LETTER

Energy Efficient Skyline Query Processing in Wireless Sensor Networks

Dongook SEONG[†], Junho PARK[†], Myungho YEO^{††}, *Nonmembers*, and Jaesoo YOO^{†a)}, *Member*

SUMMARY In sensor networks, many studies have been proposed to process in-network aggregation efficiently. Unlike general aggregation queries, skyline query processing compares multi-dimensional data for the result. Therefore, it is very difficult to process the skyline queries in sensor networks. It is important to filter unnecessary data for energy-efficient skyline query processing. Existing approaches get rid of unnecessary data transmission by deploying filters to whole sensors. However, network lifetime is reduced due to energy consumption for transmitting filters. In this paper, we propose a lazy filtering-based in-network skyline query processing algorithm to reduce energy consumption by transmitting filters. Our algorithm creates the skyline filter table (SFT) in the data gathering process which sends data from sensor nodes to the base station and filters out unnecessary data transmissions using it. The experimental results show that our algorithm reduces false positive by 53% and improves network lifetime by 44% on average over the existing method.

key words: sensor network, skyline query, filtering, priority

1. Introduction

Recently, wireless sensor networks (WSNs) have found their way into a wide variety of applications and systems with vastly varying requirements and characteristics, including environmental monitoring, smart spaces, medical applications, and precision agriculture [1]–[3]. These networks are mainly used for the systematic gathering of useful information related to the surrounding environment (e.g. temperature, humidity, seismic and acoustic data, etc.). WSNs are large-scale distributed systems that consist of thousands of tiny and low-cost nodes that are limited in terms of energy, computing power, and communication capability.

Presently, diverse techniques have been researched to save the battery power of sensor nodes. Among them, in-network query processing techniques have been suggested to reduce energy consumption during the collection of sensed data. In conventional methods, all collected data are transmitted to the base station for processing queries in the local storage of the base station. However, since they require a large amount of data transmission cost, the lifetime of the network is significantly reduced. But, in-network approaches reduce the amount of data transmission and energy consumption through the aggregation of data in the network during the data routing. In general, data aggregation techniques deal with only one-dimensional attributes. How-

ever, for environmental data collection and processing, there are many cases that the available data can be obtained only when various attribute values are combined. For example, in case of the query such as “Give the information on the areas where the air pollution is severe,” various values such as CO₂, CO, and SO₂ must be considered to determine the air pollution, and all these values must be obtained. This kind of problem can be processed through the skyline query. The skyline query finds a set of interesting objects, satisfying a set of possibly conflicting conditions [4]. In general, in sensor network for environment applications, the queries with various attributes are made frequently. Therefore, an efficient technique for processing such skyline queries is required in the sensor network.

The conventional skyline query processing techniques assume only the situations where all data are located in the same storage space. However, the data are distributed to each sensor in the in-network environment. Therefore, it is difficult to use conventional skyline query processing techniques in the sensor network environment. In this paper, we propose a lazy filtering-based in-network skyline query processing algorithm for the efficient in-network skyline query processing. The proposed algorithm significantly improves network lifetime by filtering out unnecessary data transmission.

The rest of this paper is organized as follows. Section 2 overviews problems of conventional skyline query processing techniques. In Sect. 3, we introduce our skyline query processing technique. Section 4 shows simulated experiments and compares our results with the existing techniques. Finally, we present concluding remarks in Sect. 5.

2. Related Works

In the in-network aggregation processing technique, such as TAG [5], each sensor node sends the aggregated data from the lower layer to the upper layer along a routing path. When processing in-network aggregation queries such as SUM, MAX, and COUNT, the processing can be energy-efficient because only small amount of data are merged in internal nodes without requiring whole data. However, in case of skyline query processing that considers multi-attributes and whole data, it is difficult to derive the correct skyline query result from the partial data set in the in-network aggregation process. It is necessary to collect whole data for processing a skyline query. This scheme incurs huge energy consumption. Therefore, in-network filtering scheme was pro-

Manuscript received March 12, 2010.

[†]The authors are with the Department of Information and Communication Engineering, Chungbuk National University, Cheongju, Korea.

^{††}The author is with Agency for Defense Development, Korea.

a) E-mail: yjs@chungbuk.ac.kr

DOI: 10.1587/transinf.E93.D.2854

Fig. 1 MFTAC.

posed for reducing energy consumption of processing skyline queries. In this in-network filtering scheme, removing unnecessary data is the key.

[7] suggests a filter scheme to remove the unnecessary data from query results. However, because a filter is determined by the parent node and transmitted to its child nodes, it does not affect its sibling nodes and their child nodes. [6] offers a more advanced filtering technique, called MFTAC than [7]. Figure 1 represents an example of MFTAC proposed by [6]. In this approach, first, it determines the Min-Score Filtering Tuple (MFT) to filter out most of the unnecessary data. And then, the MFT is distributed over sensor nodes. Each node filters out unnecessary data that is dominated by the MFT. Cluster heads calculate and distribute the MFT by using the HEED clustering method [8] to reduce the cost for calculating and distributing MFT. However, the MFT should be distributed to all cluster members for skyline query processing. Finally, all sensor nodes are involved in transmitting and receiving the MFT. Also, because the MFT is calculated within the cluster or routing tree that only reflects nearby data, false positives cannot be filtered out effectively. For example, clusters that consist of $C_1 = \{S_2, S_4, S_5\}$, $C_2 = \{S_1, S_3\}$ are sent as a result of the skyline query in Fig. 1. Therefore, C_1 sends $\{S_2\{0.6, 0.9\}, S_4\{0.7, 0.8\}\}$ and C_2 sends $S_1\{0.3, 0.2\}$. However, $S_1\{0.3, 0.2\}$ can be dominated by the result of C_1 . That is, sending $S_1\{0.3, 0.2\}$ was unnecessary. As a result, the MFTAC scheme has false positive and incurs a high cost in distributing filters for the skyline query processing.

In this paper, we propose a skyline query processing technique that is energy-efficient in a distributed environment. We solve the problems of conventional filtering techniques that require extra cost for filtering. Also, we propose a novel technique that improves the performance of filtering.

3. In-Network Skyline Query Processing

In this section, we describe our skyline query processing technique to reduce the filter distribution cost and the unnecessary data transmission. [7] calculates the optimal MFT

Fig. 2 Bottom-up filtering.

at the upper node of the routing tree and transmits the MFT as a filter to child nodes. Each sensor node executes the filtering through the MFT made by the parent node. However, as MFT of [7] reflects only the partial data on the transmission path, the approach by [7] produces many false positive data that reduce the network lifetime. In the scheme, we can easily see that the path for the distribution of the filter is the same as the path for the collection of data. To resolve the problem of false positives, we propose the lazy filtering technique, where the filter is setup based on the present data and is renewed during the data collection. In the following, first, our lazy filtering technique is described and then the problems in the application of lazy filtering method are discussed. Finally, we present the energy-efficient skyline query processing technique to solve these problems.

3.1 Bottom-Up Filtering

To reduce unnecessary data transmissions, our algorithm maintains the skyline filter table (SFT) at each sensor. The SFT preserves the skyline query results using the data at a node and data transmitted from its child nodes. Next, the sensor data transmitted from the child nodes are compared with the SFT to be filtered or to be transmitted to the parent node if they are judged as the result of new skyline. After that, the SFT is renewed. For example, in Fig. 2 each sensor node registers its own present data in the empty SFT. In the case of S_2 , initial entity of SFT is $\{\{0.6, 0.9\}\}$. Next, S_2 receives the data from the child node S_4 . The data of S_4 are not dominated by $\{\{0.6, 0.9\}\}$ in SFT. Therefore, SFT of S_2 is updated to $\{\{0.6, 0.9\}, \{0.7, 0.8\}\}$ and data of S_4 are relayed to S_1 . Next, S_2 receives the data of S_5 , $\{0.3, 0.4\}$. However, it is dominated by the SFT of S_2 , $\{\{0.6, 0.9\}, \{0.7, 0.8\}\}$. Therefore, the data of S_5 are filtered at S_2 . In this case, the filtering efficiency can be greatly changed according to the data transmission order of each sensor. For example, if the data are transmitted to S_1 in order of S_2 , S_4 , and S_3 , the SFT is first renewed to $\{S_2\}$ by S_2 , and S_2 is transmitted to the base station. Then, when S_4 is transmitted, the SFT in S_1 is renewed to $\{S_2, S_4\}$ and S_4 is transmitted to the base station. Finally, when S_3 is transmitted, the SFT in S_1 is renewed to $\{S_3\}$ and S_3 is transmitted to the base station.

Fig. 3 Feature of *Monotone Score*.

But if the data are transmitted to S_1 in order of S_3 , S_2 , and S_4 , the SFT in S_1 is maintained as $\{S_3\}$ and the data of S_2 and S_4 are not transmitted to the base station and therefore excluded. The most ideal case is where the skyline query results in the child nodes are transmitted first, registered in the SFT at the parent nodes, and all of the false positives are excluded. But, it is impossible for the sensor node itself to judge the falsity of the skyline results. Therefore, it is necessary to determine a data collection order.

3.2 Data Gathering Priority

Our proposed technique exploits the *Monotone Score* to raise the filtering effect. In case of the proposed lazy filtering technique, as the transmission order of the data has influence on the efficiency of the filtering, the data collection order is determined by the *Monotone Score*. As shown in Fig. 3, as the value of *Monotone Score* gets smaller, it becomes more probable for the *Monotone Score* to be included in the skyline query result [6]. Also, we applied the *Monotone Score* to the filtering technique. However, as it is applied to the single-value filter of the general type of queries, it is not appropriate to the data filtering of the skyline type of queries.

In our proposed technique, the processing to determine the transmission order of the data is as follows. First, if each attribute of environmental data collected by the sensor is represented as p , the set of j -multiple attributes can be defined as $\text{Attributes} = \{p_1, p_2, p_3, \dots, p_j\}$. All the sensors calculate the *Monotone Score* through the Eq. (1) on the basis of multiple attribute values collected by themselves.

$$\text{MonotoneScore} = \sum_{j=1}^m \frac{1}{p_j} \quad (1)$$

Each sensor calculates the transmission time within the range of one round by using Eq. (2) and transmits the data to the base station when it becomes its own transmission time. In Eq. (2), TransmitTime , $\text{Round}_{\text{current}}$, and $\text{Round}_{\text{term}}$ represent the transmission time of the relevant data, the order of rounds, and the term of one round, respectively. While the MS_{own} represents the *Monotone Score* for its own attributes, the MS_{maximum} stands for the maximum value of the *Monotone Score*.

$$\text{TransmitTime} = ((\text{Round}_{\text{current}} - 1) \times \text{Round}_{\text{term}})$$

ID	Real Data	Monotone Score
S ₁	{0.7, 0.5}	3.43
S ₅	{1.0, 0.3}	4.33
S ₇	{0.3, 1.0}	4.33
S ₄	{0.5, 0.4}	4.5
S ₂	{0.9, 0.2}	6.11
S ₆	{0.2, 0.8}	6.25
S ₃	{0.1, 0.6}	11.66

(a) *Monotone Score* of 2-dimensional attributes

Round Term : 60msec

$$\begin{aligned} S_{5 \text{ T-Time}} &= 120\text{msec} + (60\text{msec} \times S_{5 \text{ OMS}}) / 20 \\ &= 132.99\text{msec} \end{aligned}$$

(b) Transmission time

Fig. 4 Transmission time determination using *Monotone Score*.

$$+ \left(\frac{\text{Round}_{\text{term}} \times MS_{\text{own}}}{MS_{\text{maximum}}} \right) \quad (2)$$

Figure 4(a) shows the collected multiple attributes of SensorsID = $\{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$ and the *Monotone Score*. Here, the transmission of S_5 becomes 132.99 msec through the calculation of {Starting Time of Present Round: 120 msec} + ({Term of one Round: 60 msec} * {Monotone Score of S_5 : 4.33}) / {Maximum Monotone Score: 20} as shown in Fig. 4(b). If the transmission time of each sensor is calculated in this way, it is seen that, the smaller the *Monotone Score* we get, the faster the transmission time is. Through this, the efficiency of the proposed lazy filtering technique is increased.

4. Performance Evaluation

4.1 Environments

To show the superiority of our proposed method, we compare it with the existing skyline query processing method, MFTAC in various environments. The performance evaluation was carried out through the simulation parameters shown in Table 1. We assume that the sensor nodes were uniformly distributed. The model of the energy consumed for message transmission to a sensor node is $\{\text{MessageSize}\} * (\{\text{TransmissionCost}\} + \{\text{AmplificationCost}\} * \{\text{Distance}\}^2)$, where the transmission cost was 50 nJ/b and the amplification cost was set as 100 pJ/b/m². The model of the energy consumed for receiving the message from a sensor node is $\{\text{MessageSize}\} * \{\text{ReceiveCost}\}$, where the receiving cost was set as 50 nJ/b. We used real datasets, which are temperature and humidity data measured in the Washington State [9].

4.2 Experimental Results

Figure 5 and Fig. 6 show the network lifetime and the av-

Table 1 Simulation parameters.

Parameters	Values
Sensor Node Count	120
Sensor Identifier	4 byte
Size of Sensing Data	4 byte
Dimensions of Attribute	2
Initial Sensor Energy	0.1 J

Fig. 5 Network lifetime. **Fig. 6** Average of false positive.

erage false positives of our proposed technique and MFTAC as the number of sensor nodes is changed. MFTAC uses the filter based on the single value. Therefore, it is not appropriate to process the skyline query data with the multi-dimensional and multi-value characteristics. Due to this problem, in MFTAC, the unnecessary data transmissions are not appropriately removed and many false positive data are transmitted to the base station. On the other hand, the proposed technique uses the SFT, which is effective in processing skyline queries as the filter and minimizes the false positives. Furthermore, through the bottom-up filter setting in which the filters are set in the data transmission process, it removes the additional energy consumption necessary for the filter distribution. The filtering performance is also improved by determining the transmission order of filters through *Monotone Score*. As a result, the proposed method reduced the false positives by 53% and increased the total network lifetime by 44% on average.

5. Conclusions

In this paper, we have analyzed the problems of the existing

skyline query processing method and have proposed a new energy efficient method. The existing approach broadcasts MFT values to all sensor nodes in a network for filtering unnecessary data. In this process, sensors consume much energy to transmit filters and the transferred filters have poor filtering performance, which produces many false positives. Our method removes these extra energy consumptions and decreases the number of false positives. Our method does not have a filter broadcasting phase. Instead, each sensor updates its SFT using received data that was transferred by *Monotone Score* and uses multiple MFT values to filter that is suitable for result of skyline type of queries. We evaluated its performance using simulation on 8 ~ 120 node networks. Our results show that the number of false positives collected in the base station was reduced by 53% and the whole network lifetime was prolonged by 44%. Therefore, it is very suitable for conserving energy and extending the lifetime of sensor networks.

Acknowledgements

This work was supported by the Ministry of Education, Science and Technology Grant funded by the Korea Government (The Regional Research Universities Program/Chungbuk BIT Research-Oriented University Consortium) and this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST)(No. 2009-0080279).

References

- [1] R. Szewczyk, A. Mainwaring, J. Polastre, and D. Culler, "An analysis of a large scale habitat monitoring application," Proc. ACM Conf. Embedded Networked Sensor Systems (SenSys' 04), Nov. 2004.
- [2] R. Szewczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, "Habitat monitoring with sensor networks," Commun. ACM, vol.47, no.6, pp.34–40, June 2004.
- [3] A. Sharaf, J. Beaver, A. Labrinidis, and K. Chrysanthos, "Balancing energy efficiency and quality of aggregate data in sensor networks," VLDB, vol.13, no.4, pp.384–403, 2004.
- [4] D. Kossmann, F. Ramsak, and S. Rost, "Shooting stars in the sky: An online algorithm for skyline queries," Proc. International Conference on VLDB, pp.275–286, Aug. 2002.
- [5] S. Madden, M. Franklin, J. Hellerstein, and W. Hong, "TAG: A tiny aggregation service for ad hoc sensor networks," Proc. Usenix Fifth Symp. Operating Systems Design and Implementation (OSDI' 02), pp.131–146, Dec. 2002.
- [6] K. Yoon, J. Choi, Y. Chung, and S. Lee, "In-network processing for skyline queries in sensor networks," IEICE Trans. Commun., vol.E90-B, no.12, pp.3452–3459, Dec. 2007.
- [7] Z. Huang, C.S. Jensen, H. Lu, and B.C. Ooi, "Skyline queries against mobile lightweight devices in MANETs," Proc. International Conference on Data Engineering, p.66–78, 2006.
- [8] O. Younis and S. Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," IEEE Trans. Mobile Computing, vol.3, no.4, pp.366–379, 2004.
- [9] Live from Earth and Mars (LEM) Project, <http://www-k12.atmos.washington.edu/k12/grayskies/>, 2006