting cross-translation between data acquisition query and stream query in order to minimize total cost of multiple queries. Finally, our proposed framework was implemented in WSN middleware, COSMOS, developed by ETRI.

References from Journals:

Samuel R. Madden, M., W., 2005. TinyDB: An Acquisitional Query Processing System for Sensor Networks. *ACM Transactions on Database Systems*, Vol. 30, No.1, pp. 122-173.

References from Other Literature:

Daniel J. Abadi, W., S., J., 2004. An Integration Framework for Sensor Networks and Data Stream Management Systems. *In Proceedings of the 30th VLDB Conference*. Toronto, Canada. pp. 1361-1364.

Karl Aberer, M., A., 2006. The Global Sensor Networks Middleware for Efficient and Flexible Deployment and Interconnection of Sensor Networks. Technical Report, LSIR-2006-006.

Wolfgang Lindner, H., K., 2006. Data Stream Query Optimization across System Boundaries of Server and Sensor Network. *In Proceedings of the 7th International Conference on Mobile Data Management.*

P. Roy, S. Seshadri, S. Sudarshan, S. Bhobe, 2000. Efficient and Extensible Algorithmss for Multi Query Optimization. *In Proceeding of SIGMOD 2000.*

A. Silberstein, J. Yang, 2007. Many-to-Many Aggregation for Sensor Networks. *In ICDE 2007.*

U. Strivastava, K. Munagala, J. Widom, 2005. Operator Placement for In-network Stream Query Processing, *In Proceeding of PODS 2005.*

N. Trigoni, Y. Yao, J.gehrke, R., A., 2005. Multi-query Optimization for Sensor Networks. *In DCOSS 2005.*

Shili Xiang, H., K., Y., 2007. Two-Tier Multiple Query Optimization for Sensor Networks. *In Proceeding of ICDCS 2007.*

Lei Xie, C., L., Xie., Daoxu., 2006. Energy Efficient Multi-query Optimization over Large-Scale Sensor Networks. *In Lecture Notes in Computer Science*, Springer-Verlag, Vol. 4138, pp. 127-139.