
센서 네트워크를 이용한 질의 배분 기법

김동현

동서대학교 컴퓨터정보공학부

The Scheme for Distributing the Query Constraints using the Sensor Networks

Dong Hyun Kim

Division of Computer Information Engineering, DongSeo University

E-mail : pusrover@dongseo.ac.kr

요 약

센서 노드에서 수집되는 데이터는 지속적으로 삽입되는 스트림 데이터이기 때문에 효율적인 사용자 질의 처리를 위하여 노드별로 질의 색인을 구축해야 한다. 노드에서 최소 크기의 질의 색인을 구축하기 위해서는 질의 색인에 삽입되는 질의 조건을 수를 줄여야 할 필요가 있다. 이 논문에서는 삽입되는 질의 조건의 수를 줄이기 위하여 다차원 데이터 색인을 이용한 질의 조건 배분 기법에 대하여 제안한다.

ABSTRACT

Since the data collected at a sensor node is the stream data, for processing efficiently user queries, the query index should be constructed at each node. To construct the minimized query index at the node, it is required to reduce the number of query constraints inserted into the query index. In this paper, we propose the scheme of the query constraints distribution using the multi-dimensional data index in order to diminish the number of the inserted query constraints.

키워드

query index, query constraints, sensor network, query processing

1. 서 론

센서네트워크는 유비쿼터스 환경을 구축하기 위한 핵심 기술로서 다양한 위치에 설치된 센서 노드를 통하여 사람/사물 및 환경 정보를 인식하고 인식된 정보를 통합, 가공하여 사용자에게 제공하는 정보 서비스 인프라이다. 특히 인간이 개입하지 않고 현장의 데이터를 신속하게 수집, 가공하여 사용자에게 유용한 정보를 제공할 수 있기 때문에 물류, 교통, 농축산, 정보가전 및 환경 등의 다양한 산업 분야에서 활용 가능한 차세대 IT 핵심 기술이다.

기존의 데이터 검색 기술은 디스크 또는 메모리에 고정적으로 저장된 데이터에 대하여 검색을 신속히 수행하는 기술들이다. 이를 위하여 그림 1의 (가)와 같이 고정적인 데이터 집합에 대하여 데이터 및 질의 특성에 적합한 색인을 구성하고

사용자 질의가 삽입되었을 때 색인을 이용하여 고정 데이터를 효과적으로 검색하여 질의 결과를 생성하는 것이 주요 목표이다. 그러나 센싱 데이터와 같은 스트리밍 데이터는 지속적으로 데이터 집합이 삽입되는 특성을 가진다. 따라서 기존의 데이터 색인 기법을 사용하면 색인 변경 비용이 매우 커지고 검색 비용이 증가하는 단점이 있다. 지속적으로 삽입되는 데이터 집합에 대한 사용자 질의를 수행하기 위해서는 기존의 방법과는 달리 그림 1의 (나)와 같이 다수의 사용자 질의를 색인하고 지속적으로 데이터가 삽입되는 스트리밍 데이터를 질의 색인과 비교하여 삽입된 데이터가 결과 값이 될 수 있는 질의를 효율적으로 검색하는 기법을 사용할 필요가 있다.

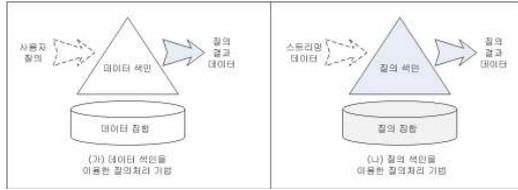


그림 1 데이터 색인 및 질의 색인을 이용한 질의 처리

센서 네트워크에서 질의 처리를 수행하기 위한 기법은 크게 2가지로 분류될 수 있다. 첫 번째는 센서 네트워크의 외부에 위치한 서버에서 사용자 질의를 처리하는 기법[1][2][3]이다. 그러나 이 기법들은 센서 네트워크에서 모든 데이터를 전송해야 하는 단점이 있다. 두 번째는 네트워크 내부에서 사용자 질의를 처리하는 기법[4][5][6][7][8]이다. 이 기법들은 수집된 데이터를 자신의 노드에서 저장하고 있다가 사용자 질의를 처리하거나 [4][5] 또는 특정 노드에 일정한 범위의 데이터를 전송하여 저장한 후에 사용자 질의를 처리하는 기법들[6][7][8]이다. 그러나 이러한 기법들은 고정적으로 저장된 데이터에 대하여 질의 처리를 수행하기 때문에 스냅샷 질의를 처리할 수 있지만 연속 질의를 처리하기 어려운 문제가 있다.

센서 네트워크 환경에서 센서 스트림 데이터에 대한 사용자 질의를 효율적으로 처리하기 위해서는 먼저 효율적인 크기의 질의 색인을 구축해야 한다. 이 논문에서는 적정 노드로 질의 조건을 전송하여 효율적인 크기의 질의 색인을 구축하기 위하여 데이터 영역을 이용한 질의 조건 배분 기법에 대하여 제안한다. 그리고 데이터 영역을 조절하기 위한 영역 조절 기법에 대하여 제시한다.

논문의 구성은 다음과 같다. 2장에서는 관련 연구를 기술하고 3장에서는 질의 조건 배분 기법에 대하여 기술한다. 마지막으로 4장에서는 결론을 제시한다.

II. 관련 연구

센서 네트워크에서 사용자 질의를 처리하기 위한 기법은 크게 두 가지로 분류할 수 있다. 첫 번째는 센서 네트워크의 외부에 위치한 서버에서 질의를 처리하는 네트워크 외부 처리 기법이다. COUGAR 시스템[1]은 센서 데이터베이스의 모델을 정의하고 일정 기간동안 유지되는 장기간 질의(Long running queries) 형태로 센서 데이터베이스에 대한 질의를 표현하였고 코넬 대학의 COUGAR 시스템에 적용하여 구현하였다. GARUDA 기법[2]은 싱크 노드로부터 센서 노드로의 포인트-투-멀티포인트(point-to-multipoint) 데이터 전달 기법을 제안하였으며 소규모 네트워크부터 클러스터화된 대규모 네트워크까지의 다

양한 크기의 네트워크에 대하여 안정된 성능을 유지하는 것을 보여주었다. TiNA 기법[3]은 사용자가 정의한 데이터 품질을 유지하면서 센서 네트워크의 에너지 소비를 감소하기 위한 데이터 합병 기법을 제안하였으며 이전 전송 데이터와 일정 수준 이상으로 차이나는 경우에만 데이터를 전송하는 기법을 제안하였다. 그러나 이러한 기법들은 센서 네트워크의 외부에서 질의를 처리하기 때문에 질의 처리를 위하여 모든 센서 노드에서 수집되는 모든 데이터를 서버까지 전송해야 하는 문제가 있다.

두 번째는 센서 네트워크의 내부 노드에서 사용자 질의를 분산하여 처리하는 방법이다. 이 방법은 센서 데이터의 저장 위치에 따라 두 가지로 분류될 수 있다. 첫 번째는 수집된 데이터를 수집한 센서 노드 자체에서 저장하는 기법이다. TAG 기법[4]은 TinyOS를 기반으로 하는 애드혹(ad hoc) 네트워크에서의 병합(agggregation) 기법을 제안하였으며 네트워크를 구성하는 모든 센서를 계층 구조의 트리로 구성한 라우팅 기법을 제시하였다. HEED 기법[5]은 에너지 효율성을 증대하기 위하여 클러스터 헤드 기반의 라우팅 기법을 제안하였으며 클러스터 헤드는 클러스터를 구성하는 노드들간의 데이터 병합 작업에 대한 조정과 클러스터 기준으로 외부 노드들과의 데이터 전송 책임을 가지도록 구성하였다.

두 번째는 데이터 값의 범위에 따라 특정 노드로 전송하여 저장하고 사용자 질의를 해당 노드로 전송하여 처리하는 기법이다. GHT 기법[6]은 모든 데이터를 센서 네트워크의 관련된 노드에 저장하는 데이터중심저장 기법을 제안하였으며 공간 해쉬 테이블(Geographic Hash Table, GHT)를 이용하여 데이터가 저장될 노드를 선정하고 질의 처리 시 이를 사용하는 알고리즘을 제시하였다. DIM 기법[7]은 이벤트가 발생하는 영역을 0 또는 하나의 센서 노드를 포함할 수 있도록 부분 영역으로 분할하고 발생하는 이벤트와 센서 노드를 KD 트리를 이용하여 연결하는 기법을 제안하였으며 각 센서노드는 고유의 비트화된 주소를 가지고 이벤트 발생 시 KD 트리의 분할 영역에 따라 이벤트의 데이터가 저장될 노드가 선정된다. 또한 GPSR(Greedy Perimeter Stateless Routing)기법을 이용한 데이터 라우팅을 사용하였다. KDDCS 기법[8]은 이벤트가 발생하는 데이터 영역의 차원에 따라 영역을 분할하고 센서 노드의 물리적인 주소와 KD 트리를 이용하여 연결하는 기법을 제안하였으며 영역 분할 시 데이터 영역을 균등 분할하지 않고 균등한 수의 센서 노드를 포함하도록 분할함. 따라서 KD 트리 구조가 편중되지 않고 트리의 레벨을 균등하게 유지하도록 설계하였다. 이러한 구조에서는 이벤트가 저장될 센서 노드를 검색하기 위한 KD 트리 검색 시 이벤트의 데이터 영역이 불균등하게 분할되는 특징을 가진다. 그리고 LSR(Logical Stateless Routing) 기법을 사용하여 노드로 이벤트 데이터

를 라우팅하였다.

III. 임계값 기반 질의 배분 기법

노드별로 질의 색인을 구축하고 사용자 질의에 해당되는 센서 값을 전송하는 방법은 크게 두 단계로 구성된다[9]. 첫 번째는 사용자 질의의 조건을 노드로 전송하여 배분하는 단계이고 두 번째는 전송된 질의 조건을 이용하여 각 노드에서 질의 색인을 구축하고 센서 데이터 스트림에 대하여 결과 값을 서버로 전송하는 단계이다.

질의 조건을 노드로 배분하는 기법은 크게 두 가지 기법이 있다. 첫 번째는 모든 노드에 동일하게 질의 조건을 배분하는 LND(all Node Distribution) 기법이다. 이 기법은 매우 단순하며 질의를 배분하기 위해 추가적인 구조나 연산이 필요없기 때문에 적은 비용으로 질의 조건을 배분할 수 있다. 그러나 질의 결과 값이 측정되지 않을 노드에도 질의 조건을 전송하기 때문에 불필요한 메시지 전송이 발생하며 특히 질의 색인의 크기가 늘어나는 단점이 있다.

두 번째는 사용자 질의를 해당 결과 값의 발생 가능성이 높은 노드에게 배분하는 AND(Appropriate Node Distribution) 기법이다. 이 기법은 노드에서 발생 가능한 데이터에 대한 질의 조건들만 유지하기 때문에 메모리 사용 및 질의 처리에 효율적이지만 질의 배분을 위한 추가적인 데이터 구조와 배분 알고리즘이 필요한 단점이 있다. 센서 노드에 장착되어 있는 메모리의 크기는 몇 백 kb의 크기로 매우 작기 때문에 질의 색인의 크기를 최소의 크기로 구성해야 할 필요가 있다. 이를 위해서는 먼저 질의 색인에 색인되는 조건의 수를 최소화해야 하므로 추가적인 비용이 소요되더라도 두 번째 기법이 효율적이다.

사용자 질의 조건을 발생 가능성이 높은 노드에 배분하기 위하여 서버에서 다차원 데이터 색인을 구성하여 유지한다. 다차원 색인 구조는 그림 2와 그림 3과 같이 KD 트리를 기반하는 계층 구조를 활용하고 노드에서 발생하는 데이터와 노드의 데이터 영역과의 적합율(hit ratio)에 따라 데이터 영역의 범위를 재조정한다.

노드의 데이터 영역을 조절하기 위하여 초기에는 각 노드별로 일정한 크기의 데이터 영역을 강제적으로 설정한다. 그리고 센서 데이터가 수집될 때 마다 노드에 설정된 데이터 영역과 비교하여 영역에 포함되는 데이터인지 검토하고 이에 따른 적합율을 계산한다. 만약 적합율이 일정 임계값이하인 경우에는 해당 노드에서 수집된 노드 센서 데이터의 범위에 맞추어 노드의 데이터 영역을 확대하거나 혹은 축소한다. 만약 적합율이 일정 임계값이상인 경우 노드의 데이터 영역을 설정한다.

만약 사용자로부터 질의가 시작되면 데이터 색인에 따라 질의 조건이 삽입될 노드를 선택한다. 그리고 네트워크 라우팅 알고리즘을 사용하여 해

당 노드로 질의 조건을 전송한다.

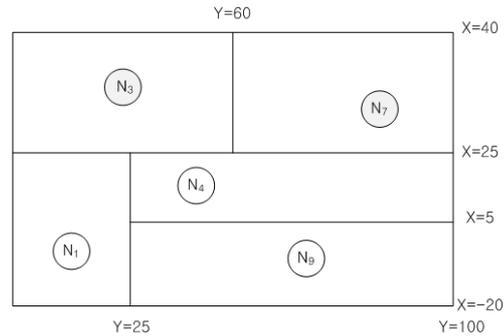


그림 2 노드 데이터 영역의 예

예를 들어 그림 2와 같이 노드 데이터 영역이 설정되고 그림 3과 같이 데이터 색인이 구성되었다고 가정해보자. 만약 질의 조건이 “ $-5 < x < 2$ & $10 < y < 20$ ”인 질의가 삽입되면 그림 2의 데이터 영역과 비교해 볼 때 N_4 가 해당되기 때문에 그림 3과 같이 해당 질의 조건을 N_4 로 전송하여 배분한다.

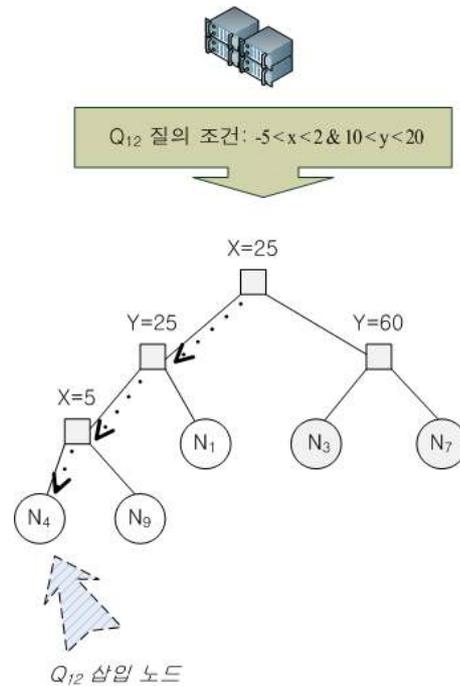


그림 3 질의 조건 배분 예

IV. 결론

센서 노드에서 수집되는 센서 데이터들은 지속적으로 삽입되는 스트림 데이터이다. 따라서 사용자 질의를 효율적으로 처리하기 위해서는 연속 질의를 처리하기 위한 질의 색인을 각 노드별로

구축할 필요가 있다.

노드별로 질의 색인을 구축하기 위해서는 먼저 질의 조건을 노드로 배분해야 한다. 이 논문에서는 저용량의 센서 저장 장치에서 최소 크기의 질의 색인을 구성하기 위하여 데이터 영역에 따른 질의 배분 기법을 제안하였다. 그리고 수집된 센서 데이터의 변화에 따라 데이터 영역의 크기를 조절하기 위한 기법을 제시하였다. 추가 연구로는 적절한 크기의 데이터 영역 설정을 위한 임계값에 대한 실험이 필요하다.

참고문헌

- [1] Philippe Bonnet, Johannes Gehrke, Praveen Seshadri, "Towards Sensor Database Systems", MDM, 2001
- [2] Seung-Jong Park, Ramanuja Vedantham, Raghupathy Sivakumar, Ian F. Akyildiz, "A Scalable Approach for Reliable Downstream Data Delivery in Wireless Sensor Networks", MobiHoc, 2004
- [3] Mohamed A. Sharaf, Jonathan Beaver, Alexandros Labrinidis, Panos K. Chrysanthis, "TiNA: A Scheme for Temporal Coherency-Aware in-Network Aggregation", MobiDE, 2003
- [4] Samuel Madden, Michael J. Franklin, Joseph Hellerstein, Wei Hong, "TAG: a Tiny AGgregation Service for Ad-Hoc Sensor Networks", OSDI, 2002
- [5] Ossama Younis, Sonia Fahmy, "HEED: A Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad-hoc Sensor Networks", IEEE Mobile Computing, 2004
- [6] Sylvia Ratnasamy, Brad Karp, Scott Shenker, Deborah Estrin, Ramesh Govindan, Li Yin, Fang Yu, "Data-Centric Storage in Sensornets with GHT, A Geographic Hash Table", MONET, 2003
- [7] Xin Li, Young Jin Kim, Ramesh Govindan, Wei Hong, "Multi-dimensional Range Queries in Sensor Networks", ACM Conf. Embedded Networked Sensor Systems, 2003
- [8] Mohamed Aly, Kirk Pruhs, Panos K. Chrysanthis, "KDDCS: A Load-Balanced In-Network Data-Centric Storage Scheme for Sensor Networks", ACM CIKM, 2006
- [9] 김동현, "센서 네트워크 환경에서의 질의 색인 기법", 한국해양정보통신학회 춘계학술대회 논문집, 14권, 1호, 2010