International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 9 Issue 3 –SEPTEMBER 2014.

Disseminated Dispensation of Queries in WSN

A.Ramanjinamma[*1], and K.Kalyani[#2]

[*]Student (M>Tech), Dept of CSE, Ananthalakshmi Institute of Technology & Sciences,
Affiliated to JNTUA University, ALITS- Anantapur

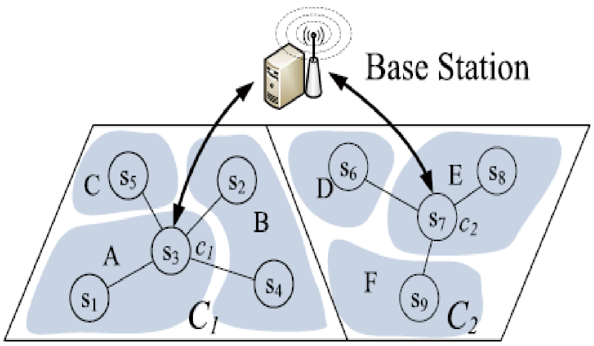
[#] Asst Professor, Dept of CSE, Ananthalakshmi Institute of Technology & Sciences,
Affiliated to JNTUA University, ALITS- Anantapur

***Abstract*— Now-a-days Wireless sensor networks (WSN) is a network that is widely used in many applications. Wireless sensor networks (WSN) is used to extract the information and collect the data from the real world or physical world. This technology is being in many fields like military, science, commerce, industry etc. As known, wireless sensor networks (WSN) has sensing quality, this quality produses the trusted results with tolerance and accuracy to external noise and tolerance. In the present system 2 various techniques are used namely, necessary set and sufficient set. But these two techniques cannot overcome that disadvantage of communication cost or transmission cost. So here we propose a new system. In the proposed system, we develop some algorithms namely SSB, NSB, BB. These algorithms are used with boundary rounds of communication for intercluster query processing. These three algorithms changes dynamically based on data distribution on the network and then minimize the transmission cost. The sufficient set of data is with two-tier and tree structure topologies, in wireless sensor networks (WSN). Here, the experimental results show that the proposed algorithm reduces the transmission of data significantly. Hence based on various conditions we present near optimal solutions. These solutions are with high performance.**

KEYWORDS: Wireless sensor networks, probabilistic databases, distributed data management, Top-k queries.

I. Introduction

Wireless sensor networks (WSN) are rapidly increasing the ways to collect the data and use that as information from real world or physical world. This new technology has resulted in rapid impacts on a wide range of applications in various fields. Some of the applications of wireless sensor networks (WSN) are military, industry, science, commerce, health care, transportation etc. but however the quality of individual sensors changes gradually in terms o their accuracy, tolerance to hardware/software noise, sensing precision etc. Many previous schemes show that distribution of noise varies widely in various photovoltaic sensors, precision and accuracy of readings usually changes significantly in humidity type sensors, and the errors in GPS devices can be up to several meters. Sensor readers are uncertain, inherently. To provide a facility of managing uncertain data, research on probabilistic databases has gathered maximum attentions in past few years. Most of the recent proposals on probabilistic data modelling introduced to associate a confidence with a data record/ data tuple to capture the uncertainty of data. Hence we can carry a possible world semantic. Here we make use of some environmental monitoring applications that are relared to wireless sensor networks (WSN). This is done to introduce some probabilistic databases. Consider a wireless sensor networks (WSN), which contains large number of sensor nodes. These nodes are employed in a geographical region. From these distributed sensor nodes we are going to collect feature readings, which are related to speed of wind gust or moisture levels.



A wireless sensor network

Due to sensing environmental and imprecision interferences, the sensor readings are usually noisy. Thus, multiple sensors are employed at certain zones in order to increase monitoring quality, gradually. In this network, sensor nodes are grouped together as clusters. For performing localized data processing, one sensor is selected as the cluster head from group of sensors. By using statistic methods, a cluster head may produce a set of data tuples for each zone within its region or monitored region. Here, we assume that each tuple is comprised of tuple id, region zone, a derived attribute value that is possible, along with a confidence that serves as a measurement of data uncertainty. Thus, the data tuples coordinates to the same zone collectively and represents the probabilistic distribution of derived possible values for the required zone. Since the existence of all possible values in these tuples is exclusive to each other, they naturally form a tuple called logical tuple. The logical tuple is also called as x-tuple. On uncertain data the Top-k queries have received increasing interests recently. When given a scoring function and a data set, a top-k query finds the maximum ranked k tuples at over all possible worlds. The answers to this posted query may vary in different possible worlds, making the problem challenging and interesting. However, various syntaxes of top-k queries can be employed in various applications.

II. RELATED WORK

A cluster head may generate a set of data tuples for each regional zone within its monitored region, while using static methods. We propose three algorithms, which are used to minimize the energy and communication overhead in data distribution in the network, responding to changing. These algorithms are sufficient set-based (SSB) algorithm. The advantages of NSB and BB take advantages of the skewed necessary sets and necessary boundaries among clusters that are local. To obtain their global boundaries, respectively, these are very effective for inter-cluster pruning. The cost of transmission increases for all algorithms because the number of tuples are increased needed for query processing.

Fig.1. Top-k process architecture

2.1 History of Top-k query processing

In different applications, various semantics of top-k queries can be deployed in different applications. In recent trends, there are several top-k query syntaxes and semantics and solutions that are proposed. This includes U-Topk and UkRank, PT-Topk in, PK-Topk in, expected rank and so on. A common way to process probabilistic top-k queries is to sort all tuples first based on the attribute that is scoring, and then process tuples in the sorted order, until some termination condition is me, to compute the final answer set. For the above mentioned queries the answer sets typically consist of highly ranked tuples in the sorted list. This is done because tuples organized lowly usually do not have the wanted ranks and confidence to be included in the answer sets. When these works produce insightful results, they are conducted on a centralized database rather than in a distributed setting as we target on in this paper. Only recently, a study on processing a probabilistic query with expected rank semantic in distributed systems has appeared.

III. INTRACLUSTER DATA PRUNING

In a cluster-based wireless sensor network, the cluster heads are reason for generating uncertain data tuples from the collected raw sensor readings within their clusters. To solve a query, it's natural for the cluster heads to prune redundant and uncertain data tuples before delivery to the home station in order to reduce communication cost and energy cost. The key issue here is how to derive a compact set of tuples that are essential for the home station to answer the probabilistic top-k queries. This is a very challenging issue for the following reasons:

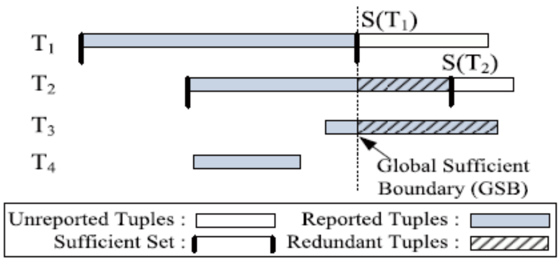
1) The interplay of probability and ranking based on the semantic of probabilistic top-k queries

2) The lack of global knowledge to determine the probability and ranking of candidate tuples locally at cluster heads.

Here, we propose the notion of sufficient set(SS) and necessary set(NS), and describe how to identify them from local data sets at cluster heads. Next, we use the PT-Topk query as a test case to derive sufficient set and necessary set and show that the top-k probability of a tuple t obtained locally is an upper bound of its true top-k probability. Thus, data tuples excluded from the sufficient sets and necessary sets in local clusters will never appear in the final answer set.

3.1 Definition of Sufficient sets and Necessary Sets:

It would be advantageous if cluster heads are used to find the minimum sets which belong to their local tuples which are sufficient for the base station, which is used to answer a given query. Ideally, sufficient set is a subset of the data set that is local. Data is excluded from the sufficient set, no matter which clusters they reside, will never be included in the final answer set nor involved in the computation of the final answer set. Here, we define the sufficient set more formally as follows:



.

Fig.2 Data distribution over the wireless sensor network

Sufficient Set. Given an uncertain data set Ti in cluster Ci, if there exists a tuple tsb €Ti (called sufficient boundary) such that the tuples ranked lower than tsb are useless for query processing at the base station, then the sufficient set of Ti, denoted as S(Ti), is a subset of Ti as specified below:

$$S(T_i) = \{t | t =_f t_{sb} \text{ort} \prec_f t_{sb}\},$$

where f is a given scoring function for ranking.

Necessary Set. Given a local data set Ti in cluster Ci, assume that Ai is the set of locally known candidate tuples for the final answer and tnb (called necessary boundary) is the lowest ranked tuple in Ai. The necessary set of Ti, denoted as N(Ti), is

$$N(T_i) = \{t | t \in T_i, t \prec_f t_{nb}\}.$$

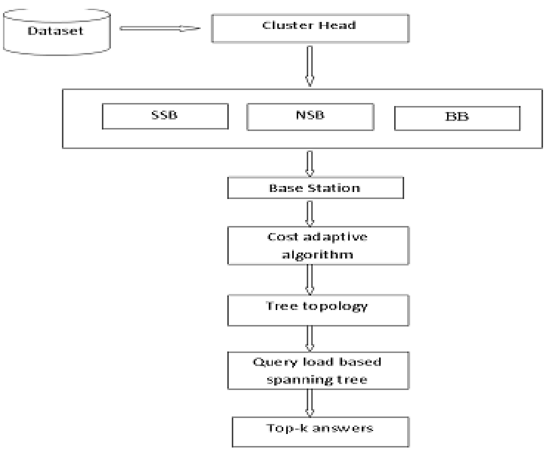


Fig3: Data Set for System Design

IV. INTERCLUSTER QUERY PROCESSING

Here, using the notion of sufficient sets and necessary sets as a basis, we propose 3 distributed algorithms for processing probabilistic top-k queries in wireless sensor networks, they are:

1) Sufficient Set-based method
2) Necessary Set-based method
3) Boundary-based method.

4.1 Sufficient Set-Based Algorithm

Algorithm 1.SSB_ALGORITHM

1:/* Cluster Head c_i Side */
2:Compute the sufficient boundary $SB(T_i)$ of (T_i)
3: if $SB(T_i)$ exists then
4: $S(T_i) \leftarrow \{ t | t \leq SB(T_i) \wedge t \varepsilon T_i$
5: $(T_i)^1 \leftarrow S(T_i)$
6:else
7: $(T_i)^1 \leftarrow T_i$
8:end if;
9:Deliver $(T_i)^1$ to the base station
1:/* Base Station side */
2: Collect $(T_i)^1$ from all c_i ($1 \leq i \leq M$)
3:$T' \leftarrow \cup_{1 \leq i \leq M} T_i'$
4:Execute centralized algorithm over T'

4.2 Necessary Set-Based Algorithm

Algorithm 2. NSB_ALGORITHM

1: /* Cluster Head c_i Side */
2: Compute the necessary boundary $NB(T_i)$ of T_i
3: $N(T_i) \leftarrow \{t | t \preceq_f NB(T_i) \wedge t \in T_i\}$
4: Deliver $N(T_i)$ to the base station
5: **if** Receive GB from the base station **than**
6: $N'(T_i) \leftarrow \{t | t \preceq_f GB \wedge t \in [T_i - N(T_i)]\}$
7: Send $N'(T_i)$ to the base station
8: **end if**

4.3 Boundary-Based Algorithm

Algorithm 3. BB_ALGORITHM

1: /* Cluster Head c_i Side */
2: Compute $NB(T_i)$ and $SB(T_i)$ of T_i
3: Send $NB(T_i)$ and $SB(T_i)$ to the base station
4: Receive GB from the base station
5: $T_i' \leftarrow \{t | (t \preceq_f GB) \wedge (t \in T_i)\}$
6: Deliver T_i' to the base station

1: /* Base Station Side */
2: Collect $NB(T_i)$ and $SB(T_i)$ from all c_i $(1 \leq i \leq M)$
3: Let $SB_{highest}$, NB_{lowest} be the highest ranked $SB(T_i)$ and the lowest ranked $NB(T_i)$ respectively, where $(1 \leq i \leq M)$
4: **if** $SB_{highest} \prec_f NB_{lowest}$ **than**
5: GB $\leftarrow SB_{highest}$
6: **else**
7: GB $\leftarrow NB_{lowest}$
8: **end if**
9: Broadcast GB to each cluster head
10: Collect T_i' from all c_i $(1 \leq i \leq M)$.
11: $T' \leftarrow \cup_{1 \leq i \leq M} T_i'$
12: Execute centralized algorithm over T'

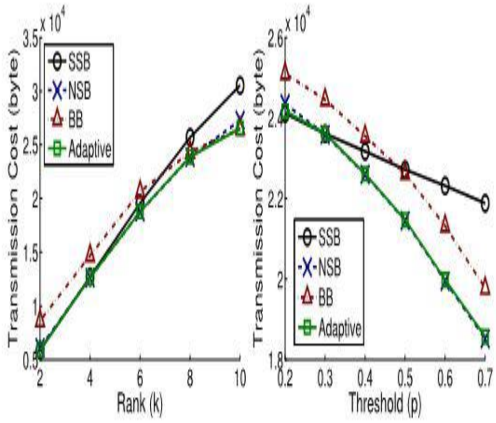
V. EVALUATION

In this section, we first conduct a simulation-based performance evaluation on the distributed algorithms for processing PTtop k queries in two-tier hierarchical cluster-based wireless sensor monitoring system. As discussed, limited energy budget is a critical issue for wireless sensor network and radio transmission is the most dominate source of energy consumption.Thus, we measure the total amount of data transmission as the performance metrics. Notice that, response time is another important metrics to evaluate query processing algorithms in wireless sensor networks. All of those three algorithms, i.e., SSB, NSB, and BB, perform at most two rounds of message exchange, thus clearly outperform an iterative approach (developed based on the processing strategy in [14]), which usually needs hundreds of iterations. Note that, there is not much difference among SSB, NSB, and BB in terms of query response time, thus we focus on the data transmission cost in the evaluation. Finally, we also conduct experiments to evaluate algorithms, SSB-T, NSB-T, and NSB-T-Opt under the tree-structured network topology. A value generated based on the node is location in order to inject spatial locality and is a random variable ranging from 0 to 1. When ¼ 1, the data distribution is highly spatial-correlated. When? ¼ 0, the data distribution is totally random. The real sensor traces are from the Intel Lab Data [35]. We treat the readings for each sensor node in the traces as for an area of interest by assigning readings to sensor nodes in an area.

Parameters	Default	Settings
k: # of top ranked tuples	6	$2 \sim 10$
p: probability threshold	0.5	$0.2 \sim 0.7$
η: data skewness factor	0.5	$0 \sim 1$
λ: mean size of X-tuple	5	$1 \sim 10$
Δ: adaptive window size	15	$1 \sim 50$
M: # of clusters	36	$4 \sim 100$
I: # of interest zones/x-tuples	500	
S_d: data tuple size (bytes)	32	$8 \sim 128$
S_q: query message size (bytes)	8	
S_b: boundary message size (bytes)	8	

Parameter Settings

We first validate the effectiveness of our proposed methods in reducing the transmission cost against two baseline approaches, including 1) a naive approach, which simply transmits the entire data set to the base station for query processing 2) an iterative approach devised based on the



(a) Varied k (b) Varied p

VI. CONCLUSION

Here, we propose the notion of sufficient set (SS) and necessary set (NS) for efficient network pruning in distributed Data that is uncertain in probabilistic top-k query processing. In the same way, we systematically derive sufficient boundaries and necessary boundaries. Here, we propose a set of algorithms, namely SSB, NSB, and BB algorithms, for in-network processing of PT-Top k queries. Additionally, we derive a cost model on communication cost of the three proposed algorithms and propose a cost-based adaptive algorithm that adapts to the application dynamics. But our work in this paper is based mainly under the setting of hierarchical network that is two-tier, the concepts of sufficient set and necessary set are universal and can be easily elaborated to a network with tree topology. The performance evaluation shows our ideas and shows that the proposed algorithms significantly reduce data transmissions. While focusing on this the developed concepts can be applied to develop algorithms to support other probabilistic top-k queries.

REFERENCES

[1]V. Bychkovskiy, S. Megerian, D. Estrin, and M. Potkonjak, “ACollaborative Approach to in-Place Sensor Calibration,” Proc.Second Int’l Conf. Information Processing in Sensor Networks (IPSN),pp. 301-316, 2003.

[2] E. Elnahrawy and B. Nath, “Poster Abstract: Online Data Cleaning in Wireless Sensor Networks,” Proc. First Int’l Conf. Embedded Networked Sensor Systems (SenSys ’03), pp. 294-295, 2003.

[3] A. Deshpande, C. Guestrin, S.R. Madden, J.M. Hellerstein, and W. Hong, “Model-Driven Data Acquisition in Sensor Networks,” Proc. 13th Int’l Conf. Very Large Data Bases (VLDB ’04), pp. 588-599, 2004.

[4] R. Cheng, D.V. Kalashnikov, and S. Prabhakar, “Evaluating Probabilistic Queries over Imprecise Data,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD ’03), pp. 551-562, 2003.

[5] S. Abiteboul, P. Kanellakis, and G. Grahne, "On the Representation and Querying of Sets of Possible Worlds," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '87), pp. 34- 48, 1987.
[6] R. Cheng, Y. Xia, S. Prabhakar, R. Shah, and J.S. Vitter, "Efficient Indexing Methods for Probabilistic Threshold Queries over Uncertain Data," Proc. 30th Int'l Conf. Very Large Data Bases (VLDB'04), pp. 876-887, 2004.