Journal of the Korea Convergence Society
Vol. 7. No. 6, pp. 43-48, 2016
http://dx.doi.org/10.15207/JKCS.2016.7.6.043

A Data Driven Index for Convergence Sensor Networks

Jeong-Seok Park
Dept. of Medical IT Engineering, Korea National University of Transportation

융합 센서 네트워크를 위한 데이터 기반 색인

박정석
한국교통대학교 의료IT공학과

Abstract Wireless sensor networks (WSN) can be more reliable and easier to program and use with the help of sensor database management systems (SDMS). SDMS establish a user-friendly SQL-based interface to process declarative user-defined queries over sensor readings from WSN. Typical queries in SDMS are ad-hoc snapshot queries and long-running, continuous queries. In SDMSs queries are flooded to all nodes in the sensor net, and query results are sent back from nodes that have qualified results to a base station. For query flooding to all nodes, and result flooding to the base station, a lot of communication energy consuming is required. This paper suggests an efficient in-network index solution, named Distributed Information Gathering (DIG) to process range queries in a sensor net environment that can save energy by reducing query and result flooding.

• **Key Words** : Convergence sensor networks, Sensor database management systems, Query processing, Range queries, In-node indexing

요 약 무선센서 네트워크는 센서 데이터베이스 관리 시스템을 통해 보다 효율적으로 개발 및 운용될 수 있다. 센서 데이터베이스 관리 시스템은 무선센서 입력에 대해 선언된 사용자 정의 질의를 처리하기 위해 사용자들에게 익숙한 SQL 유형의 사용자 접속을 지원한다. 무선센서 네트워크상의 전형적 질의 유형은 임의의 스냅 샷 값 검색이나 오래도록 지속되는 연속 질의 형태를 갖는다. 무선센서 네트워크상에서 질의 처리는 베이스스테이션으로부터 여러 노드들로 질의를 보내는 과정과 여러 노드에서 얻어지는 질의 결과를 베이스스테이션으로 회수하는 과정이 있는데 이러한 질의의 파급이나 베이스스테이션으로의 결과 전송은 많은 에너지 소모를 요구한다. 이 논문은 무선센서 네트워크상에서 영역 질의를 처리함에 있어 질의 및 결과를 파급시키는데 소모되는 에너지를 절약시켜 주기 위한 분산정보수집(DIG: Distributed Information Gathering)이라고 이름붙인 효율적 색인 방법을 제안한다.

• **주제어** : 융합 센서 네트워크, 센서 데이터베이스 관리 시스템, 질의처리, 영역질의, 인-노드 색인

1. INTRODUCTION

Nowadays, Wireless sensor networks (WSN) are applied to various field to survey the physical world. Sensor database management systems (SDMS) model sensor motes as distributed data generators and query processing units. Users can use declarative SQL queries to deploy their observation and monitoring

---

*Corresponding Author : 박정석(jpark@ut.ac.kr)

Received October 21, 2016 Revised November 23, 2016
Accepted December 20, 2016 Published December 31, 2016

tasks [1, 2].

The objective of SDMS is to route queries to corresponding sensor motes and route qualified sensor data back to query issuers with the minimized communication cost. Due to the constraints of WSNs, the query processing efficiency of SDMS mainly depends on the communication cost. The communication cost of query processing in SDMS comprises two types of communication, query communication and update communication to route query messages to corresponding sensor motes, and query results to query issuers.

To minimize the communication cost, SDMSs favor aggregation queries[3, 4], since the aggregation query results are usually compact and easy to transmit. Different in-network update data reduction technologies, such as data compression[5], update filtering[6], update suppression[7], function approximation[8, 9], and query result sharing[10, 11], have been developed to reduce the update communication cost.

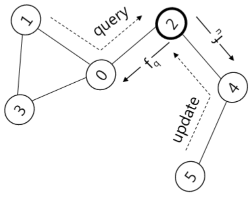
Range queries[12] target a specific range from the sensor value space. The difficulty of processing range queries is how to deliver the query messages, especially when the query rate is high. When routing queries, if the query issuer is able to know which motes detect qualified results, delivery of queries can be reduced and communication cost can be saved[13]. In traditional DBMSs, index structures are utilized to direct queries and speed-up query processing.

This paper presents an in-network index technology, called Distributed Information Gathering (DIG), to efficiently process range queries in constrained WSNs. DIG first divides the sensor value space into subranges, and designates one or more sensor motes as index nodes for each subrange. Instead of flooding every query message, DIG only needs to route query messages to index nodes. The experiment results shows that DIG can save up to 80% in the total communication cost compared to the approach of flooding queries.

2. IN-NETWORK INDEX

To process range queries, a SDMS has to build the link between local sensor readings to the incoming queries. Transmitting all raw readings to a central base and process the queries in the central base, or flooding queries to all sensor motes and process queries at local motes, can build the links. However, both approaches consume unnecessary energy. By partitioning the sensor value space into several subranges with a predefined hash function, we can designate one or more sensor motes as index nodes for each value subrange. Those index nodes work as bridges between incoming queries and updating sensor data. With the help of index nodes, we neither flood queries to all sensor motes nor send all raw readings to a central base.

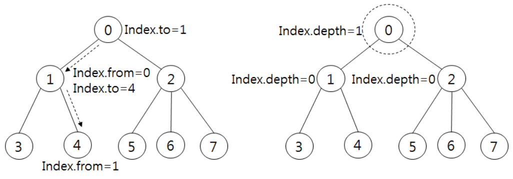
When detecting a new sensor reading, a sensor mote simply send the reading to a nearby index node responsible for the reading value. When issuing a new query, a query issuer only send the query message to the corresponding index node(s). Fig.1 illustrates an example of how to process range queries based on a in-network index structure. In Fig.1, the circles indicate the locations of sensor motes, the straight line indicate the communication link between motes and the mote 2 is an index node. The mote 5 needs to update its current reading to the node 2. The update message is routed to the index node through the mote 4. The mote 1 is interested in the value range covered by the mote 2. Instead of flooding the query message to every sensor motes, we only need to route the query message to the node 2. After receiving the query from the mote 1 and caching the update reading from the mote 5, the mote 2 send qualified sensor readings to query issuers.



[Fig. 1] An example of index node

3. DESIGN OF DIG

The detailed design of DIG depends on the particular underlying protocol. We've implemented DIG over the routing tree based protocol[14] which is generally used by current SDMSs. Fig.2(a) illustrates the index structure of DIG over the routing tree, in which the senor mote 0 is the root mote. We set a single index mote for every value subrange, and designate the root mote for all index nodes when DIG starts. Index nodes should "move" to minimize the communication cost. Statistical reports about sensor readings help index nodes to find where to 'go'. The expensive centralized collection and analysis of statistical reports counteract the benefit gained by the in-network index. An index node receives update and query messages, and returns qualified sensor data back. Thus, the distributed index nodes are able to find the optimal location for themselves. Since index nodes receives messages only through the immediate neighbor motes, a global information about the network is unnecessary. For the routing tree protocol which ensures the parent-to-children communication links, DIG should only enable the index movement within parents and children. In Fig.2(a), if the mote 1 routes most update messages to the root mote, the root mote should passed the index node to the mote 1. Additional to the update rate, the update route length is also important to designate the index nodes. If the mote 2 only update its local reading to the root, while the mote 1 routes update messages from the mote 4 and 5, the root mote should pass the index node to the mote 1 instead of the mote 2. Senor motes keep the movement by index.to and index.from locally in successive index nodes as shown by Fig.2(a). In this way, the index designation message is kept only between the successive index nodes. Index nodes are able to "move" further into the network. By tracing the index.to and index.from links, DIG can keep the trace of current index nodes for every value subrange, and direct update and query messages correctly.



(a) Move Index Node (b) Split Index Node

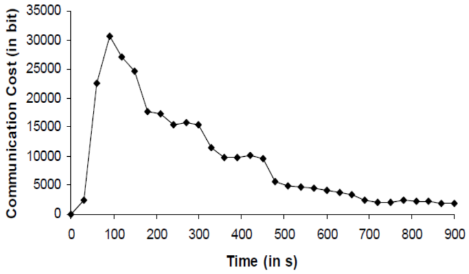
[Fig. 2] Index Structure of DIG over a Routing Tree

Statistics about sensor readings and queries are important for DIG to choose the optimal number and location of index nodes. DIG uses two fixed time periods, T1 and T2, to collect the statistical information. The first time period allows index nodes to analyze the routing cost of their immediate neighbors. At the end of T1, a index node can chose one of its immediate neighbors to be the index node for the next T1 time. The second time period, T2, allows index nodes to "move" close enough to minimize the communication cost. If only "moving" cannot optimize the communication cost, when T2 expires, the network can decide to choose more or less index nodes for a value range. To select the next index nodes, current index node should know the routing cost of their immediate neighbors. Both the message and hop counts are important statistics. When routing a message to an index node, local sensor motes increase the hop count in the message along the route path. After receiving a message from an immediate neighbor, an index node accumulates the hop count to a cost entry of the immediate neighbor. In this way, an index node keeps only a small size of statistical reports about its immediate neighbors in its local memory. Based on the statistics collected during T1 periods, index nodes can "move" close enough to the sensor motes which have more similar readings within the value subranges covered by the index nodes. In this way, DIG shortens the length of update route.

4. EXPERIMENT RESULT

We implemented DIG in TinyOS[15], and tested DIG in TOSSIM. Current DIG uses the routing tree based protocol provided by TinyOS to maintain the routing tree among sensor motes. We used a video clip of flames as the input data to test DIG. We tested DIG on different sized networks in different geographical layouts with different query patterns. For all tests, we set T1 = 30s and T2 = 150s. So we can move index nodes in every 30 seconds and do split/merge operation in every 150 seconds if necessary.

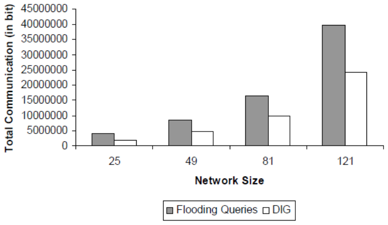
Fig.3 illustrates a result of update communication cost on the 49 grid network over the first 900 seconds since DIG starts. DIG can decrease the update communication cost in a few minutes after the update communication cost reaches the peak.



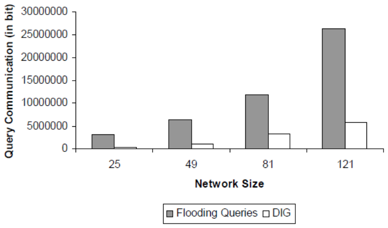
[Fig. 3] A result of update communication cost

For the grid layout, we tested DIG in different sized networks and set the root mote at the center. The network size starts from 25 motes to 121 motes, which simulates different WSN applications. For this set of experiments, we set the query rate fixed at 1 new query per 5 seconds, and compared DIG with the approach of flooding queries to every sensor mote.

Fig.4(a) explains the total communication costs of different approaches in 4 hours. DIG can save around 50% of the total communication cost, compared with the cost of flooding queries for the four different network sizes, as shown by Fig.4(a). DIG improves the communication cost for disseminating query by a factor of 5 for uniformly distributed queries. Fig.4(b) also illustrates the scalability of DIG for the query communication. For the grid layout with 49 motes, we also tested DIG with different query rates. We set a new query coming in every 5, 10 or 20 for this set of tests.

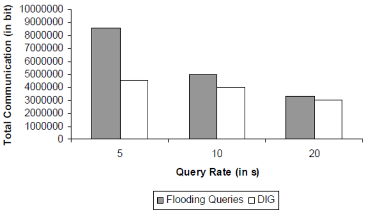


(a) Total Communication Cost

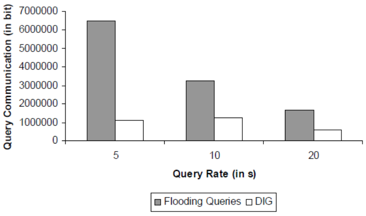


(b) Query Communication Cost

[Fig. 4] Different Network Sizes



(a) Total Communication Cost



(b) Query Communication Cost

[Fig. 5] Different Query Rates

Fig.5(a) shows the total communication cost of DIG and the approach of flooding queries for different query rates. The cost of DIG can be identical to the cost of flooding queries under some circumstances, as shown by Fig.5(a). As the query rate decreases, the query communication decreases as illustrated by Fig.5(b). DIG can significantly decreases the communication cost for query messages, when the query rate is high. Even for the query rate of 1 new query per 20 seconds, DIG can still save around 60% in disseminating queries.

5. CONCLUSION

We present a new in-network index technique, named DIG, for constrained WSNs. DIG provides an efficient query and data driven index to shorten both the sensor data update and query delivery route paths. By utilizing index nodes for different value subranges, DIG avoids the expensive query flooding used by current SDMSs. By designating index nodes at different locations and different number of index nodes for different value subranges, DIG minimizes the total cost of update and query communication. By maintaining the index structure within network, DIG avoids the expensive centralized index designation chosen by other in-network index techniques. There are many and various ongoing sensor network related researches[16, 17, 18] and DIG can be applied for in-network query efficiency for such applications.

ACKNOWLEDGMENTS

This was supported by Korea National University of Transportation in 2016.

REFERENCES

[1] P. Bonnet, J. Gehrke, and P. Seshadri, “Towards sensor database systems”, MDM, pp. 3-14, Springer-Verlag, 2001.

[2] S. R. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “Tinydb: an acquisitional query processing system for sensor networks”, ACM Trans. Database Syst., Vol. 30, No. 1, pp. 122-173, 2005.

[3] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, “Tag: a tiny aggregation service for ad-hoc sensor networks”, ACM SIGOPS, Vol. 36, Issue SI, pp. 131-146, 2002.

[4] M. Sharifzadeh and C. Shahabi, “Supporting spatial aggregation in sensor network databases," in ACM GIS, pp. 166-175, 2004.

[5] J. Hellerstein and W.Wang, “Optimization of in-network reduction”, VLDB-Workshop DMSN, pp. 166-175, 2004.

[6] A. Jain, E. Y. Chang, and Y.-F. Wang, “Adaptive stream resource management using Kalman filters”, ACM SIGMOD, pp. 11-22, 2004.

[7] A. Silberstein, R. Braynard, and J. Yang, “Constraint chaining: on energy-efficient continuous monitoring in sensor networks”, SIGMOD '06: Proceedings of the 2006 ACM SIGMOD international conference on Management of data, (New York, NY, USA), pp. 157-168, ACM Press, 2006.

[8] A. Deligiannakis, Y. Kotidis, and N. Roussopoulos, “Compressing historical information in sensor networks“, in ACM SIGMOD, pp. 527-538, 2004.

[9] C. Guestrin, P. Bodik, R. Thibaux, M. A. Paskin, and S. Madden, “Distributed regression: an efficient framework for modeling sensor network data.”, IPSN, pp. 1-10, 2004.

[10] N. Trigoni, Y. Yao, A. J. Demers, J. Gehrke, and R. Rajaraman, “Multi-query optimization for sensor networks”, DCOSS (V. K. Prasanna, S. S. Iyengar, P. G. Spirakis, and M. Welsh, eds.), vol. 3560 of Lecture Notes in Computer Science, pp. 307-321, Springer, 2005.

[11] N. Trigoni, Y. Yao, A. J. Demers, J. Gehrke, and R. Rajaraman, “Multi-query optimization for sensor networks”, In TR2005-1989, Cornell University, 2005.

[12] X. Li, Y.-J. Kim, R. Govindan, and W. Hong, “Multi-dimensional range queries in sensor

networks", in Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, SenSys 2003, Los Angeles, California, USA, November 5-7, 2003, pp. 63-75, 2003.

[13] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion for wireless sensor networks", IEEE/ACM Transactions on Networking, Vol. 11, 2003.

[14] A. Woo, T. Tong, and D. E. Culler, "Taming the underlying challenges of reliable multihop routing in sensor networks", in Proceedings of the 1st International Conference on Embedded Networked Sensor Systems, SenSys 2003, Los Angeles, California, USA, November 5-7, 2003, pp. 14-27, 2003.

[15] TinyOS, "http://www.tinyos.net"

[16] Hae-Won Choi, Hyunsung Kim, "Impersonation Attacks on Anonymous User Authentication and Key Agreement Scheme in Wireless Sensor Networks", Journal of Digital Convergence, Vol. 14, No. 10, pp. 287-293, Oct. 2016.

[17] Bong-Hi Shin, Hye-Kyoung Jeon, Kyung-Yong Chung, "An Energy Efficient Clustering Method Based on ANTCLUST in Sensor Network", Digital Convergence, Vol. 10, No. 1, pp. 371-378, Feb. 2012.

[18] Yong-Tae Kim, Yoon-Su Jeong, "Optimization Routing Protocol based on the Location, and Distance information of Sensor Modes", Journal of Digital Convergence, Vol. 13, No. 2, pp. 127-133, Feb. 2015.

저자소개

박 정 석(Jeong-Seok Park) [정회원]

- Ph. D., Computer Science, Feb. 2000, Chungbuk National University, Korea
- M.S., Computer Science, Aug. 1983, Soongsil University, Korea
- B.S., Computer Science, Feb. 1981, Soongsil University, Korea
- 1996/8~present: Professor at Korea National Univ. of Transportation
- 2015/1~2015/12: Visiting Scholar at Univ. of Hartford, CT, U.S.A.
- 2006/8～2007/8: Visiting Scholar at Univ. of Maine, Orono, ME, U.S.A.
- 1983/1～1996/8: Senior Researcher at Korea Atomic Energy Research Inst. in South Korea

<Research Interest> : Query processing on wireless sensor network, Bio-sensor data processing