ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

Data Management in Wireless Sensor Network: A Survey

Mandar Mokashi[1], Dr.A.S.Alvi[2]

Final Year Master of Engineering, Department of CSE, PRMIT & R, Bandera (M.S), India[1]

Professor and Head, Department of IT, PRMIT & R, Badnera (M.S), India[2]

Abstract: A wireless sensor network (WSN) is a network of distributed sensors grouped together to monitor physical or environmental conditions, like temperature, pressure, sound etc. and to pass their sensed values through the network to a main location (sink) cooperatively. The development of wireless sensor networks was first motivated by military applications such as battlefield surveillance; today such networks are used in several industrial non industrial and consumer applications, such as industrial process monitoring and control, machine observation, health monitoring, and so on.

Keywords: Sensor network, Wireless network, sensor node, data management in wireless sensor networks, Query processing

I. INTRODUCTION

A sensor network is viewed as a distributed database that collects physical measurements about the environment indexes them, and then serves queries from users. Each sensor node typically generates a stream of data items that are obtained from the sensing devices on the node [6]. Generally, each sensor consists of a small node with sensing, computing, and communication capabilities.

II. STRUCTURE OF SENSOR NODE

sensor nodes are distributed (typically, randomly) in a geographical region. each sensor node is equipped with sensing devices, a short-range radio, and a limited battery. two sensor nodes can either communicate with each other directly (if within each other's transmission radius) or indirectly using intermediate nodes. the data generated in a sensor network is simply the readings of the sensing devices on the nodes, and can be modeled as relational data streams

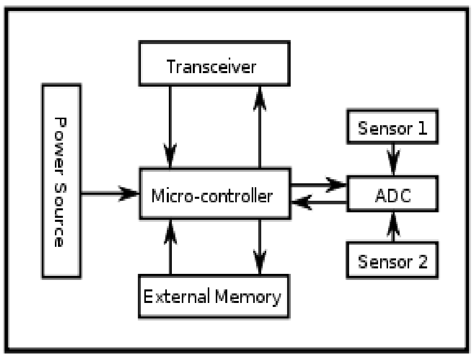

Fig 1. The typical architecture of the sensor node. [27]

[17] is used to processing aggregation in in-network manner. Also, Tiny DB supports metadata management, in-network persistent storage, and multiple concurrent

III. BASIC APPROACHES TO DATA MANAGEMENT IN WSNS

First, we introduce some early data management approaches. In [14], a scalable and robust communication paradigm, directed diffusion, is proposed. Attribute-value pairs are used to name data generated by sensor nodes. A request node sends its interests of named data to destination sensor nodes or regions. Then, data satisfying the interest are returned along the reverse path of interest dissemination to the request node. To improve the performance and save energy, intermediate nodes can cache data and might aggregate the data.[15] extends and improves the directed diffusion method especially on experiments. Tiny OS [16] is a free and open source operating system designed for WSNs. Tiny OS is an embedded operating system which is written in the nesC. The component library of Tiny OS consists of sensor drivers, network protocols, distributed services, and data acquisition tools. Therefore, Tiny OS can support basic data requests. Moreover, users can develop their own applications based on Tiny OS. Tiny DB [4] is a data management system for WSNs based on Tiny OS. It can extract information from WSNs by sending queries. Importantly, Tiny DB allows users describe the data they want to acquire by writing a simple, SQL-like query. For answering a query, Tiny DB requests the data from sensor nodes in the network and routes it back to a PC. In the phase of processing queries, filtering and aggregation algorithms might be used. Tiny DB uses intelligent in-network energy-efficient processing algorithms to prolong the network lifetime. For instance, tree-based routing is used for query delivery, data collection, and in-network aggregation, and queries. REED [18] extends Tiny DB with the ability to process joins operations between sensing data and static tables which is built outside the WSN. Filter conditions is

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

stored in static tables, and then those tables are distributed throughout the network. Join operation is executed in in-network manner. REED is also suitable for various event detection applications which Tiny DB and data collection systems cannot handle. Cougar [19] is another distributed database system to sensor networks. It considered query languages, aggregation processing, query optimization, catalog management, and multi-query optimization. When a new query is coming, query optimizer of Cougar either merge it with an existing query or generate a new query plan. A leader is chosen to control the execution of the query plan.

III. DATA COLLECTION

Data collection is widely used for applications, which collects all sensed data continuously (SELECT! query). In, Chu et al. [23] have proposed a mechanism, Ken, using conditional data transmission to conserve energy by reporting only if the difference between the sensed Data collection is widely used for applications, which collects all sensed data continuously (SELECT ! query). In [23], Chu et al. have proposed a mechanism, Ken, using conditional data transmission to conserve energy by reporting only if the difference between the sensed value and the predicted value is beyond certain bounds. The replicated dynamic probabilistic model for a sensor node is installed on both the sink and that sensor node. A sensor node does not need to report sensed value normally if the predicted error is within the apparently, it is easy to store and access the data at Base station (BS). However, such methods (such as [31] [17] [32]) might be more applicable to. Templates for LaTeX and Microsoft Word. The LaTeX templates depend on the official IEEE trancls and IEEEtran.bst files, whereas the Microsoft Word templates are self-contained. Causal Productions has used its best efforts to ensure that the templates have the same appearane. value and the predicted value is beyond certain bounds. The replicated dynamic probabilistic model for a sensor node is installed on both the sink and that sensor node. A sensor node does not need to report sensed value normally if the predicted error is within the threshold user specified, thus saving reporting communication cost. Once the predication error is greater than the error bound, the sensed data is reported to the sink and the parameters of the model are adjusted to match current data changing intend. The effectiveness of this scheme in reducing the communication cost is relied on the predication model chosen. As shown in [24], the prediction models based on the temporal and spatial correlations of data work extremely well in WSNs. A novel disjoint-Cliques Model is also proposed to use special correlation to reduce data transmission and improve prediction accuracy. In [25],a data-driven approach was presented. Model-based suppression is used to provide continuous data without continuous reporting. In addition, a key problem for data suppression, link failure, is addressed. A mobile filtering approach for error-bounded data collection was proposed in [26]. By migrating filters wisely the number of data reporting is reduced significantly. Jain et al. have built dynamic procedures that employ maximum filtering of data using a technique called stochastic recursive data filtering, to conserve resources subject to meeting precision standards [27]. Primarily, these methods are aimed at reducing energy consumption by substituting data acquisition using data estimation. Another method of data estimation is based on collection of data samples for a relatively long time and calculating the autocorrelation of the vector of samples [28]. This approach aims at enabling nodes to identify patterns in the behaviour of sensed processes and report only uncommon observable data to conserve energy. Two key issues in Data Management in Sensor Networks are Data Storage (how to store data efficiently) and Query Processing (how to achieve fast and accurate information retrieval)

V. DATA STORAGE

Many approaches have been proposed to describe how to store data generated by WSNs. One category of such storage solutions is that base station collects and stores all data as [31] [17] [32]) might be more applicable to answer continuous queries. Obviously, the mortal drawback of collecting all data is shortening the limited power supply since sending packets costs large amount of energy. Also, shipping all data out of network may be not necessary because users are not interested in all data. Sensor nodes near BS become the bottleneck as a result of forwarding packets. Moreover, the WSNs may not able to transfer all data continuously generated by sensor nodes due to the limitation of bandwidth. Since centralized storage is not practical, naturally distributed in-network data storage is considered. Nevertheless, in-network data storage and data retrieve of WSNs are challenging issues for WSNs because each sensor node just has limited memory space and there are a number of sensor nodes in a WSN normally. For improving network lifetime, in-network storage techniques have been addressed to solve ad-hoc queries. These frameworks are primarily based on the Data-Centric Storage (DCS) concept [33]. In DCS, relevant data are categorized and named according to its meanings (e.g., tiger sightings). All data with the same general name will be stored at the same sensor node. Then, when users query the data with a particular name, it can be sent directly to the sensor node who stores those named data. The major difference among in network DCS schemes is using different events-to-sensors mapping methods. The mapping was designed using hash tables in DHT [33] and GHT [34], or using k-d trees in DIM [35],KDDCS [36], and STDCS [11]. STDCS use sensor location as data indexing instead of the sensed values. Hence, STDCS addresses a sensors-to-sensors mapping instead of the readings/events-to-sensors mapping. Moreover, a switching-time is defined to be the time duration after which the mapping function changes. STDCS uses a spatiotemporal indexing to balance query load among sensors.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

As we know, indexing techniques can significantly improve the data acquiring/query performance. For WSNs, another benefit of using index is reducing cost of data request dissemination since the destination of data request can be obtained from A Data Management Tool called ES3N[1]uses Semantic Web techniques to manage and query network lifetime for WSNs with the index. The works in [37], [35], and [36] use a spatially distributed hashing index technique to solve range queries in a multi-dimensional space. The work in [12] proposed a distributed spatial temporal index structure to track moving objects. The work in [38] addressed a time-based index arrangement for event query processing. For saving energy, the index is adjusted according to the event frequency. Specially, most of these approaches just store partial data, which satisfy conditions or present events and moving objects, generated by sensors [34], [36], [35], [37], [12], and [38].To our best knowledge, no in-network distributed historical data storage, indexing, and query processing schemes have been presented in the literature. Since sensors have limited memory space, storing all historical data on sensors might not be applicable for some applications with a large quantity of sensing data. However, most applications desire statistical results, events, the trend of value changing of historical data rather than specific values. This characteristic provides us a chance to store sketch information to answer queries based on historical data of sensor networks. For enhancing query processing performance and saving energy, a distributed index is necessary to guide query forwarding. In [39], we proposed a in-network historical data storage and query processing scheme based on distributed index.

This scheme stores historical data locally and processes queries energy-efficiently by using a distributed index tree. Regression techniques can be used to save memory capacity according to data characteristics and users' requirements. In order to process queries quickly and energy-efficiently, in-network distributed tree-based indexes are constructed and maintained. Using indexes to process queries can significantly reduce the number of involved sensors, thus conserving energy consumption. Furthermore, for avoiding load skew, index data are partitioned on different nodes according to their time stamp

IV. QUERY PROCESSING

Data collected from a minidome Sensor Network. Our tool supports complex queries on both continuous and archival data, by capturing important associations among data, collected and stored in a distributed dynamic ontology Sensor Networks are typically able to process a vast range of queries fairly efficiently using existing techniques. The Tiny DB project is based on a query language that supports basic, aggregate, temporal aggregate, event based and even lifetime [1]proposes SPARQL query language For Query Processing, RDF is a directed, labelled graph data format for representing information in the Web. This specification defines the syntax and semantics of the SPARQL query language for RDF. SPARQL can be used to express queries across diverse data sources, whether the data is stored natively as RDF or viewed as RDF via middleware. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. SPARQL also supports extensible value testing and constraining queries by source RDF graph. The results of SPARQL queries can be results sets or RDF graphs. In fact, SPARQL is a Semantic based querying capabilities. This language supports a range of query types, including monitoring, network health, exploratory, actuation and offline delivery queries. From a Semantic Web perspective, these querying types can be further leveraged by discovering and using Semantic Associations between fragments of data, to present greater query richness. The results of SPARQL queries can be results sets or RDF graphs. In fact, SPARQL is a Semantic Web candidate recommendation presently SPARQL is embedded in Jena, which is a Java framework for building Semantic Web applications that provides a programmatic environment for RDF, RDFS and OWL, including a rule-based inference engine [8]. Our Ontology Schema, once imported in main memory, creates an ontology model6, to which formatted streaming data are added as *resources*. SPARQL queries capture relationships among data triples.

VII. CONCLUSION

Wireless sensor networks are enabling applications that previously were not practical. As new standards based networks are released and low power systems are continually developed, we will start to see the widespread deployment of distributed databases in wireless sensor networks. Sensor nodes can be imagined as small computers, extremely basic in terms of their interfaces and their components. All of this sensor network research is producing a new technology which is already appearing in many practical areas. Now a days we can observe a vide scope and various implementations of distributed database management in wireless sensing devices.

REFERENCES

[1] Micah Lewis, Delroy Cameron, Shaohua Xie2, I. Budak Arpinar1 "ES3N: A Semantic Approach to Data Management in Sensor Networks"pp.8,2004
[2] S. R. Madden, M. J. Franklin, J. M. Hellerstein, andW. Hong, "Tinydb: an acquisitional query processing system for sensor networks," ACM Trans. Database Syst., vol. 30, no. 1, pp. 122–173, 2005
[3] Jena – A Semantic Web Framework for Java. jena.sourceforge.net
[4] M. Aly, A. Gopalan, J. Zhao, and A. Youssef, "Stdcs: A spatio-temporal data-centric storage scheme for real-time sensornet applications," in Sensor, Mesh and Ad Hoc Communications and Networks, 2008. SECON '08. 5th Annual IEEE Communications Society Conference on, June 2008, pp. 377–385.
[5] A. Meka and A. Singh, "Dist: a distributed spatio-temporal index structure for sensor networks," in CIKM '05: Proceedings of the 14th

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 3, March 2013

ACM international conference on Information and knowledge management. NewYork,NY,USA:ACM,2005,pp.139–146.
[6] R. G. C. Intanagonwiwat and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks," in Proceedings of the sixth annual international conference on Mobile computing and networking, Boston, MA USA, 2000, pp.56–67.
[7] J. Heidemann, F. Silva, C. Intanagonwiwat, R. Govindan, D. Estrin, and D. Ganesan, "Building efficient wireless sensor networks with low-level naming," in SOSP '01: Proceedings of the eighteenth ACM symposium on Operating systems principles. NewYork,NY, USA: ACM, 2001, pp. 146–159.
[8] "Tinyos website," available: http://www.tinyos.net/ [Accessed August 15, 2009].
[9] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "Tag: a tiny aggregation service for ad-hoc sensor networks," in SIGOPS Oper. Syst. Rev., vol. 36, no. SI, June 2002, pp. 131–146.77
[10] D. J. Abadi, S. Madden, and W. Lindner, "Reed: robust, efficient filtering and event detection in sensor networks," in VLDB '05: Proceedings of the 31st international conference on Very large data bases. VLDBEndowment,2005,pp.769–780.
[11] Y. Yao and J. Gehrke, "The cougar approach to in-network query processing in sensor networks," SIGMOD Rec., vol. 31, no. 3, pp. 9–18, 2002.
[12] D. Chu, A. Deshpande, J. M. Hellerstein, and W. Hong, "Approximate data collection in sensor networks using probabilistic models," in ICDE '06: Proceedings of the 22nd International Conference on Data Engineering. Washington, DC, USA: IEEE Computer Society, 2006, p. 48.
[13] A. Deshpande, C. Guestrin, S. R. Madden, J. M. Hellerstein, and W. Hong, "Modeldriven data acquisition in sensor networks," in VLDB '04: Proceedings of the Thirtieth international conference on Very large data bases. VLDBEndowment,2004,pp.588–599.
[14] A. Silberstein, R. Braynard, G. Filpus, G. Puggioni, A. Gelfand, K. Munagala, and J. Yang, "Data-driven processing in sensor networks," in CIDR '07: Proceedings of the 3rd Biennial Conference on Innovative Data Systems Research. Asilomar, California,USA: IEEE Computer Society, 2007.
[15] D. Wang, J. Xu, J. Liu, and F. Wang, "Mobile filter: Exploring migration of filters for error-bounded data collection in sensor networks," Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on, pp. 1483–1485, April 2008.
[16] A. Jain, E. Y. Chang, and Y. Wang, "Adaptive stream resource management using kalman filters," in SIGMOD Conference, 2004, pp. 11–22.
[17] D. de O. Cunha, R. P. Laufer, I. M. Moraes, M. D. Bicudo, P. B. Velloso,and O. C. M. B. Duarte, "Bio-inspired field estimation with wireless sensor networks," Universidade Federal do Rio de Janeiro, Tech. Rep.,http://www.gta.ufrj.br/ftp/gta/TechReports/CLMB05a.pdf, accessed November 03,2007.
[18] P. B. Johannes, J. Gehrke, and P. Seshadri, "Towards sensor database systems," 2001,pp. 3–14.
[19] J. Gehrke and S. Madden, "Query processing in sensor networks," in Proceedings of the First Biennial Conference on Innovative Data Systems Research. CIDR' 03, 2003.
[20] S. Shenker, S. Ratnasamy, B. Karp, R. Govindan, and D. Estrin, "Data-centric storage in sensornets," SIGCOMM Comput. Commun. Rev., vol. 33, no. 1, pp. 137–142, 2003.
[21] S. Ratnasamy, B. Karp, L. Yin, F. Yu, D. Estrin, R. Govindan, and S. Shenker, "Ght:a geographic hash table for data-centric storage," in WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications. NewYork,NY, USA: ACM, 2002, pp. 78–87.
[22] X. Li, Y. J. Kim, R. Govindan, and W. Hong, "Multi-dimensional range queries in sensor networks," in SenSys '03: Proceedings of the 1st international conference on Embedded networked sensor systems. NewYork,NY,USA:ACM,2003,pp.63–75.
[23] M. Aly, K. Pruhs, and P. K. Chrysanthis, "Kddcs: a load-balanced in-network datacentric storage scheme for sensor networks," in CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management. NewYork,NY,USA: ACM, 2006, pp. 317–326.
[24] B. Greenstein, D. Estrin, R. Govindan, S. Ratnasamy, and S. Shenker, "Difs: a distributed index for features in sensor networks," in Sensor Network Protocols and Applications,2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on, May 2003, pp. 163–173.
[25] G. Li, J. Li, and Y. Li, "Time: Time-based index management for event query processing in wireless sensor networks," in Performance, Computing and Communications Conference, 2008. IPCCC 2008. IEEE International, Dec. 2008, pp. 1–8.
[26] C. Ai, R. Du, M. Zhang, and Y. Li, "In-network historical data storage and query processing based on distributed indexing techniques in wireless sensor networks," in WASA '09: Proceedings of the International Conference on Wireless Algorithms, Systems and Applications. Boston, MA, USA: Springer Berlin / Heidelberg, August 2009,pp. 264–273.
[27] "Sensor network". Wikipedia, the free encyclopedia. en.wikipedia.org (2006)

BIOGRAPHY

Prof.Dr.A.S.Alvi is currently working as a Professor & Head, in Department of Information and Technology, at Prof.Ram Meghe institute of Technology and Research, Badnera(M.S.),India. He has published several research papers in reputed National and International journals. His areas of interests are Discrete Structures, DBMS, and Netwok Security.

Mr.M.K.Mokashi is currently pursuing his Master of Engineering from Prof. Ram Meghe Institute of Technology, Badnera (India) from Sant Gadge Baba Amaravati University, Amaravati. His subjects of interests are data ase management systems, software engineering and computer networks