
1. N.V.R.D. Swathi 2. K. Naga Bhargavi
1, 2 Dept., of CSE, Sri Sai Aditya Institute of Science & Technology, Surampalem, EG.dist, AP, India.

Abstract: A wireless sensor Network consists of spatially dispersed self-directed sensors to monitor the environment conditions. This work supports in-network top-k query process over uncertain data in a distributed wireless sensor network. We develop the notion of necessary Set for efficient in-network pruning of uncertain data in a distributed setting. Based on the necessary sets, we propose, necessary set-based, algorithm for in-network processing of PT-Topk queries in a two-tier hierarchical sensor network. For providing privacy proposing locked Q protocol that prevents attackers from gaining information from both sensor collected data and sink issued queries.

Index terms: Top-k Query Process, SecureQ protocol, NSB

1. Introduction

Wireless Sensor Network usually collects the information from the physical environment. The information that is collected from those physical environments is of uncertain data and there is a presence of noise. Sometimes, many sensor nodes are deployed in an environment to avoid the data uncertainty for sensing precision. This is used in various applications such as military, commerce and healthcare etc. In this networks, data uncertainty and energy consumption is main issue when we consider in the sensor networks. Recent research on probabilistic data has received renewed attentions and they are measured by confidence values that are associated with it. It is measured by using fixing the threshold limit for removing uncertainty. Then after the process in sensor nodes the information is delivered to the base station, sometimes it takes many rounds of communication to complete the process. So, the energy consumption will be high as it takes many rounds of communication. They have a number of accessed tuple and materialized search states. So, it will take high memory to process each state. The ranking across the major models of uncertain data, such as attribute and tuple-level uncertainty are done. The imprecision in the data often lead to a large number of answers of low quality.

2. Related Work

An extensive number of research works in this area has appeared in the literature. Due to the limited energy budget available at sensor nodes, the primary issue is how to develop energy-efficient techniques to reduce communication and energy costs in the networks. TAG is one of the first studies in this area. By exploring the semantics of aggregate operators (e.g., sum, avg, and top-k), in-network processing approach is adopted to suppress redundant data transmissions in wireless sensor networks. Approximate-based data aggregation techniques have also been proposed. The idea is to tradeoff some data quality for
improved energy efficiency. Silberstein et al. develop a sampling-based approach to evaluate approximate top-k queries in wireless sensor networks. Probabilistic ranked queries based on uncertainty at the attribute level are studied. A unique study that ranks tuple by their probabilities satisfying the query. Finally, uncertain top-k query is studied under the setting of streaming databases where a compact data set is exploited to support efficient slide window top-k queries.

For our first idea, we consider the semantics of the query and the properties/attributes of the sensor nodes when configuring the sensor network. We have observed that the length of the messages sent by a node when processing Group-By queries depends on the number of groups existing in the routing sub tree rooted at that node. This observation leads to the principle that reducing the number of groups considered at a sensor node (performing in-network aggregation) will reduce the size of the transmitted data and hence incur less energy cost for transmitting them. Based on this principle, our first contribution is a group-aware network configuration method that “clusters” along the same path sensor nodes that belong to the same group.

For our second idea, we exploit temporal correlation in streams of sensor readings to suppress insignificant readings which can be tolerated. Further, suppressing such readings potentially allow nodes that do not have to transmit data to switch into doze or sleep mode, powering down their antennas. Doze mode offers the maximum possible saving in energy. Towards this, our second contribution is a framework, called TiNA (short for temporal coherency-aware in-Network Aggregation). TiNA works on top of existing in-network aggregation schemes such as TAG and Cougar and aims to balance the reduction in energy with the loss of QoD by adhering to user-specified QoD requirements. TiNA reduces energy consumption by using temporal coherency tolerances in conjunction with in-network aggregation to save energy at each sensor node, while maintaining the specified quality of data. These tolerances are based on user preferences or can be dictated by the network in cases where the network cannot support the current tolerance level.

In this paper we focus on optimizing top-k queries in sensor networks. A top-k query returns the k nodes with the highest sensor readings in the network. Top-k queries are both useful in practice and theoretically interesting. They serve as a case study to illustrate the challenges and difficulties in reasoning with models of joint distributions, and demonstrate the advantages of our proposed optimization framework.

Consider the simple example of ornithologists studying various bird species in a forest who do not know where birds are most likely to be found at any given time. They place bird feeders in various locations in the forest, each of which includes a sensor that detects weight changes to count the number of times birds land at the feeder. Before entering the forest, the ornithologists want to know the best locations to find birds. They run a top-k query over the network to determine the feeders that have attracted the most birds recently and therefore are likely to be good points of observation. Unpopular feeders attracting few birds should be omitted from the query result. A top-k query meets exactly this goal. Of course, in practice, the basic top-k query may need to be more tailored to this setting, e.g., the researchers might want to group nearby feeders into clusters for purposes of
observation, and obtain the top k clusters ordered by average bird count. Nevertheless, the basic form of the query remains top-k.

**Literature Survey**

In wireless sensor networks are revolutionizing the ways to collect and use information from the physical world. This new technology has resulted in significant impacts on a wide array of applications in various fields, including military, science, industry, commerce, transportation and health-care. However, the quality of sensors varies significantly in terms of their sensing precision, accuracy, tolerance to hardware/external noise and so on. To facilitate management of uncertain data, researches on probabilistic databases have received renewed attentions in the past few years. Most of the recent works on probabilistic data modeling propose to associate a confidence with a data record/tuple to capture the data uncertainty and thus carry possible worlds semantic. We explore the problem of processing probabilistic top-k queries in distributed wireless sensor networks.

Ranking queries are a powerful concept in focusing attention on the most important answers to a query. To deal with massive quantities of data such as multimedia search, streaming data, web data and distributed systems, tuple from the underlying database are ranked by a score, usually computed based on a user-defined scoring function. Only the top-k tuple with the highest scores are returned for further inspection. Within these motivating application domains distributed, streaming, web and multimedia applications data arrives in massive quantities, underlining the need for ordering by score. But an additional challenge is that the data is also typically inherently fuzzy or uncertain. Data items in the output of such operations are usually associated with a confidence, reflecting how well they are matched with other records from different data sources. In recognition of this aspect of the data, there have been significant research efforts devoted to producing probabilistic database management systems, which can represent and manage data with explicit probabilistic models of uncertainty. With a probabilistic database, it is possible to represent huge number of possible realizations of the data an exponential blow-up from the size of the relation representing the data. A key problem in such databases is how to extend the familiar semantics of the top-k query to this setting, and how to answer such queries efficiently.

3. **Existing System**

This new technology has resulted in significant impacts on a wide array of applications in various fields, including military, science, industry, commerce, transportation, and health-care. However, the quality of sensors varies significantly in terms of their sensing precision, accuracy, tolerance to hardware/external noise, and so on.

**Disadvantages**

- We explore the problem of processing probabilistic top-k queries in distributed wireless sensor networks.
- The wind station very slowly.
- Data is not accuracy purify.

4. **Proposed System**

Based on the notion of necessary sets, we propose a Necessary set based algorithm for in-network processing of PT-Topk queries in a two-tier hierarchical sensor network.
These algorithms exploit individual and combined strengths of sufficient and necessary sets in query processing. For privacy Adopted Secure protocol which prevents Attackers to gain data.

**Advantages**

- Additionally, NSB and BB take advantage of the skewed necessary sets and necessary boundaries among local clusters to obtain their global boundaries, respectively, which are very effective for intercluster pruning.

- We apply sufficient set and necessary set to sensor networks with tree topology, to further improve query processing performance by facilitating sophisticated in-network filtering at the intermediate nodes along the routing path to the root.

- The transmission cost increases for all algorithms because the number of tuple needed for query processing is increased.

3. **System Architecture:**

**Fig:** Evaluation of SecureQ Protocol and NSB Algorithm

**Necessary Set-Based Algorithm**

After receiving all the necessary sets, the base station merges all the received tuple into a table and finds the necessary boundary called the global boundary (GB)). If GB is ranked higher than the highest ranked necessary boundary, it is concluded that all the necessary data have been delivered to the base station. Thus, the base station computes the final answer. Otherwise, entering the second phase, the base station sends the GB back to the cluster heads, which return the supplementary data tuple ranked between its local necessary boundary and GB. Then, the base station computes the final answer.

**Algorithm: NSB ALGORITHM**

**AT CLUSTER HEAD:**

1. Compute the necessary boundary  
   \[ NB(T_i), N(T_i) \leftarrow \{ x | x \leq_f NB(T_i) \land x \leq T_i \} \]

2. Deliver N(Ti) to the base station

3. if cluster head receive GB from the base station
   then  
   \[ N'(Ti) \leftarrow \{ x | x \leq GB \land x \leq [Ti - N(Ti)] \} \]
   Now, N'(Ti) is send to the base station.
   end if

**AT BASESTATION:**

1. It receives the tuple N (Ti) from the cluster head.  
   \[ 1 \leq i \leq N \]
   \[ T' \leftarrow U1 \leq i \leq N N(Ti) \]
2. Now, it will calculate the global boundary.
3. if global boundary GB is less than that of NB(Ti), then
   It calculate the final necessary boundary else
   It will broadcast GB to ci and once again it collects necessary tuple
   \[ T' \leftarrow U_{1 \leq i \leq N'}(T_i) \]
   end if
Where, \( x \) is the tuple ci is the cluster head
\( N(T_i) \) is the necessary set
\( NB(T_i) \) is the necessary boundary
\( T_i \) is the records collected from the sensor
\( N \) is the number of clusters in the zone
\( T' \) is the aggregation of data sets received from the clusters

**SecureQ protocol:**

SecureQ is a Proto-filter that prevents attackers from gaining information from both sensor collected data and sink issued queries. SecureQ also allows a sink to detect compromised storage nodes when they misbehave. To preserve privacy, SecureQ uses a novel technique to encode both data and queries such that a storage node can correctly process encoded queries over encoded data without knowing their values. To preserve integrity, Truth confirmation algorithm to generate integrity verification information so that a sink can use this information to verify whether the result of a query contains exactly the data items that satisfy the query. Operations of SecureQ Proto-filter as a Protocol

1. Select any one of the value from the original data, which is already stored into the storage node.
2. Selected data will be computed (any basic arithmetic operation and it will be selected randomly) with remaining data.
3. Finally, generating new encoded data set, this will be viewed by the unauthorized user.

Operations of SecureQ Proto-filter as a filter

1. Identify the different types of sensor.
2. Dynamically observe the number of active sensors in the sensor networks.
3. Based on the user request, Dynamically Storage nodes stimulate the filter to rearrange the data according to the sensor id.
4. Dynamically SecureQ filter rearrange the data based on the value of the data.
5. Dynamically storage node sends the accurate response data to the user with the help of Proto-filter.

**Truth confirmation algorithm:**

After the sink receives Query Result and Verification Object, it verifies the integrity of Query Result as follows.

First, the sink verifies that every item in Query Result satisfies the query. Second, the sink verifies that the storage node has not excluded any item that satisfies the query.

Let \( \{(En1|1|En1)k_i,\ldots,(Ej-1|Ej)k_i,\ldots,(En2|1|En2)k_j\} \) be the correct query result and Query Result be the query result from the storage node. We consider the following four cases.

1) If there exists \( n1 \leq j < n2 \) such that, \( (Ej_1|1|Ej_1k_i \notin QUERY RESULT \), the sink can detect this error because the
items in QUERY RESULT do not form truth information.
2) If \((E_{n1-1} | E_{n1})_k \notin QUERY RESULT\), the sink can detect this error because it knows the existence of \(E_{n1}\) from \((E_{n1} | E_{n1+1})_k\) and \(E_{n1}\) satisfies the query.
3) If \((E_{n2-1} | E_{n2})_k \notin QUERY RESULT\), the sink can detect this error because it knows the existence of \(E_{n2}\) from the item \((E_{n2} | E_{n2+1})_k\) in VERIFICATION OBJECT and \(E_{n2}\) satisfies the query.
4) If QUERY RESULT=\(\emptyset\), the sink can verify this fact because the item \((E_{n2} | E_{n2+1})_k\) in VERIFICATION OBJECT should satisfy the property \(E_{n2} < a \ L < E_{n2+1}\).

6. Modules
PT-Topk Query Processing

The PT-Topk queries in a centralized uncertain database, which provides a good background for the targeted distributed processing problem. The query answer can be obtained by examining the tuple in descending ranking order from the sorted table (which is still denoted as T for simplicity). We can easily determine that the highest ranked k tuple are definitely in the answer set as long as their confidences are greater than p since their qualifications as PT-Topk answers are not dependent on the existence of any other tuple.

Sensor Networks

The extensive number of research work in this area has appeared in the literature. Due to the limited energy budget available at sensor nodes, the primary issue is how to develop energy-efficient techniques to reduce communication and energy costs in the networks. Approximate-based data aggregation techniques have also been proposed. The idea is to tradeoff some data quality for improved energy efficiency. Silberstein et al. develop a sampling-based approach to evaluate approximate top-k queries in wireless sensor networks. Based on statistical modeling techniques, a model-driven approach was proposed in to balance the confidence of the query answer against the communication cost in the network. Moreover, continuous top-k queries for sensor networks have been studied in and. In addition, a distributed threshold join algorithm has been developed for top-k queries. These studies, considering no uncertain data, have a different focus from our study.

Data pruning

The cluster heads are responsible for generating uncertain data tuple from the collected raw sensor readings within their clusters. To answer a query, it’s natural for the cluster heads to prune redundant uncertain data tuple before delivery to the base station in order to reduce communication and energy cost. The key issue here is how to derive a compact set of tuple essential for the base station to answer the probabilistic top-k queries.

Structured network topology

To perform in-network query processing, a routing tree is often formed among sensor nodes and the base station. A query is issued at the root of the routing tree and propagated along the tree to all sensor nodes. Although the concepts of sufficient set and necessary set introduced earlier are based on two-tier
hierarchical sensor networks, they are applicable to tree-structured sensor networks.

**Data transmission**

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 there is not much difference among SSB, NSB, and BB in terms of query response time, thus we focus on the data transmission costing the evaluation. Finally, we also conduct experiments to evaluate algorithms, SSB-T, NSB-T, and NSB-T-Opt under the tree-structured network topology.

**Performance evaluation**

The performance evaluation on the distributed algorithms for processing PT-top 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.

**7. Conclusion**

In this paper, we propose the notion of sufficient set and necessary set for efficient in-network pruning of distributed uncertain data in probabilistic top-k query processing. Accordingly, we systematically derive sufficient and necessary boundaries and propose a suite of algorithms, namely SSB, NSB, and BB algorithms, for in-network processing of PT-Topk 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. Although our work in this paper is based mainly under the setting of two-tier hierarchical network, the concepts of sufficient set and necessary set are universal and can be easily extend to a network with tree topology. The performance evaluation validates our ideas and shows that the proposed algorithms reduce data transmissions significantly. While focusing on PT-Topk query in this paper, the developed concepts can be applied to other top-k query variants. We plan to develop algorithms to support other probabilistic top-k queries in the future.

**References**


Ms. N.V.R.D. Swathi is a student of Sri Sai Aditya Institute of Science & Technology, surampalem. Presently she is pursuing M.Tech [CSE] from this college and she received her B.Tech degree in Computer Science & Engineering from Chaitanya Institute of Science & Technology, affiliated to JNTUK Kakinada in the year 2011. Her area of interest includes Object Oriented languages, Software Engineering and current trends in Computer Science.

Mrs. K. Naga Bhargavi obtained her B.Tech degree in Computer Science & Engineering from pragati Engineering College, affiliated to Jawaharlal Nehru Technological University Kakinada and M.Tech degree in Computer Science & Engineering from SRKR Engineering College, Bhimavaram, affiliated to Andhra University Visakhapatnam, India. She is working as a Assistant Professor of Computer Science & Engineering at Sri Sai Aditya Institute of Science & Technology, Surampalem, India. She has one international Journal paper. Her areas of interests include Network Security, Cloud Computing and Data mining.