International Journal For Technological Research In Engineering
Volume 1, Issue 9, May-2014 *ISSN (Online): 2347 - 4718*

PROCESSING OF PRESUMPTION TOP-K INQUIRE IN A TREE STRUCTURED NETWORKS

Subhranshu Sekhar Tripathy[1] (MTech), Dr. C. Nalini[2] (Guide)
Department of Computer Science, Bharath University, Tamil Nadu, India.

Abstract— For distributed processing of probabilistic top-k queries in cluster-based tree structured wireless sensor networks we are applying the term of sufficient set, boundary set and necessary set .These two notation have very wide usability which can applicable in localized data pruning in cluster based wireless sensor networks. By keeping these terms, we develop a suite of algorithms, namely, sufficient set-based (SSB), necessary set-based (NSB), and boundary-based (BB), for intercluster query processing with bounded rounds of communications.
Keywords— Top-k inquire, distributed data management, probabilistic databases, wireless sensor networks, distributed query processing, data sniffing, tree structured wireless sensor network.

I. INTRODUCTION

WIRELESS sensor networks are redefining the processes to collect and use data and information from the physical world. This advanced technology has proved in significant impacts on a wide array of applications in various fields, including military, science, industry, commerce, transportation, and health-care. Ranking and aggregation queries used in data exploration, data analysis and decision making scenarios. While most of the currently proposed ranking and aggregation techniques focus on deterministic data, several emerging applications involve data that is unclean and uncertain. Further more uncertainty imposes probability as a new ranking dimension that does not exist in the traditional settings. In wireless sensor networks, sensor nodes play an important role in terms of collection of real word data in to processed output. But the quality of sensors depends upon their sensing precision, accuracy, tolerance to hardware/external noise, and so on. For example, studies show that the distribution of noise varies widely in different photovoltaic sensors, precision and accuracy of readings usually vary significantly in humidity sensors, and the errors in GPS devices can be up to several meters. Thus, sensor readings are inherently uncertain. Process is not visible. In many current applications, data involves uncertainty and impression. The impression of real world data entails different semantics for top k query and top k aggregate queries. It also motivates the need for new query formulation and efficient processing techniques and motivates the need for new query formulations and efficient processing techniques to address ranking and aggregation from a probabilistic perspective. Managing uncertain data via probabilistic database (PDBS) has evolved as an established field of research in recent years with plethora of application ranging from scientific data management, sensor networks, and data integration to knowledge management systems. The basic idea is to perform data pruning and aggregation at cluster heads such that only the data tuples required for final processing are transferred to the base station. Recently top k queries on uncertain data draws sharp attentions in research fields. In uncertain database each data is associated with a scoring function, according to that scoring functions it will top k query will gives the answer for each query. There are a lot of semantics of top-k queries which are being employed in different applications. Among them top-k query semantics and solutions proposed recently, including U-Topk and UkRanks in, PT-Topk, expected rank. The process through which all top k queries are processed is that each query is being associated with scoring attribute and are being processed until satisfies the condition.

II. RELATED WORK

A. A Collaborative Approach to in-Place Sensor Calibration Author: V. Bychkovskiy, S. Megerian

Numerous factors contribute to errors in sensor measurements. In order to be useful, any sensor device must be calibrated to adjust its accuracy against the expected measurement scale. In large scale sensor networks, calibration will be an exceptionally difficult task since sensor nodes are often not easily accessible and Manual device-by-device calibration is intractable. In this paper, we present a two-phase post-deployment calibration technique for large-scale, dense sensor deployments. In its 1st phase, the algorithm derives relative calibration relationships between pairs of co-located sensors, while in the second phase; it maximizes the consistency of the pair-wise calibration functions among groups of sensor nodes. The key idea in the rst phase is to use temporal correlation of signals received at neighboring sensors when the signals are highly correlated (i.e. sensors are observing the same phenomenon) to derive the function relating their bias in amplitude. We formulate the second phase as an optimization problem and present an algorithm suitable for localized implementation. We evaluate the performance of the 1st phase of the algorithm using empirical and simulated data.

B. Poster Abstract: Online Data Cleaning in Wireless Sensor Networks

Author: E. Elnahrawy and B. Nath

We present our ongoing work on data quality problems in sensor networks. Specifically, we deal with the problems of outliers, missing information, and noise. We propose an approach for modeling and online learning of spatial-temporal correlations in sensor networks. We utilize the

learned correlations to discover outliers and recover missing information. We also propose a Bayesian approach for reducing the effect of noise on sensor data online.

C. Evaluating Probabilistic Queries over Imprecise Data
Author: R. Cheng, D.V. Kalashnikov
Many applications employ sensors for monitoring entities such as temperature and wind speed. A centralized database tracks these entities to enable query processing. Due to continuous changes in these values and limited resources (e.g., network bandwidth and battery power), it is often infeasible to store the exact values at all times. A similar situation exists for moving object environments that track the constantly changing locations of objects. In this environment, it is possible for database queries to produce incorrect or invalid results based upon old data. However, if the degree of error (or uncertainty) between the actual value and the Database value is controlled; we can place more confidence in the answers to queries. More generally, query answers can be augmented with probabilistic estimates of the validity of the answers. In this chapter we study probabilistic Query evaluation based upon uncertain data. A classification of queries is made based upon the nature of the result set. For each class, we develop algorithms for computing probabilistic answers. We address the important issue of ensuring the quality of the answers to these queries, and provide algorithms for efficiently pulling data from relevant sensors or moving objects in order to improve the quality of the executing queries. Extensive experiments are performed to examine the effectiveness of several data update policies.

D. On the Representation and Querying of Sets of Possible Worlds
Author: S. Abiteboul, P. Kanellakis
We represent a set of possible worlds using a k-anonymity privacy protection model. We introduce kpro-anonymity model, a space-efficient and complete representation system for finite sets of worlds. We study the problem of efficiently evaluating queries on sets of possible worlds represented by kpro-anonymity model.

E. Efficient Query Evaluation on Probabilistic Databases
Author: N. Dalvi and D. Suciu
We describe a system that supports arbitrarily complex SQL queries on probabilistic databases. The query semantics is based on a probabilistic model and the results are ranked, much like in Information Retrieval. Our main focus is efficient query evaluation, a problem that has not received attention in the past. We describe an optimization algorithm that can compute efficiently most queries. We show, however, that the data complexity of some queries is *#P*-complete, which implies that these queries do not admit any efficient evaluation methods. For these queries we describe both an approximation algorithm and a Monte-Carlo simulation algorithm.

III. PROPOSED SYSTEM

There are three proposed algorithms to minimize the transmission cost. We show the applicability of sufficient set and necessary set to wireless sensor networks in tree-structured network topologies. There are several top-k query semantics and solutions proposed recently, including U-Topk and UkRanks in PT-Topk in PK-Topk in expected rank in and so on. A common way to process probabilistic top-k queries is to first sort all tuples based on the scoring attribute, and then process tuples in the sorted order to compute the final answer set. Nevertheless, while focusing on optimizing the transmission bandwidth, the proposed techniques require numerous iterations of computation and communication, introducing tremendous communication overhead and resulting in long latency. As argued in this is not desirable for many distributed applications, e.g., network monitoring, that require the queries to be answered in a good response time, with a minimized energy consumption. In this paper, we aim at developing energy efficient algorithms optimized for fixed rounds of communications. In this paper, we explore the problem of processing probabilistic top-k queries in distributes wireless sensor network having tree topology. Here, we first use an environmental monitoring application of wireless sensor network to introduce some basics of probabilistic databases. Due to sensing impression and environment interfaces, the sensor readings are usually noisy. An extensive performance evaluation is conducted to compare the proposed algorithms along with two baseline approaches. Experimental result validates our ideas and shows that the proposed algorithms reduce data transmissions significantly without incurring excessive rounds of communications.

IV. ILLUSTRATION OF FEEDBACK SESSIONS

A. Presumption Top-K Inquire
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 tuples 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 tuples 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 tuples.

B. Data snipping
The cluster heads are responsible for generating uncertain data tuples 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 tuples 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 tuples essential for the base station to answer the probabilistic top-k queries.

C. Tree 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 network. This work is based on two tier architecture and in further it can be applicable in the tree structured architecture. This provides a wide platform of adoptability in different type of topology network.

D. Information filtering

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.

E. Performance evaluation

The performance evaluation on the distributed algorithms for processing PT-top k queries in tree 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 which is an important point to be remembered. Based on that remembered. An extensive performance evaluation is conducted to compare the proposed algorithms along with two baseline approaches. Experimental result validates our ideas and shows that the proposed algorithms reduce data transmissions significantly without incurring excessive rounds of communications. We will 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

V. ASSOCIATED WORK

In recent years, many works have been done to Here; we review representative work in the areas of 1) top-k Query processing in wireless sensor networks, and 2) top-k query processing on uncertain data. Top-k query processing in sensor networks. An extensive number of research works in this area has appeared in the literature [21], [24], [25], [26]). 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 [21] 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 [27], [25]. The idea is to tradeoff some data quality for improved energy efficiency. Silberstein et al. develop a sampling-based approach to evaluate approximate on statistical modeling techniques, a model-driven approach was proposed in [5] 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 [28] and [29]. In addition, a distributed threshold join algorithm has been developed for top-k queries [24]. These studies, considering no uncertain data, have a different focus from our study. Top-k query processing on uncertain data. While research works on conventional top-k queries are mostly based on some deterministic scoring functions, the new factor of tuple membership probability in uncertain databases makes evaluation of probabilistic top-k queries very complicated since the top-k answer set depends not only on the ranking scores of candidate tuples but also their probabilities [8]. For uncertain databases, two interesting top-k definitions (i.e., U-Topk and U-kRanks) and A_-like algorithms are proposed [17]. U-Topk returns a list of k tuples that has the highest probability to be in the top-k list over all possible worlds. U-kRanks returns a list of k tuples such that the ith record has the highest probability to be the ith best record in all possible worlds. In [13], PT-Topk query, which returns the set of tuples with a probability of at least p to be in the top-k lists in the possible worlds, is studied. Inspired by the concept of dominate set in the top-k query, an algorithm which avoids unfolding all possible worlds is given. Besides, a sampling method is developed to quickly compute an approximation with quality guarantee to the answer set by drawing a small sample of the uncertain data. In [19], the expected rank of each tuple across all possible worlds serves as the ranking function for finding the final answer. In [30], U-Topk and U-kRank queries are improved by exploiting their stop conditions. In [31], all existing top-k semantics have been unified by using generating functions. Recently, a study on processing top-k queries over a distributed uncertain database is reported in [14] and [23]. Li et al. [14] only support top-k queries with the expected ranking semantic. On the contrary, our proposal is a general approach which is applicable to probabilistic top-k queries with any semantic. Furthermore, instead of repeatedly requesting data which may last for several rounds, our protocols are guaranteed to be completed within no more than two rounds. These differences uniquely differentiate our effort from [14]. Our previous work [23] as the initial attempt only includes the concept of sufficient set. In this paper, besides of sufficient set, we propose another important concept of necessary set. With the aid of these two concepts, we further develop a suite of algorithms, which show much better performance than the one in [23]. Probabilistic ranked queries based on uncertainty at the attribute level are studied in [32], [33], and [19]. A unique study that ranks tuples by their probabilities satisfying the query is presented in [12]. 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 [18]. We will 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.

VI. EXPERIMENTAL SETUP AND RESULT

Here, we first conduct a simulation-based performance evaluation on the distributed algorithms for processing PT-topk queries in tree structured 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. To perform the necessity test we have to take a system, which should be having the features for hardware part Processor Pentium –IV Speed will be 1.1 GHz RAM 256 MB (min) Hard Disk must be20 GB Key Board Standard Windows Keyboard Two or Three Button Mouse Monitor SVGA need. The software configuration should be. Operating System on which simulation is going to be performed is the Windows XP Programming Language we are using is JAVA Version JDK 1.6 & above. DATABASE, MYS, tool is Net beans IDE 7.0 to perform the desire experiment first we have to collect local data from the environment through local sensors. After receiving the necessity data these are processed to the cluster head. There it's having probabilistic data base. Which functions to store all the processed data? Networks, sensor nodes are grouped into clusters, where cluster heads are responsible for local processing and to report aggregated results to the base station. Each tuple records a possible wind speed corresponding to a location. The data base is consisting of most probable of four columns and the number of rows is depending on the number of local sensors. Columns are named as location, speed of the wind and confidence values associated with each tuple. Confidence value Associated with a tuple indicates the existence probability of that particular wind speed. For example, there are two data Tuples generated for Location A. The wind speeds in these Two tuples are both valid (i.e., with measured confidences) But their presence is exclusive. In probabilistic databases, the exclusive presence of tuple instances in an x-tuple _ is Dictated by the x-relation rule, denoted in the form of ¼ ftuple1; tuple2; . . . g. In our running environment. Each Location is associated with an x-relation rule. For instance, Corresponding to location A, _A ¼ ft1; t3g, i.e., _A is an xtuple for A and t1 and t3 are alternative tuple instances (or simply called alternatives) of _A. That the overall confidence of an x-tuple is the sum of its alternatives. A number of experiments is conducted to validate the proposed algorithms for a two-tier network in the following aspects: 1) overall performance,2) sensitivity tests3) adaptiveness. Additionally, overall performance under the tree topology is evaluated. We first validate the effectiveness of our proposed methods in reducing the transmission cost against two baselines 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 Parameter Settings processing strategy explored. The iterative approach runs as follows: in each round, each cluster head delivers a single data tuple with the highest score in its local data set and the information of current local highest score (after removing the delivered data) to the base station. In response, the base station derives necessary set upon the data collected so far. If the score of necessary boundary is higher than the current local highest score of all cluster heads, the base station can determine the final result based on the available data and thus terminate the algorithm; otherwise, the algorithm goes for the next round. The experiments use both the synthetic data and real traces. Here we randomize all parameter settings within their ranges to show the general performance. Both results on synthetic and real data indicate that all the algorithms proposed in this paper significantly reduce data transmissions against the two baseline approaches by achieving about 70 percent of saving against the Iterative approach in synthetic data sets and 80 percent of saving against the Iterative approach in real traces . One important reason that our approaches outperform the Iterative approach is due to the small and constant query processing rounds in our approaches. In our experiment, our algorithms complete within two rounds; while the iterative approach incurs about 60-200 rounds. Note that the 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. Next, we examine the impact of a variety of query and system parameters on the performance of the proposed algorithms. In the plots, we do not show the result of baseline approaches for clarity of presentation. We also omit the plots of experiments on real traces due to space limitation. Here we first show the impact of query parameters, i.e., kind p, on performance. Figure shows the trend of transmission cost by varying k from 2 to 10. As shown, the transmission cost increases for all algorithms because the number of tuples needed for query processing is increased. Among the SSB; NSB, and BB algorithms, BB does not perform as well as others when k is small but it becomes a good choice when k becomes larger. It shows the trend of transmission cost by varying query threshold p from 0.2 to 0.7. Performance of all algorithms improves as the threshold p increases because a high threshold in general reduces the size of the result set and hence the amount of data transmitted to the base station. Among the algorithms, SSN performs the best when p is small but becomes significantly worse than the others when p becomes large. This is because NSB and BB are able to use the tighter threshold to filter out more unqualified tuples and reach a tighter necessary boundary. On the other hand, the

sufficient boundary does not benefit as much. Finally, we observe that the adaptive algorithm matches very well with NSB. Next; we present results of our sensitivity tests on a number of system Parameters and data distribution. We examine how the algorithms scale up by increasing the number of clusters in the system from 4, 16, 36, 64, to 100.The result, shows scales up better than NSB and SSB. As the number of clusters increases, the total amount of data transmission rises accordingly. Realizing intracluster and intercluster pruning before data transmission is an important factor for BB to outperform others in this examination. The same observation can also be made in the sensitivity test on the size of data tuples. As the primary cost in the network is tuple transmission, SSB exhibits a linear increase of transmission cost at a higher rate than NSB and BB. Again, the result shows that BB is the best choice when the tuple. The experimental result aligns with conclusion from experiments on synthetic data. Size is large. Overall, the adaptive algorithm matches up the best algorithm under all circumstances. The impact of data skewness on transmission cost. In the experiments, when the skewness is high the difference among readings in different zones is increased. As shown, the data skewness helps intracluster data pruning as the performance of every algorithm improves when the data distribution becomes more skewed. 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 forinter cluster pruning. Particularly, when _ ¼ 1, BB shows the best performance in terms of transmission cost. Thus, adaptive algorithm approaches BB under this scenario. Shows the impact of the size of x-tuples on the Performance. When _ is small, SSB shows a good Performance because the probability of highly ranked tuples in an x-tuple is higher. Thus adaptive algorithm chooses SSB as the best candidate. As _ grows, there are more tuples in an x-tuple. Thus, the size of sufficient set is linearly increased as more tuples need to be considered to set the sufficient boundary. On the other hand, although the size of necessary set also increases, it does not increase linearly since many low-probability tuples are filtered out by the probability threshold. By utilizing both sufficient and necessary boundaries, the performance of BB also gets worsen but in a slower pace due to the smaller necessary set. However, due to the overhead for exchanging boundary information, BB may not perform as well as NSB, particularly with light data skewness. Based on the above results, we find that NSB and BB generally exhibit better performance than SSB does. However, there are no clear winners for NSB and BB. As we discussed above, with larger network scale (M), larger tuplesize (Sd), or extreme skewed data distribution, BB usually outperforms NSB. However, in other cases, NSB Shows better performance over BB. The adaptive algorithm reaches our expectation by achieving the least transmission Cost under all circumstances. In this section, we further test its addictiveness to dynamic sensor network environments. One important factor that has an impact on the adaptive algorithm is the size of tuning window. To figure out the optimal setting of tuning window size, we conduct an Experiment by varying the window size from 1; 10; 20; . . . , to 90. To mock the dynamics of readings, we change _ setting among f0; 0:5; 1:0g every 50 round. As shown in Fig. 10a, there exists an optimal window size for adapting to the dynamic readings. If the window size is too small, the adaptation will be too sensitive and thus fail to get the optimal performance. On the other hand, if the window size is too large, the adaptation may lag behind the change of readings. Additionally, we conduct an experiment to show the behaviors of algorithms under a dynamic environment we simulate. Fig. 10b shows that the adaptive algorithm switches timely to match the best algorithms. As shown in the figure, adaptive algorithm switches from the NSB algorithm to the BB algorithm at about time 10, then returns back to NSB at about time 20, and finally switches to the BB algorithm again at about time 30. While the different algorithms may outperform each other at different time, the adaptive algorithm adapts to the dynamic changes to achieve the least transmission cost. Finally, we evaluate SSB-T, NSB-T, and NSB-T-Opt in a tree structured network. Experimental settings are the same as the ones used in two-tier hierarchical network. As shown in Fig. 11, SSB-T, NSB-T, and NSB-T-Opt show much better performance against the two baseline approaches. Notice that, SSB-T outperforms NSB-T, as NSB-T incurs extra overhead for the delivery of supplementary data in the second round. On the other hand, NSB-T-Opt address this issue through in-network suppressing of unnecessary supplementary data Transmission. From the experiment, we observe that the query for the supplementary data never reaches the leaf nodes, i.e., requests for supplementary data are fulfilled by some intermediate nodes along the routing paths. In this way, NSB T-Opt indeed optimizes the performance of NSB-T and shows excellent performance.

VII. CONCLUSION

Our experimental study shows the efficiency of our techniques under different data distributions with orders of magnitude improvement over native methods. By implementing the concept of sufficient set, boundary set and necessary set for efficient in –network pruning of distributed uncertain data in probabilistic top-k query processing. . Based on the proposed problematic statement, design a cost model based on communication of three proposed algorithms and propose a cost based adaptive algorithm that automatically adjust to the application dynamics. In this work we are using tree structured network, the concepts of sufficient set and necessary set are universal and can easily implement in two tire hierarchical topology. The performance evaluation satisfies our ideas and proved that the proposed algorithm decrease data transmissions significantly.

REFERENCES

[1] V. Bychkovskiy, S. Megerian, D. Estrin, and M. Potkonjak, "A Collaborative Approach to in-Place Sensor Calibration," Proc. Second Int'l Conf. Information Processing in Sensor Networks (IPSN), pp.

301-316, 2003.
[2] http://www.veriteq.com/, 2012.
[3] 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.
[4] 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.
[5] 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.
[6] 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.
[7] N. Dalvi and D. Suciu, "Efficient Query Evaluation on Probabilistic Databases," Proc. 30th Int'l Conf. Very Large Data Bases (VLDB '04), pp. 864-875, 2004.
[8] A.D. Sarma, O. Benjelloun, A. Halevy, and J. Widom, "Working Models for Uncertain Data," Proc. 22nd Int'l Conf. Data Eng. (ICDE '06), p. 7, 2006
[9] 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.
[10] Y. Tao, R. Cheng, X. Xiao, W.K. Ngai, B. Kao, and S. Prabhakar, "Indexing Multi-Dimensional Uncertain Data with Arbitrary Probability Density Functions," Proc. 31st Int'l Conf. Very Large Data Bases (VLDB '05), pp. 922-933, 2005
[11] C. Re, N. Dalvi, and D. Suciu, "Efficient Top-k Query Evaluation on Probabilistic Data," Proc. Int'l Conf. Data Eng. (ICDE '07), pp. 896-905, 2007.
[12] M. Hua, J. Pei, W. Zhang, and X. Lin, "Ranking Queries on Uncertain Data: A Probabilistic Threshold Approach," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '08), 2008.
[13] F. Li, K. Yi, and J. Jestes, "Ranking Distributed Probabilistic Data," Proc. 35th SIGMOD Int'l Conf. Management of Data (SIGMOD '09), 2009.
[14] H.W. Rabiner, C. Anantha, and B. Hari, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," Proc. 33rd Hawaii Int'l Conf. System Sciences (HICSS '00), 2000.
[15] Y. Diao, D. Ganesan, G. Mathur, and P.J. Shenoy, "Rethinking Data Management for Storage-Centric Sensor Networks," Proc. Conf. Innovative Data Systems Research (CIDR '07), pp. 22-31, 2007.
[16] M.A. Soliman, I.F. Ilyas, and K.C. Chang, "Top-k Query Processing in Uncertain Databases," Proc. Int'l Conf. Dat Eng (ICDE '07), 2007.
[17] C. Jin, K. Yi, L. Chen, J.X. Yu, and X. Lin, "Sliding-Window Top-k Queries on Uncertain Streams," Proc. Int'l Conf. Very Large Data Bases (VLDB '08), 2008.
[18] G. Cormode, F. Li, and K. Yi, "Semantics of Ranking Queries for Probabilistic Data and ExpectIEEE Int'l, 2009.
[19] P. Cao and Z. Wang, "Efficient Top-k Query Calculationin Distributed Networks," Proc. 23rd Ann. ACM Symp. Principles of Distributed Computing (PODC), pp. 206-215, 2004.
[20] S. Madden, M.J. Franklin, J. Hellerstein, and W. Hong, "TAG: A Tiny A Ggregation Service for Ad-Hoc Sensor Networks," (OSDI '02), 2002.
[21] Y. Xu, W.-C. Lee, J. Xu, and G. Mitchell, "Processing Window Queries in Wireless Sensor Networks," Proc. IEEE 22nd Int'l Conf. Data Eng. (ICDE '06), 2006.
[22] M. Ye, X. Liu, W.-C. Lee, and D.L. Lee, "Probabilistic Top-k Query Processing in Distributed Sensor Networks," Proc. IEEE Int'l Conf. Data Eng. (ICDE '10), 2010.
[23] D. Zeinalipour-Yazti, Z. Vagena, D. Gunopulos, V. Kalogeraki, V. Tsotras, M. Vlachos, N. Koudas, and D. Srivastava, "The Threshold Join Algorithm for Top-k Queries in Distributed Sensor Networks," Proc. Second Int'l Workshop Data Management for SensorNetworks (DMSN '05), pp. 61-66, 2005
[24] A. Sharaf, J. Beaver, A. Labrinidis, and K. Chrysanthis, "Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks," Int'l J. Very Large Data Bases, vol. 13, no. 4, pp. 384-403,2004.
[25] A.S. Silberstein, R. Braynard, C. Ellis, K. Munagala, and J. Yang, "A Sampling-Based Approach to Optimizing Top-k Queries in Sensor Networks," Proc. 22nd Int'l Conf. Data Eng. (ICDE'06),p. 68, 2006.
[26] Q. Han, S. Mehrotra, and N. Venkatasubramanian, "Energy Efficient Data Collection in Distributed Sensor Environments," Proc. 24th Int'l Conf. Distributed Computing Systems(ICDCS'04),pp. 590-597, 2004.
[27] M. Wu, J. Xu, X. Tang, and W.-C. Lee, "Top-k Monitoring in Wireless Sensor Networks," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 7, pp. 962-976, July 2007.
[28] D. Wang, J. Xu, J. Liu, and F. Wang, "Mobile Filtering for Error-Bounded Data Collection in Sensor Networks," Proc. 28th Int'l Conf. Distributed Computing Systems (ICDCS '08), pp. 530-537, 2008.
[29] K. Yi, F. Li, G. Kollios, and D. Srivastava, "Efficient Processing of Top-k Queries in Uncertain Databases with X-Relations," IEEE pp. 1669 1682, Dec.2008.
[30] J. Li, B. Saha, and A. Deshpande, "A Unified Approach to Ranking in Probabilistic Databases," Proc. Int'l Conf. Very Large Data Bases (VLDB), vol. 2, no. 1, pp. 502-513, 2009.
[31] X. Lian and L. Chen, "Probabilistic Ranked Queries in Uncertain Databases," Proc. 11th Int'l Conf. Extending Database Technology (EDBT '08), pp. 511-522, 2008.
[32] M.A. Soliman and I.F. Ilyas, "Ranking with Uncertain Scores," Proc. IEEE Int'l Conf. Data Eng. (ICDE '09), 2009.
[33] X. Liu, J.Xu, and W.-C. Lee, "A Cross Pruning Framework for Top-k Data Collection in Wireless Sensor Networks," Proc. 11th

[34] Distributed Processing of Probabilistic Top-k Queries in Wireless Sensor Networks Mao Ye, Wang-Chien Lee, Member, IEEE, Dik Lun Lee, Member, IEEE, and Xingjie Liu, Student Member, IEEE , VOL. 25, NO. 1, JANUARY 2013

