Research on Practical Wireless Sensor Network Query Processing Technology

Changjian Guo\(^1\), Xiaoxuan Wu\(^{1,2,*}\)

\(^1\)School of Artificial Intelligence and Big Data, Hefei University, Hefei, China
\(^2\)Anhui Engineering Lab of Big Data Technology Application for Urban Infrastructure, Hefei University, Hefei, China

*Corresponding author e-mail: kexinyufan@163.com

Abstract. WSN (Wireless Sensor Network) is intended for obtaining sensory data so that the query processing technology of high efficiency is the precondition of widely used WSN. Existing query processing technology assumes that the WSN deployed in two-dimensional or three-dimensional free space are not suitable for those deployed in such limited circumstances as roads, pipelines and indoor environments. This paper explores the problems of existing query processing technology, does research on perception data query processing method and technology, and seeks out the existing problem and its solution by means of using the WSN actually deployed in a building as an experimental platform.

1. Introduction

With the rapid development and maturity of embedded technology and wireless communication technology, sensor nodes have the ability of sensing, computing, storage and wireless communication, enabling sensor nodes to be organized into wireless sensor networks in many ways, effectively sensing, monitoring and Handle various information in the network monitoring area and provide a large amount of useful information to users. Therefore, wireless sensor networks have broad application prospects in the fields of environmental monitoring, national defense construction, smart home and intelligent transportation. In recent years, they have been a hot topic of research\(^{[1]}\).

Unlike other networks, wireless sensor networks are networks for the purpose of obtaining sensory data, such as collecting water quality and pollution data for a certain area of a river. The wireless sensor network data query system provides users with a simple, easy-to-use, SQL-like sensing data query interface. Users can query sensor data just like traditional relational database systems, which greatly reduces the difficulty of wireless sensor network application development\(^{[2]}\).

Although the traditional database query processing technology has achieved good research results, these research results are mainly in the PC environment, and are not suitable for wireless sensor network environment. In recent years, many scholars have considered the characteristics of limited sensor nodes, frequent changes in network topology, and easy failure of nodes. A new query processing algorithm for sensor networks is proposed. For example, the TinyDB system designed by the University of California at Berkeley. Therefore, it is very useful to study energy-efficient and resource-limited query processing techniques\(^{[3]}\).
2. **Query Processing System Structure**

Figure 1 shows the structure of a wireless sensor network query processing system. It consists of two main components: query processing software (DBMS) running on a base station or computer and software running on the sensor side. The main functions of the former part are query parsing, checking, query optimization and distributing the optimized query results to the sensor network in a "multi-hop" routing manner, then collecting the sensing data and returning the data to the base station or "multi-hop" routing. On the computer. The main function of the latter part is to perform operations such as query receiving, processing, sampling and communication. It is a component of a group.

![Figure 1. Wireless sensor network query processing system structure](image)

3. **Query processing key technology**

Wireless sensor network query processing technology is based on data model, query language and data storage. Its key technologies mainly include the following points.

3.1 **Data model and data storage**

The data model of the wireless sensor network data management system is an extension of the traditional relational data model. It sees the data perceived by the sensor network as an ever-increasing list. The attributes of the data in the table mainly include two types: the first type is the data sensed by various sensors (such as temperature and pressure); the second type is the description of the attributes of the perceived data, such as the time and location of the perceived data acquisition. The data generated by each sensor node corresponds to a row in the table.

There are three main methods of sensor network data storage: external storage, local storage, and data-centric storage.

3.2 **Query language and query processing**

At present, most of the data query languages of wireless sensor networks continue the traditional SQL language form and extend the SQL language. Typical is the query language of TinyDB.

There are several methods for query processing:

a) Data aggregation query;

b) Data collection query;

c) Time and space range query;

d) K Nearest Neighbors (KNN, K Nearest Neighbors).

3.3 **Prototype System.**

The most representative of the wireless sensor network data query prototype system is the TinyDB system developed by the University of California at Berkeley. TinyDB studies wireless sensor networks from the perspective of relational databases and virtualizes the sensory data of the entire sensor network into a database system. Data communication is performed using the IEEE 802.15.4 standard, and intra-network query processing is supported. After the user submits the SQL query request to TinyDB, TinyDB sends the user's query request to each node. Then the results returned by
each node are merged and returned to the user, and the user does not need to know the underlying
details of the sensor network, such as node communication and failure\textsuperscript{[6,7]}.

4. Query processing research problems

The existing research on wireless sensor network query processing technology mainly has the
following problems:

(1) Strong assumptions have been made in wireless communication links, network topology, node
deployment, etc., which are inconsistent with the actual situation

(2) The existing data collection and aggregation algorithms assume that the network deployment
area is an unconstrained space, ignoring the mutual interference of different routing paths and the
self-interference of a single routing path. For wireless sensor networks deployed in confined spaces
such as roads, pipelines, coal mines, indoors, etc., the number of network routing paths in confined
spaces is small, which worsens the competition of wireless channels and exacerbates message
collisions\textsuperscript{[8]}.

(3) The existing K-nearest neighbor query processing algorithm assumes that the wireless sensor
network node communication model is an ideal disk model, that is, if the node's communication radius
is \( r \), when the distance between nodes \( n_1 \) and \( n_2 \) is \( D(n_1, n_2) > r \), \( n_1 \) The probability of successfully
transmitting a message packet to \( n_2 \) \( P(n_1, n_2) = 0 \); when \( D(n_1, n_2) \leq r \), \( P(n_1, n_2) = 1 \). In the actual
sensor network, the value of \( P(n_1, n_2) \) is in the interval \([0,1]\), and the quality of the communication
link of the node is often dynamically changed\textsuperscript{[9,10]}.

(4) Existing spatio-temporal and K-nearest query processing algorithms are not suitable for
wireless sensor networks deployed in restricted environments because the distance measurement
functions of unrestricted and restricted deployment spaces are different, and the unconstrained space
uses Euclidean distance. The confined space uses network distance.

(5) There is a cut-point detection method based on depth-first search in graph theory, but simply
moving the method to the sensor network is not effective. The reason is that the method needs to
traverse all sensor nodes in depth, and the cut-point detection delay is large.

(6) The existing wireless sensor network data query system only supports simple data aggregation
and collection of queries, but is inefficient in a limited space, and does not support spatio-temporal
range query, K-nearest neighbor query, and cut-point query.

5. Main research contents and treatment methods

Aiming at the problems existing in the existing query processing research, a TelosB-based sensor
hardware platform is built to support sensors including temperature, humidity and pressure.
Specifically, 50 sensor nodes are deployed on three floors of a building to monitor the environment
information of the building, perform personnel positioning and scene perception to construct an
experimental platform. On the experimental platform, the existing data collection query algorithm and
the wireless sensor network data query system TinyDB were tested. After the experiment, the
following phenomena were found: Firstly, although the bandwidth of the wireless communication link
can theoretically reach 256KB/S, The existing query processing algorithm returns the query result to
the base station and the throughput rate does not exceed 2 KB/S; Secondly, after the query processing
algorithm runs for a period of time, the sensing data of some nodes in the network cannot be returned
to the base station; the phenomenon of the phenomenon 1 occurs that the existing query processing
algorithm assumes that the network is deployed in the unconstrained space compared to the
unrestricted environment. In a wireless sensor network, in a restricted network, the sensor node has
fewer data forwarding paths. In an extreme case, only one routing path returns the query result (such
as a sensor node deployed in a corridor), so that the wireless communication channel competes.
Intensified, the probability of packet collision increases, and the throughput of query result
transmission decreases. Symptom 2 occurs because there are cut points in the network. Due to factors
such as poor deployment environment, software failure, and insufficient power, node failures occur
frequently, and the failure of the cut point causes the network to be disconnected.

To solve the above phenomenon, we must study and deal with the following aspects:

(1) Bad topology query technology for wireless sensor networks.
For the wireless sensor network cut-point query, there is a cut-point detection method based on depth-first search in graph theory, but this method requires deep traversal of all sensor nodes, and the cut-point detection delay is large. In addition, the method needs to be in the network. Each node performs a cut point determination, which is not efficient. It is proposed to use the first filtering and accurate detection query technology, that is, first design the wireless sensor network cut-point filtering strategy, and divide the sensor nodes into two categories: nodes that may become cut points and nodes that cannot become cut points, and then may become cuts. The node set of points performs cut point detection. There are three main methods of sensor network data storage: external storage, local storage, and data-centric storage.

Specifically, since the leaf nodes of the tree cannot be cut points, the network is organized into a tree topology, the sensor nodes that become leaves are filtered out, and only the non-leaf nodes are subjected to cut point detection. In addition, since wireless sensor networks can organize multiple trees, in order to maximize the number of filtering nodes, how to find the tree with the largest number of leaf nodes in these trees becomes a key issue. This problem is equivalent to constructing the minimum connected dominating set problem, and it is proposed to solve it by the theory and algorithm of the minimum connected dominating set. In addition, in order to reduce the time of the cut point detection, it is proposed to use the breadth-first search to detect the cut point.

(2) Data collection and aggregation query algorithm for wireless sensor networks in an efficient and restricted environment.

The research idea of the existing wireless sensor network data collection and aggregation query algorithm is to construct a routing path with low packet loss rate to improve the throughput of the query result returned to the base station and reduce energy consumption. However, since the order of sending and receiving the sensing data is not effectively planned, the mutual interference of different routing paths and the self-interference of a single routing path are neglected, resulting in serious congestion of the query result and high packet loss rate. In a limited wireless sensor network, the data transmission path is reduced, which further degrades performance. There are several methods for query processing.

Specifically, many effective link quality estimation algorithms have been proposed. Based on the link quality information calculated by these algorithms, the correspondence between communication link quality, routing path and energy consumption is found. The problem is to optimize and solve the problem difficulty analysis; the existing wireless sensor network communication module has multiple channels, different channel interference is small, design node wireless communication channel planning strategy, set different channels for neighboring nodes, thus Reduce wireless communication interference.

(3) Space-time query algorithm for wireless sensor networks in an efficient and restricted environment.

The existing location routing protocol for the two-dimensional and three-dimensional free space is improved. For the sending node located near the fork, the sending node calculates the next hop forwarding node, first selects the forwarding road according to the network deployment space information, and then sends the path. The node calculates its set of neighbor nodes on the forwarding road, and selects a neighbor node with the lowest cost according to the network distance of the neighbor node from the destination node, the link quality of the sending node and the neighbor node. The spatio-temporal range and K-nearest neighbor query messages will be sent using the above-mentioned routing protocol applicable to the confined space.

The following method is used to process the spatio-temporal range query: as shown in Figure 2, the rectangular EFGH is the query area, and S is the base station node. In order to return the sensing data of all nodes in the query area EFGH to S, these nodes are organized into several groups, and each group is assigned a sink node (such as nodes a and b in Figure 2), and the sensing data of the nodes in the group is Collected to the aggregation node, and then returns the query result to the base station. The energy consumption of the above process is deduced to obtain the energy-optimized grouping node setting method and grouping strategy.
6. Conclusions

Query processing is an important part of wireless sensor networks. In some special applications, there are advantages that traditional technologies do not have. Although some achievements have been made, there are still many problems, which still need deeper levels of research and discussion. This paper abstracts the wireless sensor network into a distributed database to study and design an efficient cut-point query algorithm to provide good decision support for network redeployment and performance diagnosis. At the same time, the node data transmission and reception scheduling strategy for the limited deployment space is proposed to improve the throughput of data aggregation and data collection and query algorithms, so that the wireless sensor network has a wider application space and better adaptability.

7. Acknowledgement

This work was partially supported by grants from the Key University Science Research Project of Anhui Province(Nos. KJ2019A0836) and the Talent research project of Hefei University(Nos. 16-17RC16). The authors would like to thank the comments and suggestions from the reviewers.

References
[1] F.Y Ren, H.N Huang, W. Lin, Wireless Sensor Networks, Journal of Software. 14(2003):1282-1291.
[2] C. Li, H.L Yan, M. Yong et al, Research progress in wireless sensor networks, Computer Research and Development. 42(2005): 163-174.
[3] J.Z Li, J.B Li, S.F Shi, The concept, problem and progress of sensor networks and their data management, Journal of Software. 14(2003): 1717-1727.

[4] R. Sylvia, K. Brad, S. Scott, et al. Data-Centric storage in Sensornets with GHT, a geographic hash table, Mobile Networks and Applications. 8(2003):427-442.

[5] S. Scott, R. Sylvia, K. Brad, et al. Data-Centric storage in sensor nets, ACM SIGCOMM Computer, Communication Review. 33(2003):137-142.

[6] G. Abhishek, G. Jens, C. John, Resilient data-centric storage in wireless ad-hoc sensor networks, Proc. of the 4th Int’l Conf. on Mobile Data Management, (2003):45-62.

[7] Z. Wensheng, C. Guohong, L.P Tom, Data dissemination with ring-based index for wireless sensor networks, IEEE Int’l Conf. on Network Protocols, (2003):305-314.

[8] G. Benjamin, E. Deborah, G. Ramesh, et al. DIFS: a distributed index for features in sensor networks, Proc. Of the 1st IEEE Int’l Workshop on Sensor Network Protocols and Applications Anchorage.(2003):163-173.

[9] L. Xin, J.K Young, G. Ramesh, et al. Multi-Dimensional range queries in sensor networks, Proc. of the 1st Int’l Conf. on Embedded Networked Sensor Systems.(2003):509-517.

[10] R.H Wendi, C. Anantha, B. Hari, Energy-Efficient communication protocol for wireless microsensor networks, Proc. of the 33rd Hawaii Int’l Conf. on System Sciences. (2000):8020-8029.