International Journal of Innovation Sciences and Research
Vol.4, No, 7, pp.335-339, July- 2015

Available online at
http://www.ijisr.com

RESEARCH ARTICLE

OPTIMIZING TRANSMISSION COST BY USING PTOP-K QUERIES IN WIRELESS SENSOR NETWORK

Dr. K. Madhavi and *Kavitha, S.

Department of Computer Science and Engineering, JNTU College Of Engineering, Anatanpuramu, India

Accepted 15th June, 2015; Published Online 31st July 2015

ABSTRACT

:-Wireless Sensor Networks are widely used in real world applications. Such networks produce uncertain data. Such data can be queried in WSN. However, achieving top-k query processing is a challenging task. This also needs to reduce transmissions. In other words, optimization of data transmissions can reduce the communication cost. This will improve the performance of WSN. Top-k queries are the means of optimizing communications in WSN. In this paper we proposed an algorithm named Energy Efficient algorithm that performs the top-k query processing. We also built a prototype application to demonstrate the proof of concept. We used the bench mark dataset available over Internet pertaining to Adult Daily Life (ADL). This dataset captures the daily life of adults with respect to various events like sleeping, toileting and so on. The empirical results reveal the usefulness of the application.

Key Words: Wireless Sensor Network, Top-k query processing.

INTRODUCTION

Wireless Sensor Networks are widely used in the real world for collecting and using information in both civilian and military applications. The applications are in the fields of healthcare, transportation, commerce, industry, science and military. The data is collected from sensor devices and the quality of sensors improves the quality of data collected from sensors. The precision and other qualities play an important role in sensing. Based on the work done, the sensors can be divided into multiple kinds such as voltic sensors, humidity sensors, and binary sensors and so on. The sensors can also make use of GPS technology so as to sense the data correctly. The sensors in WSN can produce huge amount of uncertain data which can be queried. However, query processing when taken place normally the communication cost is more. This will lead to the performance issues in WSN. To overcome this problem top-k query processing concept is used in this paper. A typical WSN is presented in Figure 1.

As can be seen in Figure 1, it is evident that there are many sensors running in different zones. They need to transmit data to base station. The base station can be subjected to querying. Thus the query with top-k probability is very important for WSN to reduce transition power. In this paper we proposed an algorithm for effective communication in WSN. The algorithm is known as energy efficient algorithm which is responsible to produce top-k results will less cost. Thus the communications in WSN can become optimized for good performance. The remainder of this paper is structured as follows. Section II provides review of relevant literature. Section III presents proposed system. Section IV presents implementation. Section V presents experimental results while section VI concludes the paper.

***Corresponding author: Kavitha, S.**
Department of Computer Science and Engineering, JNTU College Of Engineering, Anatanpuramu, India.

Related works

In this section review of literature is made on top-k query processing in WSN and top – k queries on uncertain data. The researchers in (Silberstein *et al.,* 2006) (Sharaf *et al.,* 2004), (Zeinalipour-Yazti *et al.,* 2005) and (Madden *et al.,* 2002) made extensive study on these areas. Due to the limitations in resources the WSN is vulnerable to various attacks. However, in this paper its effectiveness in communication is considered. When data is more and the transactions involve huge amount of data, the lifetime of the network is lost. That is the reason top-k queries can reduce the amount of data to be transferred. In (Han *et al.,* 2004) and (Sharaf *et al.,* 2004) approximation based data aggregation techniques came into existence. In the same fashion, later on, approximation based top-k queries concept was introduced in (Silberstein *et al.,* 2006). A model driven approach for the same is presented in (Xu *et al.,* 2006) while further study is found in (Wang *et al.,* 2008) and (Wu *et al.,* 2007).

Many researches were made recently. For instance in (Ye *et al.,* 2010), (Li *et al.,* 2009), (Cormode *et al.,* 2009), (Lian and Chen, 2008), (Soliman *et al.,* 2007) and (Jin *et al.,* 2008) there was research on top-k query processing. The query processing was made in WSN especially for top-k query processing. As this kind of query processing bestows advantages besides increasing the life of network, it is very useful. In this paper a new algorithm is proposed towards optimizing transmissions in WSN.

Proposed solution for Top-K Query Processing

As the WSN is widely used in many real world applications, it became imperative that the communication cost of the network is optimized. Top-k query processing can reduce the communication cost.

In this paper we proposed an algorithm named "Energy Efficient Algorithm" for top-k query processing. The algorithm is as follows.

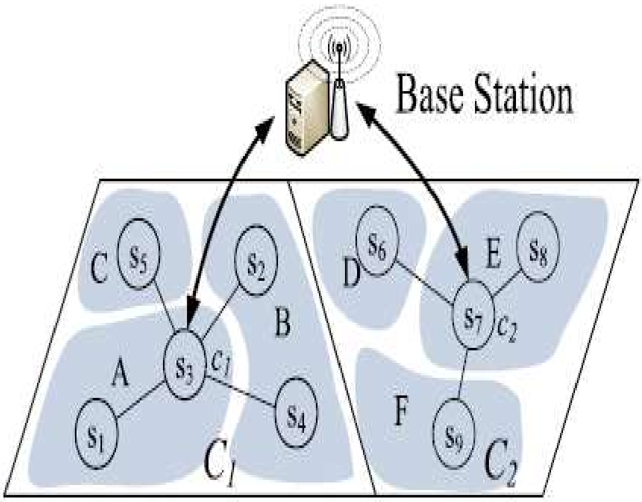


Figure 1. Wireless sensor network with multiple zones

Proposed Algorithm

Algorithm : Energy Efficient Algorithm
Purpose :To process top – k queries on dataset generated by sensors
Inputs : dataset, top-k query
Outputs : Top – K results
STEP 1: PRE-PROCESSING
Load given dataset into a reader object
Write the dataset content to a relational table
STEP 2: IDENTIFY PROBABLE TOP K COLUMN
For Each Column in Dataset
IF column is top k probable
Choose column as Candidate for Top K Processing
END IF
END
Initialize F to hold unique field values
For all values in chosen Column
Add unique value to ***F***
END
STEP 3: : PROCESSING TOP –K QUERIES
Initialize ***R*** for holding top K Results
Populate F on UI
Take Top K Input from End User
Extract all rows from dataset that satisfy user selection
Compute ranking for rows
Sort rows in ascending order by rank
Populate top K rows into ***R***
Return ***R***

Listing 1. Proposed algorithm for top – k search

As can be seen in Listing 1, it is evident that the algorithm has multiple steps that can be used to complete the processing of Top-k queries. In step 1, pre-processing takes place. Step 2 focuses on identifying probable top-k columns. Step 3 is meant for actually processing top-k queries can be shown in below Figure 2. The dataset details are as given below.

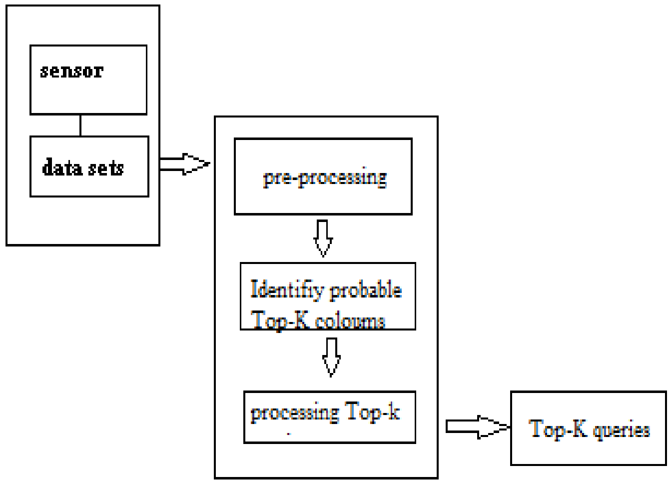


Figure 2. Top-K Query Processing

Data Set Used for Experiments

Activities of Daily Living (ADLs) Recognition Using Binary Sensors are captured to form this dataset. This dataset comprises information regarding the ADLs performed by two users on a daily basis in their own homes. This dataset is composed by two instances of data, each one corresponding to a different user and summing up to 35 days of fully labeled data. Each instance of the dataset is described by three text files, namely: description, sensors events (features), activities of the daily living (labels). Sensor events were recorded using a wireless sensor network and data were labeled manually.

Start time	**End time**	**Activity-**
-------------------	--------------------	
2011-11-28 02:27:59	2011-11-28 10:18:11	Sleeping
2011-11-28 10:21:24	2011-11-28 10:23:36	Toileting
2011-11-28 10:25:44	2011-11-28 10:33:00	Showering
2011-11-28 10:34:23	2011-11-28 10:43:00	Breakfast
2011-11-28 10:49:48	2011-11-28 10:51:13	Grooming
2011-11-28 10:51:41	2011-11-28 13:05:07	Spare_Time/TV
2011-11-28 13:06:04	2011-11-28 13:06:31	Toileting
2011-11-28 13:09:31	2011-11-28 13:29:09	Leaving
2011-11-28 13:38:40	2011-11-28 14:21:40	Spare_Time/TV
2011-11-28 14:22:38	2011-11-28 14:27:07	Toileting
2011-11-28 14:27:11	2011-11-28 15:04:00	Lunch
2011-11-28 15:04:59	2011-11-28 15:06:29	Grooming
2011-11-28 15:07:01	2011-11-28 20:20:00	Spare_Time/TV
2011-11-28 20:20:55	2011-11-28 20:20:59	Snack
2011-11-28 20:21:15	2011-11-29 02:06:00	Spare_Time/TV

2011-11-29 02:16:00	2011-11-29 11:31:00	Sleeping
2011-11-29 11:31:55	2011-11-29 11:36:55	Toileting
2011-11-29 11:37:38	2011-11-29 11:48:54	Grooming
2011-11-29 11:49:57	2011-11-29 11:51:13	Showering
2011-11-29 12:08:28	2011-11-29 12:18:00	Breakfast
2011-11-29 12:19:01	2011-11-29 12:22:00	Grooming
2011-11-29 12:22:38	2011-11-29 12:24:59	Spare_Time/TV
2011-11-29 13:25:29	2011-11-29 13:25:32	Snack
2011-11-29 13:25:38	2011-11-29 15:12:26	Spare_Time/TV
2011-11-29 15:13:28	2011-11-29 15:13:57	Toileting
2011-11-29 15:14:33	2011-11-29 15:45:54	Lunch
2011-11-29 15:49:51	2011-11-29 15:50:54	Grooming
2011-11-29 15:52:04	2011-11-29 16:17:58	Spare_Time/TV
2011-11-29 16:18:00	2011-11-29 16:31:27	Toileting
2011-11-29 16:34:17	2011-11-29 17:08:07	Spare_Time/TV

Table 1. Shows a portion of ADL dataset

Implementation

We built a prototype application in order to demonstrate the proof of concept. The proposed algorithm is applied on real time dataset which is captured on Adult Daily Life. The dataset has data pertaining to various activities or events of adults and their starting and ending times. The aim of the paper is to implement a solution for optimizing the top – k queries in wireless sensor networks. The methodology used to implement is described here. First of all sensor dataset is obtained on Adult Daily Life, and then the proposed algorithm is applied on the dataset using a web based application. The algorithm is able to produce top-k results on making queries based on selected activities. As can be seen in Figure 3, the energy efficient algorithm is able produce events and their corresponding frequencies as an intermediary step. Afterwards, the algorithm produced final top-k results based on selected event. As can be seen in Figure 4, it is evident that the top-k results are produced for the event breakfast. This way for every event top-k results can be generated by the algorithm.

Experimental Results

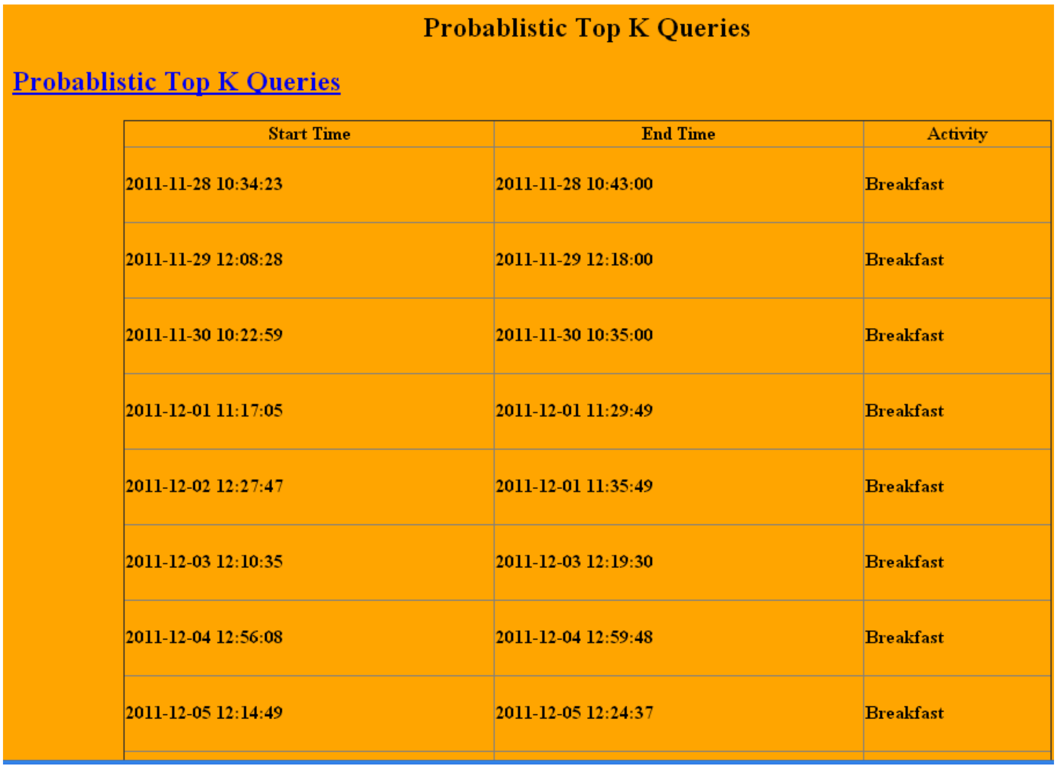
Experiments are made with the prototype application which facilitates the user to view top-k results. For each activity the algorithm can produce frequency of the events and produce the top-k results. As can be seen in Figure 5, it is evident that the events are presented in the graph that reflects the event dynamics of the two families of people studied by deploying two sets of sensors. The sensor produced data is analyzed and mined so as to produce the event wise count of individuals.

Senget Activity	Count
Breakfast	56
Grooming	204
Leaving	56
Lunch	36
Showering	56
Sleeping	52
Snack	44
Spare_Time/TV	308
Toileting	176

Graph

butedProcessing/index.jsp/DatasetDisplay.jsp?id=Breakfast

Figure 3. Activities of adults and the frequency of events



Probablistic Top K Queries

Probablistic Top K Queries

Start Time	End Time	Activity
2011-11-28 10:34:23	2011-11-28 10:43:00	Breakfast
2011-11-29 12:08:28	2011-11-29 12:18:00	Breakfast
2011-11-30 10:22:59	2011-11-30 10:35:00	Breakfast
2011-12-01 11:17:05	2011-12-01 11:29:49	Breakfast
2011-12-02 12:27:47	2011-12-01 11:35:49	Breakfast
2011-12-03 12:10:35	2011-12-03 12:19:30	Breakfast
2011-12-04 12:56:08	2011-12-04 12:59:48	Breakfast
2011-12-05 12:14:49	2011-12-05 12:24:37	Breakfast

Figure 4. Shows top-k results for the event "Breakfast"

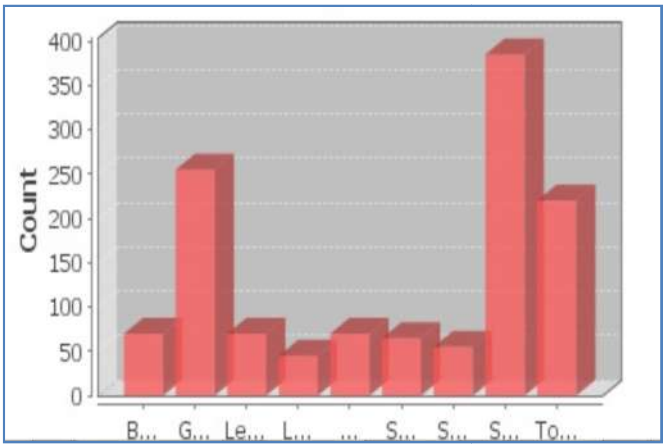


Figure 5. Shows Adult Daily Life events

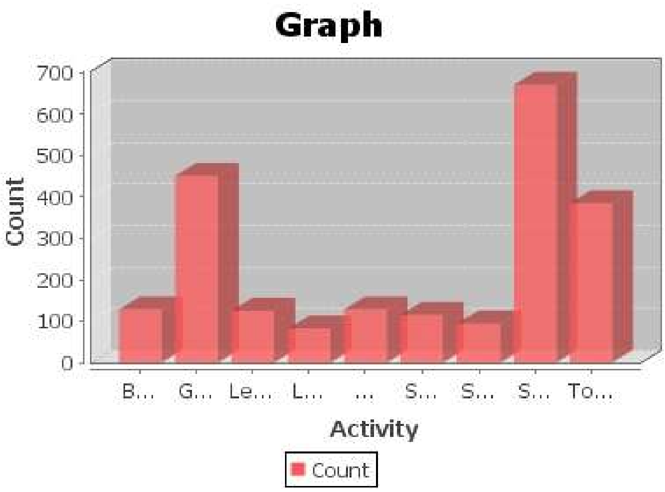


Figure 6. Shows Adult Daily Life events (Experiment 2)

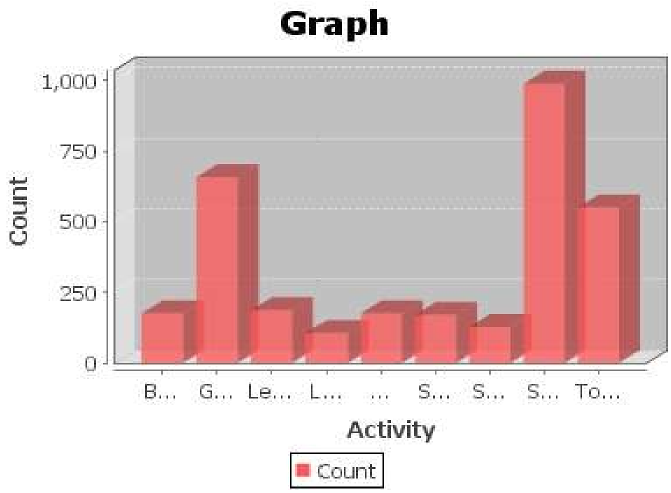


Figure 7. Shows Adult Daily Life events (Experiment 3)

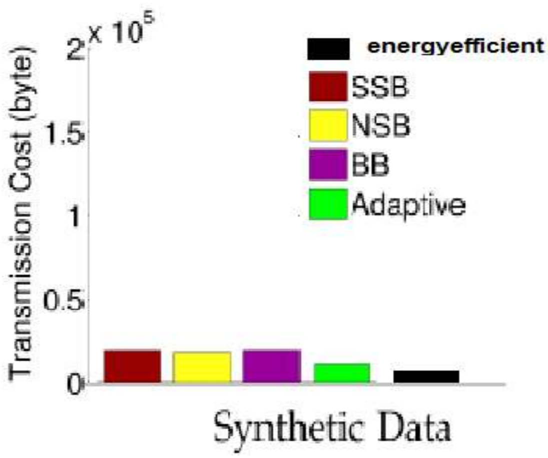


Figure 7. Cost Optimization using synthetic data

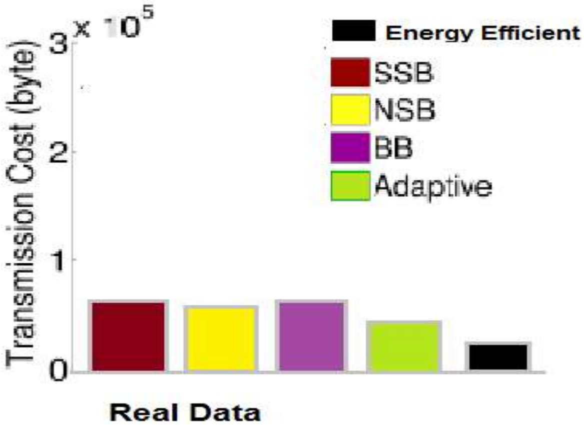


Figure 8. Cost Optimization using real data

As can be seen in Figure 6, it is evident that the events are presented in the graph that reflects the event dynamics of the two families of people studied by deploying two sets of sensors. The sensor produced data is analyzed and mined so as to produce the event wise count of individuals. The results are of experiment 2. As can be seen in Figure 7, it is evident that the events are presented in the graph that reflects the event dynamics of the two families of people studied by deploying two sets of sensors. The sensor produced data is analyzed and mined so as to produce the event wise count of individuals. The results are of experiment 3.

Performance Evaluation

The previous experiments on adaptive algorithm are conducted on a setting that exhibits dynamic changes with certain temporal locality. Since the algorithm dynamically adapts to the changes by switching to appropriate methods, it provides an additional saving over the other algorithms. When compare to existing method the proposed method the cost optimization is reduced. First validate the effectiveness of proposed methods in reducing the transmission cost using synthetic data and real data .the cost optimization clearly shown in below graphs.

Conclusion

In this paper we studied the problem top-k queries in WSN. The nodes in WSN can produce huge amount of uncertain data. Querying such data in a cost-effective fashion is the need of the hour. Therefore our study focused on processing top-k queries in WSN. Towards this end we proposed an algorithm for achieving top-k results. The proposed algorithm also incurs less cost for producing top-k results. We built a prototype application for demonstrating proof of concept. The application provides user-friendly interface and the users can have queries on the data. The ADL dataset is used for experiments. The dataset has details of adults' daily life which reflects the priorities or the time spending patterns of adults. The experimental results reveal that the system is useful and can be used in real world applications.

REFERENCES

Cormode, G. Li, F. and Yi, K. 2009. "Semantics of Ranking Queries for Probabilistic Data and Expected Ranks," Proc. IEEE Int'l Conf. Data Eng. (ICDE '09).

Han, Q., Mehrotra, S. and Venkatasubramanian, N. 2004. "Energy Efficient Data Collection in Distributed Sensor Environments," Proc. 24th *Int'l Conf. Distributed Computing Systems (ICDCS '04),* pp. 590-597.

Jin, C., Yi, K., Chen, L., Yu, J.X. and Lin, X. 2008. "Sliding-Window Top-k Queries on Uncertain Streams," Proc. Int'l Conf. Very Large Data Bases (VLDB '08).

Li, F., Yi, K. and Jestes, J. 2009. "Ranking Distributed Probabilistic Data,"Proc. 35th SIGMOD *Int'l Conf. Management of Data (SIGMOD '09).*

Lian, X. and Chen, L. 2008. "Probabilistic Ranked Queries in Uncertain Databases," Proc. 11th *Int'l Conf. Extending Database Technology (EDBT '08),* pp. 511-522.

Madden, S. Franklin, M.J. Hellerstein, J. and Hong, W. 2002. "TAG: A Tiny AGgregation Service for Ad-Hoc Sensor Networks," Proc. Fifth Symp. Operating Systems Design and Implementation (OSDI '02).

Sharaf, A., Beaver, J., Labrinidis, A. and Chrysanthis, K. 2004. "Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks*," Int'l J. Very Large Data Bases,* vol. 13, no. 4, pp. 384-403,

Silberstein, A.S., Braynard, R., Ellis, C., Munagala, K., and Yang, J. 2006. "A Sampling-Based Approach to Optimizing Top-k Queries in Sensor Networks," Proc. 22nd *Int'l Conf. Data Eng.,* (ICDE '06), p. 68,

Soliman, M.A., Ilyas, I.F. and Chang, K.C. 2007. "Top-k Query Processing in Uncertain Databases," *Proc. Int'l Conf. Data Eng. (ICDE'07).*

Wang, D., Xu, J., Liu, J. and Wang, F. 2008. "Mobile Filtering for Error Bounded Data Collection in Sensor Networks," Proc. 28th *Int'l Conf. Distributed Computing Systems (ICDCS '08),* pp. 530-537.

Wu, M., Xu, J., Tang, X. and Lee, W.C. 2007. "Top-k Monitoring in Wireless Sensor Networks," IEEE Trans. Knowledge and Data Eng.,vol. 19, no. 7, pp. 962-976, July.

Xu, Y., Lee, W.C., Xu, J. and Mitchell, G. 2006. "Processing Window Queries in Wireless Sensor Networks," Proc. *IEEE 22nd Int'l Conf. Data Eng.,* (ICDE '06).

Ye, M., Liu, X., Lee, W.C. and Lee, D.L. 2010. "Probabilistic Top-k Query Processing in Distributed Sensor Networks," Proc. *IEEE Int'l Conf. Data Eng., (ICDE '10).*

Zeinalipour-Yazti, D., Vagena, Z., Gunopulos, D., Kalogeraki, V., Tsotras, V., Vlachos, M., Koudas, N. and Srivastava, D. 2005. "The Threshold Join Algorithm for Top-k Queries in Distributed Sensor Networks," Proc. Second *Int'l Workshop Data Management for Sensor Networks (DMSN '05)*, pp. 61-66.

****