

센서 네트워크를 위한 데이터 인지 우선순위 기반의 에너지 효율적인 Top-k 질의 처리

(Data-Aware Priority-Based Energy Efficient Top-k Query Processing in Sensor Networks)

여명호[†] 성동욱[†] 유재수^{**}
(Myung Ho Yeo) (Dong Ook Seong) (Jae Soo Yoo)

요약 Top-k 질의는 많은 센서 네트워크 응용 분야에서 중요하게 사용된다. 기존에 제안된 필터 기반의 Top-k 질의 처리 기법은 질의 결과에 영향을 미치는 값만을 수신하기 위해 각 센서 노드의 측정 허용 범위를 필터로 설정함으로써 센서 데이터의 전송 횟수를 줄인다. 하지만, 센서 데이터의 재검증과 필터의 재배포 과정에서 추가적인 에너지를 소모한다. 또한, 질의 결과값이 빈번히 변경될수록 필터의 재배포 횟수가 증가하여 에너지의 소모가 더욱 증가한다. 본 논문에서는 Top-k 질의를 처리하기 위해서 측정된 데이터를 기반으로 우선 순위를 결정하고, 우선 순위에 따라 최소한의 센서 데이터를 수집하는 새로운 Top-k 질의 처리 기법을 제안한다. 제안하는 기법의 우수성을 보이기 위해서 시뮬레이션을 통한 성능 평가를 수행하였으며, 그 결과 기존 필터 기반의 기법에 비해 네트워크의 수명이 크게 향상되었다.

키워드 : 센서 네트워크, Top-k 질의 처리, 에너지 효율, 모니터링

Abstract Top-k queries are important to many wireless sensor applications. Conventional Top-k query processing algorithms install a filter at each sensor node and suppress unnecessary sensor updates. However, they have some drawbacks that the sensor nodes consume energy extremely to probe sensor reading or update filters. Especially, it becomes worse, when the variation ratio of top-k result is higher. In this paper, we propose a novel Top-k query processing algorithm for energy-efficiency. First, each sensor determines its priority as the order of data gathering. Next, sensor nodes that have higher priority transmit their sensor readings to the base station until gathering k sensor readings. In order to show the superiority of our query processing algorithm, we simulate the performance with the existing query processing algorithms. As a result, our experimental results show that the network lifetime of our method is prolonged largely over the existing method.

Key words : Sensor network, Top-k query, Energy-Efficiency, Monitoring

· 본 연구는 교육과학기술부와 한국산업기술재단의 지역혁신인력양성사업의 지원과 정부(교육과학기술부)의 개원으로 한국과학재단의 연구지원을 받아 수행된 연구임 (No. R01-2006-000-1080900)

† 학생회원 : 충북대학교 정보통신공학과
mhyeo@netdb.chungbuk.ac.kr
sergei@netdb.chungbuk.ac.kr

** 종신회원 : 충북대학교 전기전자컴퓨터공학부 교수
yjs@chungbuk.ac.kr
(Corresponding author)

논문접수 : 2008년 5월 23일

심사완료 : 2009년 3월 13일

Copyright©2009 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 데이터베이스 제36권 제3호(2009.6)

1. 서론

무선 센서 네트워크 기술의 발전은 환경 감시, 스마트 빌딩, 군사 응용 등과 같은 여러 응용 분야로부터 다양한 데이터를 수집하는 기회를 제공한다[1,2]. 이때, 일반적으로 센서 노드는 배터리에 의해 동작하기 때문에 에너지 사용에 많은 제약은 받는다. 또한, 센서 노드들이 배포되는 환경이 매우 가변적이므로 센서 노드를 일일이 재충전하거나 교체를 하는 것 역시 쉽지 않다. 이러한 문제점을 해결하기 위해서 제한된 에너지를 효율적으로 사용하기 위한 기법에 대한 연구가 활발히 진행되고 있다.

센서 네트워크의 많은 응용 분야들은 병합된 형태의 센서 데이터를 수집하고, 모니터링 하는 것을 목표로 한다[3-9]. 여러 질의 형식 중 특히 Top-k 질의는 여러 모니터링 분야에 유용하게 사용된다. Top-k 질의는 센서 데이터 중 가장 높거나 가장 낮은 k 개의 값을 요청하는 질의이다. 센서 네트워크에서 연속적인 Top-k 질의를 처리한다고 할 때, 가장 직관적이고, 단순한 전략은 모든 센서 데이터를 기지국으로 전송하고, 기지국으로부터 k 개의 노드를 탐색하는 방법이다. 하지만, 배터리에 의해 동작하는 센서 노드들은 주기적으로 데이터를 전송하기 위해서 많은 에너지를 소모하며, 이는 네트워크 수명을 단축시킨다. 이러한 문제점을 해결하고 데이터 수집시 네트워크의 트래픽을 줄이기 위해서 TAG [10]와 같은 인-네트워크 병합(In-Network Aggregation) 기법이 제안되었다. TAG는 라우팅 트리를 만들고, 라우팅 트리를 기반으로 각 센서 노드는 질의 결과에 해당하는 결과값만을 상위 노드로 전송한다. 상위 노드는 다시 하위 노드의 결과값과 자신의 값을 비교하여 질의에 해당하는 결과값을 상위 노드로 전송한다. 이러한 일련의 과정을 통해 모든 센서 노드로부터 데이터를 수집하지 않고, Top-k 질의에 해당하는 결과값만 수집한다. 하지만, TAG를 이용하여 연속적인 질의를 처리하는 경우, 모든 센서 노드가 데이터를 전송하는 단점이 있다.

좀 더 향상된 질의 처리 방법으로 FILA[11]와 같은 필터 기반의 기법이 있다. FILA 역시 TAG 기법과 같이 라우팅 트리를 만들고, 라우팅 트리를 기반으로 센서 노드로부터 데이터를 수집한다. 그 다음, 각 센서 노드가 수집한 센서 데이터의 최대/최소값으로 필터를 설정하고, 각 센서 노드는 자신의 필터의 범위를 벗어나는 센서 데이터만 기지국으로 전송한다. 그 결과 메시지 전송량이 감소하고, 네트워크의 수명이 늘어난다. 하지만, Top-k 질의 결과의 정확성을 보장하기 위한 재검증 과정과 필터의 갱신을 위한 필터값의 재전송은 네트워크의 수명을 감소시키는 요인이 된다.

본 논문에서는 재검증과 필터 전송에 따른 에너지 소모를 제거하기 위한 새로운 접근으로 데이터 수집 우선순위를 이용한 연속적인 Top-k 질의 처리 기법을 제안한다. 제안하는 기법의 주된 아이디어는 Top-k에 해당하는 센서 노드만 데이터 수집에 참여시키자는 것이다. Top-k 질의의 목적은 최고 혹은 최저 값을 수집한 k 개의 센서 노드를 탐색하는 것이다. 따라서, 측정된 데이터의 크기에 따라 수집의 우선순위를 부여하고, k 개의 노드로부터 데이터를 수집한다. 이외 Top-k에 해당하지 않는 센서 노드는 데이터 수집에 참여하지 않고 수면 상태(Sleep Mode)를 유지하기 때문에 네트워크의

수명(Network Lifetime)이 연장된다.

본 논문의 구성은 다음과 같다. 제2장에서는 Top-k 질의 처리를 위한 기존 기법들의 특징을 기술하고, 기존 기법의 문제점을 분석한다. 제3장에서는 제안하는 질의 처리 기법의 특징과 질의 처리 과정을 기술한다. 제4장에서는 성능평가와 분석을 통해 제안하는 기법의 우수성을 보이고, 제5장에서는 결론과 향후 연구 방향을 제시한다.

2. 관련연구

TAG는 데이터 수집시 네트워크의 트래픽을 줄이기 위해서 대표적인 인-네트워크 병합(In-Network Aggregation) 기법이다. TAG는 센서 네트워크를 라우팅 트리로 구성하고, 라우팅 트리를 통해 데이터를 병합하고, 기지국으로 전송함으로써 네트워크 트래픽을 감소시킨다. 예를 들면, 임의의 센서 노드들이 그림 1과 같은 라우팅 트리를 구성한다고 가정할 때, 단말노드 B, C로부터 A로 전달된 값은 루트 노드 A에 의해 Top-1 질의의 결과에 해당하는 데이터만 기지국으로 전송된다. 결국 모든 센서 데이터가 아닌 질의의 결과값만 기지국으로 전송하기 때문에 전체적인 네트워크 수명이 연장된다. 하지만, 데이터 수집 과정에서 B의 값은 Top-1 질의의 결과값이 아니기 때문에 다른 노드의 관점에서 B의 데이터 전송은 불필요한 에너지 낭비를 초래한다.

TAG 기법은 질의 결과에 영향을 미치지 않는 불필요한 데이터의 전송을 제거하기 위해 트리 기반 라우팅 기

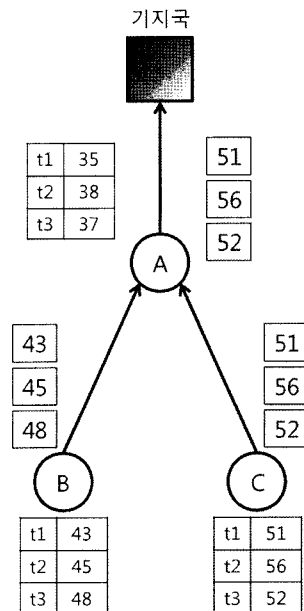


그림 1 TAG를 이용한 Top-k 질의 처리

법의 특성을 이용한다. 라우팅 경로상에서 부모에 해당하는 노드는 자식 노드로부터 수신한 데이터와 자신이 수집한 데이터를 질의의 목적에 따라 병합하거나 선별하게 된다. 해당 데이터는 다시 자신의 상위 노드로 전송되고 상위 노드는 이 일련의 과정을 반복한다. 최종적으로 최상위 노드인 기지국에서 질의 결과에 해당하는 데이터를 수신한다. TAG 기법을 통해 Top-k 질의를 처리하기 위해서는 모든 센서 데이터를 적어도 한번 이상 수집해야 한다. 즉, Top-k 질의 결과에 포함되지 않는 데이터라고 할지라도 기지국까지 전송해야 한다. 예를 들어 그림 2와 같은 라우팅 트리에서 Top-1 질의를 처리한다고 할 때, 단말 노드로부터 데이터 전송을 시작하기 때문에 일부 불필요한 데이터 전송이 발생한다. 즉, 루트 노드 A에서 측정한 데이터 "55"가 Top-1 질의의 결과로 선택되지만 단말노드로부터 센서 데이터를 수집하기 때문에 {B, C, D, E, F, G} 노드의 데이터 전송은 질의 결과와 상관없는 불필요한 데이터 전송이 된다.

FILA는 이러한 문제점을 해결하기 위해서 각 센서 노드에 데이터의 수집 범위를 나타내는 필터를 내장하고, 필터의 범위를 벗어나는 데이터가 발생할 때만 데이터를 전송하여 불필요한 데이터 전송을 제거한다. 그림 3과 같이 A, B, C 노드에 각각 (20, 39), (39, 47), (47, 80)의 필터가 지정되었다고 가정할 때, 각 센서 노드들은 필터의 범위를 벗어나는 값을 측정할 경우만 데이터를 전송한다. 데이터가 전송되지 않으면, 현재의 데이터 수집 범위를 유지하며 Top-k 질의의 결과값은 변하지 않았다는 것을 의미한다. t3 시점에서 B 노드가 48을

측정하였기 때문에 자신의 데이터를 A 노드로 전송한다. A 노드는 해당 값이 Top-k 질의의 결과값인지 판단하기 위해서 기지국으로 전송한다. 하지만, 해당 값의 범위는 데이터를 전송하지 않은 C 노드의 필터 범위에 포함되기 때문에 정확성을 보장하기 위해서 C 노드의 측정 값을 검증해야 한다. 이 과정에서 B 노드의 값은 C 노드로 다시 전송되고, C 노드의 센서 데이터와 비교하여 검증 과정을 수행한다. C 노드는 B 노드의 값보다 더 큰 값을 측정했으므로 센서 데이터 52를 A를 통해 기지국으로 전송하고 Top-k 질의의 결과값을 결정한다. 그 다음, 기지국에서 B, C 노드의 필터를 갱신하기 위한 메시지를 송신한다.

FILA 기법은 연속적인 Top-k 질의를 효율적으로 처리하기 위해 필터 기반의 기법을 이용한다. 각 노드마다 과거 수집한 데이터를 기반으로 필터를 설정하고, 필터 범위를 벗어나는 센서 데이터만 전송하여 불필요한 데이터 전송을 방지한다. 하지만 FILA에서 이용하는 필터 기반 기법의 특성상 Top-k에 해당하는 센서 데이터라 할지라도 필터 범위에 속하는 경우 기지국으로 전송하지 않는다. 즉, Top-k에 해당하는 노드는 알 수 있지만, 실제 센서 데이터는 알지 못한다. 그리고 특정 노드에서 자신의 필터 범위를 벗어나는 데이터를 수집한 경우 해당 데이터가 질의 결과에 영향을 미치는지 검증하는 단계와 필터를 갱신하는 단계가 필요하다. 이 단계에서 해당 데이터를 기지국으로 전송해야 하며, 해당 데이터를 수집한 노드의 필터를 갱신함으로써 필터의 범위에 영향을 받는 다른 노드들의 필터 또한 갱신해야 한다. 예를

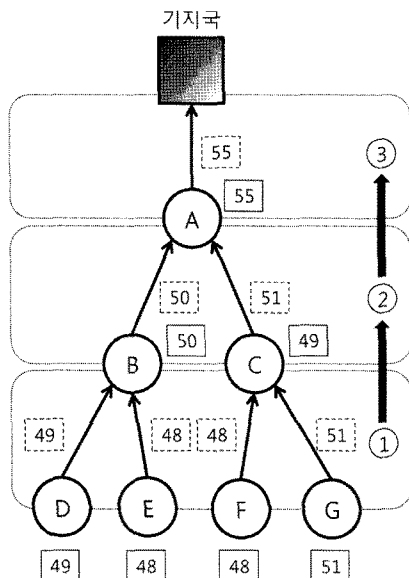


그림 2 TAG 기법을 이용한 Top-k 질의 처리

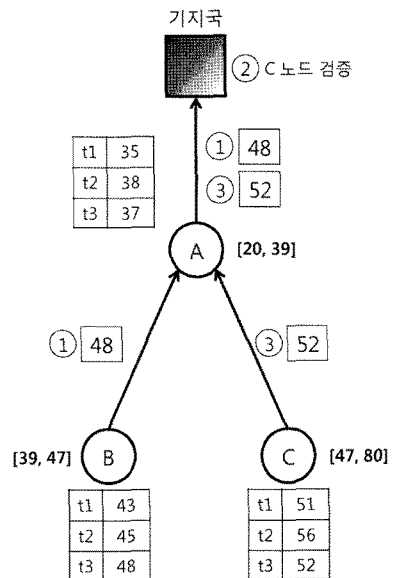


그림 3 FILA를 이용한 Top-k 질의 처리의 예 (k=3)

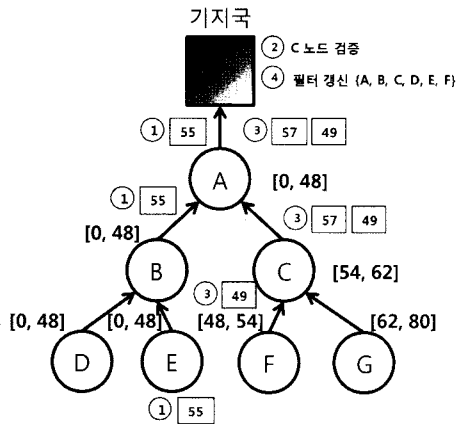


그림 4 FILA 기법을 이용한 Top-3 질의 처리

들어 그림 4와 같이 연속적인 Top-3의 질의가 내려진 상황에서 E 노드가 55값을 수집했을 경우 Top-3 노드가 변경된다. 이때 E 노드의 데이터는 기지국으로 전송되고 기지국에서는 결과 재검증을 위해 Top-3 노드였던 {C, F, G} 노드로 탐색 메시지를 전송하여 최신 데이터를 수집한다. 이후 Top-3 결과를 갱신하고 {A, B, C, D, E, F} 노드의 변경된 필터를 갱신한다. 여기서 해당 노드가 라우팅 트리에서 하위에 위치할수록 갱신비용은 기하급수적으로 증가하는 문제점을 가지고 있다.

3. 제안하는 연속적인 Top-k 질의 처리 기법

본 장에서는 재검증 과정과 필터 제조정에 따른 에너지 소모를 제거하기 위한 새로운 접근을 제안한다. 제안하는 질의 처리 기법은 수집한 데이터를 기반으로 각 센서 노드에 수집 우선 순위를 부여하고, 이 우선 순위에 따라 Top-k를 만족하는 최소 개의 센서 데이터를 수집함으로써 불필요한 데이터 전송을 줄인다. 먼저 데이터 수집을 위한 우선 순위 부여 기법과 단일 Top-k 질의 처리 기법을 설명하고, 그 다음 연속적인 질의 처리를 위한 확장을 기술한다.

3.1 데이터 인지 우선 순위 부여

TAG 기반의 데이터 수집은 단말 노드로부터 상위노드로 순차적으로 데이터를 수집한다. 정적인 수집 순서에 따라 Top-k 질의와 같은 일부 센서 노드의 결과값을 요구하는 질의를 처리하는 경우, 불필요한 데이터 전송이 발생하게 된다. 본 논문에서는 측정값의 크기에 따라 수집 순서를 동적으로 부여하기 위해서 “데이터 인지 우선 순위(Data-Aware Priority)” 부여 기법을 제안한다. 먼저 센서 네트워크의 데이터 수집을 위한 TDMA(Time Division Multiple Access) 스케줄을 그림 5와 같이 N 개의 부분 TDMA 프레임으로 나눈다.

그 다음, 각 TDMA 프레임을 실제 노드가 데이터를 전송하는 시간 슬롯 단위로 나눈다. Top-k 질의 처리 기법을 수행하기 이전에 모든 센서 노드는 식 (1)을 이용하여 중앙의 어떠한 제어 없이 스스로 TDMA 프레임을 결정한다.

$$\{TDMA\ \text{프레임의\ 번호}\} = \left\lfloor \frac{V_{\max} - V_{\text{node}}}{V_{\max} - V_{\min}} \times C_{\text{slot}} \right\rfloor \quad (1)$$

이때, V_{\min} 과 V_{\max} 는 각각 센서가 측정할 수 있는 최소, 최대 값을 나타내고, V_{node} 는 현재 측정값을 나타낸다. C_{slot} 은 미리 정의된 TDMA 슬롯의 수를 나타낸다. 결국, 센서 스스로 측정값에 따라 TDMA 순서를 결정함으로써 큰 값을 측정된 센서 노드가 일찍 데이터 수집에 참여한다. 이때, 데이터 충돌과 멀티홉 전송을 고려하여 TDMA 프레임은 모든 센서 노드가 기지국까지 데이터를 전송할 수 있는 충분한 길이를 가진다.

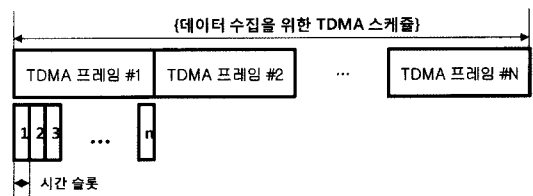


그림 5 데이터 인지 우선 순위 부여 기법을 위한 TDMA 스케줄 구조

3.2 우선 순위 기반의 단일 Top-k 질의 처리

3.1절에서 제안한 데이터 인지 우선 순위 부여 기법을 통해 각 센서 노드는 큰 값을 가진 순서대로 데이터를 전송한다. Top-k 질의의 목적은 가장 높은 값을 가진 k 개의 센서 노드를 찾는 것이다. 따라서, 큰 값을 가진 센서 노드부터 내림차순으로 데이터를 수집하고, 수집을 멈추면 된다. 이를 보장하기 위해서 모든 센서 노드가 데이터 전송 횟수를 기록하는 카운터 변수를 가지므로써 Top-k 질의 처리를 쉽게 구현할 수 있다. 하지만, 4.1절에서 제안된 TDMA 스케줄 구조를 그림 6과 같이 다중 홉으로 데이터를 수집하는 센서 네트워크에 그대로 사용하는 경우 잘못된 값을 Top-k 질의의 결과로 선정하는 문제가 발생한다. 그림 6과 같은 라우팅 트리가 구성되고, B, G 센서 노드는 같은 프레임에 할당되었다고 가정하자. TAG 기반의 데이터 병합은 다중홉(Multi-hop) 방식으로 수행되기 때문에 센서 노드마다 기지국까지 데이터를 전송하는 시간이 다르다. 따라서, G 센서 노드가 더 높은 값을 측정하여 더 빠른 시간 슬롯을 할당 하더라도 기지국에 가까운 B 센서 노드의 데

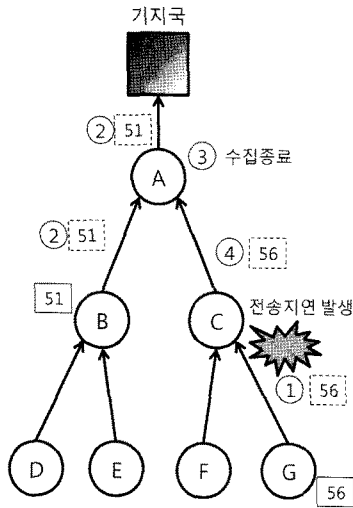


그림 6 전송지연에 따른 질의 처리 문제

이타가 먼저 전송되고, 기지국은 Top-1 질의의 결과로 B를 선택한 후 수집을 종료할 수 있다. G 센서 노드의 값이 기지국까지 전송되는 것을 보장하기 위해 가장 멀리 있는 센서 노드가 기지국까지 데이터를 전송할 수 있을 정도로 시간 슬롯을 길게 설정할 수 있다. 하지만, 데이터 전송의 지연은 데이터의 충돌에 의해 지연될 수 있기 때문에 적절한 해결책이 아니다.

본 논문에서는 이를 해결하기 위해 그림 7과 같은 TDMA 스케줄 구조를 사용한다. 각 TDMA 프레임이 종료되는 시점에서 낮은 종료 검사를 수행하여 라우팅 트리의 최상위 노드가 k개 이상의 데이터를 전송했는지 검사한다. 본 논문에서는 이러한 검사 단계를 “K검사”라고 정의한다. 즉, k개의 데이터를 전송하는 시간 슬롯을 종료하자마자 데이터 수집을 중단하는 것이 아니라 해당 시간 슬롯을 포함하고 있는 TDMA 프레임의 종료시점에서 전송 횟수를 검사한다.

최종 수정된 우선 순위 기반의 Top-k 질의 처리 기법은 그림 8과 같은 순서로 진행된다. 그림 9는 제안하는 Top-k 질의 처리 의사코드를 나타낸다. 먼저 TAG와 같은 방법으로 라우팅 트리를 생성하고, 각 센서 노

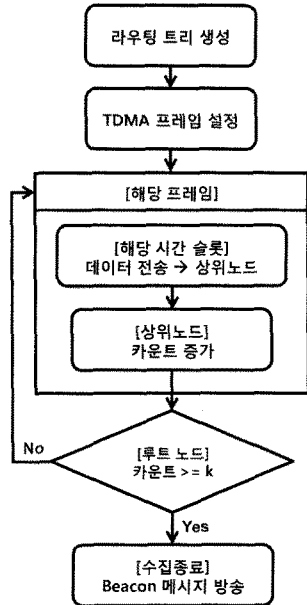


그림 8 우선순위 기반의 Top-k 질의 처리 순서

드는 데이터 인지 우선 순위 부여 기법을 통해 TDMA 프레임을 선정한다. 그 다음, 해당 프레임, 해당 시간 슬롯에 데이터를 상위 노드로 전송하고, 상위 노드는 자신의 카운터 값을 하나 증가시킨다. 해당 TDMA 프레임이 모두 종료되는 시점에서 루트 노드는 카운트 값이 k개 이상인지 검사하고, k보다 큰 경우 수집을 종료하기 위해서 제어 신호를 네트워크로 전송한다. 이때, 제어 신호는 단순히 전체 네트워크의 동작을 정지하기 위한 신호로 일반 메시지에 비해 크기가 작다. 만약 전송된 데이터의 수가 k보다 작은 경우, 다음 TDMA 프레임을 진행하여 더 많은 데이터를 수집을 시도한다.

3.3 센서 데이터의 압축

일반적인 Top-k 질의의 결과값으로 해당 센서 노드는 식별자와 함께 센서 데이터를 함께 전송한다. 제안하는 질의 처리 기법의 경우, 센서 데이터의 범위를 이용하여 TDMA 프레임의 번호를 계산하기 때문에 TDMA 프레임에 포함되는 각각의 데이터 값은 양자화가 가능

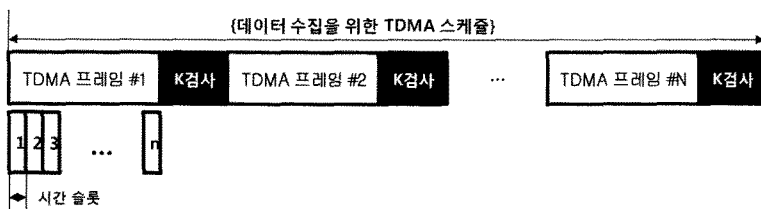


그림 7 데이터 수집을 위한 TDMA 스케줄

```

// 수집 데이터의 범위와 시간 주기를 고려한 TDMA 스케줄 생성
tdmaSchedule = createTDMA(Vmax, Vmin, tduration)
sensors = {s1, s2, s3, ..., sn}
// 배포된 센서들을 바탕으로 라우팅 트리 구축
IF createRoutingTree(sensors) THEN
    // 전체 센서들에게 TDMA 스케줄 배포
    broadcast(tdmaSchedule)
    WHILE network == live DO
        Round++
        // 모든 센서들의 데이터 수집
        sensors.sensing
        FOR EACH s = sensors[i] DO
            v = value of s
            // 각 센서마다 자신이 수집한 데이터에 따른 송신할 TDMA Frame 번호 계산
            tdmaFrameNo = (Vmax-V)/(Vmax-Vmin)*Cslot
            s.sendSlot = tdmaFrameNo
        END FOR
        // 각 센서노드는 자신의 송신 TDMA 슬롯에 맞춰 수집 데이터를 기지국으로 전송
        FOR EACH slot = tdmaSlot[i] DO
            FOR EACH s = sensors[i] DO
                IF s.sendSlot == slot THEN
                    sendToBS(value of s)
                    s.parentNodes.count++
                END IF
            END FOR
            // 기지국은 수신된 데이터가 k개를 만족하면 전체 네트워크에 송신정지 메시지 전송
            IF root.receivedDataCount >= top-k THEN
                broadcast(beacon)
                Exit
            END IF
        END FOR
    END WHILE
END IF
    
```

그림 9 우선순위 기반의 Top-k 질의 처리 의사코드

TDMA 프레임 #1	K검사	TDMA 프레임 #2	K검사
{데이터의 범위}		{데이터의 범위}	
100~91		90~81	데이터 범위
100, 95, 91		90, 85, 81	데이터의 예
0xF, 0x8, 0x0		0xF, 0x8, 0x0	양자화 변환

그림 10 양자화 과정

하다. 그림 10은 제안하는 질의 처리 기법에서 양자화 과정의 예를 나타낸다. TDMA 프레임 #1은 (100,91) 범위의 데이터를 수집한다고 할 때, 각 데이터 {100,95,91}은 {0xF, 0x8, 0x0}으로 4레벨 양자화가 가능하다. 양자화 값은 각 프레임별로 독립적으로 사용되며, 기지국은 TDMA 프레임 번호를 이용하여 수집한 양자화 값을 해당 범위의 값으로 복원한다.

3.4 우선 순위 기반의 연속적인 Top-k 질의 처리

제안하는 질의 처리 기법은 연속적인 Top-k 질의를 처리하기 위해서 가장 직관적이고 간단한 방법을 사용한다. 즉, 매 데이터 측정 주기(Time Unit)마다 앞서 제안한 Top-k 질의를 반복하여 실행한다. 복잡한 기법을 수행하기 위한 여러 과정을 수행하지 않는다는 점에서 직관적이고 간단한 방법은 비교적 컴퓨팅 성능이 낮은 센서 네트워크 환경에 더욱 적합하다.

4. 성능 평가

4.1 실험 환경

제안하는 연속적인 Top-k 질의 처리 기법의 우수성을 보이기 위해 기존에 제안된 연속적인 Top-k 질의 처리 기법인 FILA와 시뮬레이션을 통해 성능을 비교평가하였다. 시뮬레이션 툴은 JAVA로 직접 구현을 하였으며, 그림 11과 같이 TAG 기반의 네트워크 모양을 구성

표 1 성능 평가 환경

파라미터	기본값	변화
센서 개수	120(레벨=5)	48(레벨=3) ~440(레벨=10)
센서 식별자	4byte	-
측정 데이터의 크기	4byte	-
필터 윈도우의 크기	8byte	-
검증 메시지의 크기	4byte	-
Beacon 메시지의 크기	1bit	-
측정 데이터 양자화	16레벨	-
k	1	1~10
TDMA 프레임 수	70	10~100
Top-k 변동률	1.5%	1~5%, 10%, 100%

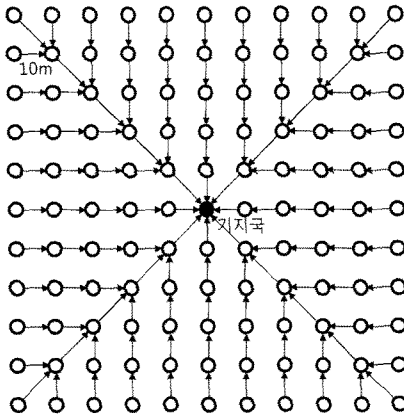


그림 11 다중-홉 네트워크 토폴로지

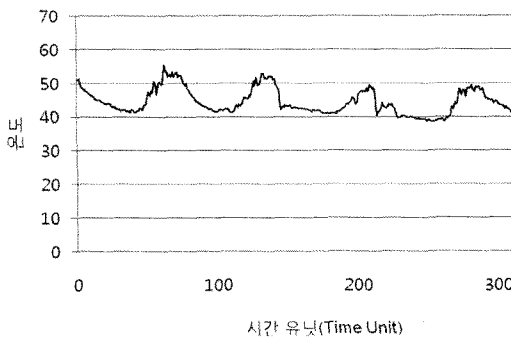


그림 12 시뮬레이션에서 사용한 데이터 모델

하였다. 실험은 표 1과 같이 센서의 수를 48개~440개로 변화시키면서 네트워크 수명을 측정하였다. 센서 노드의 메시지 전송에 소모되는 에너지 모델은 {메시지 크기} × ((전송 비용) + {증폭 비용} × {거리})이며, 전송 비용은 50nJ/b, 증폭 비용은 100pJ/b/m²으로 설정하였다. 메시지 수신에 소모되는 에너지 모델은 {메시지 크기} × {수신 비용}이며, 수신 비용은 50nJ/b으로 설정하였다[12,13].

시뮬레이션에서 사용한 데이터 모델은 LEM 프로젝트[14]에 의해 수집된 데이터로서 미국 워싱턴주의 환경 정보(온도, 습도, 광량 등)를 측정한 것이다. 여러 개의 센서 데이터 샘플을 추출하기 위해서 FILA기법의 실험 방법과 동일하게 데이터의 위상차를 두어 각 센서 노드가 서로 다른 데이터를 수집하도록 설정하였으며, 위상차의 크기를 조절하여 각 센서에서 수집되는 데이터의 변동률을 조절하였다.

4.2 최적의 TDMA 프레임 수

그림 13은 TDMA 프레임 수에 따른 제안하는 기법의 네트워크 수명의 변화를 나타낸다. TDMA 프레임 수에 따라 네트워크 수명이 다르게 나타나며, TDMA 프레임 수가 70인 경우 가장 좋은 성능을 보인다. TDMA 프레임 수는 k 검사 시점에 영향을 끼친다. 직관적으로 TDMA 프레임 수가 많을수록 단일 TDMA 프레임 동안 수집하는 데이터의 범위가 좁아지고, False Positive의 수가 줄어들 것으로 예상된다. 하지만, 실제 데이터의 발생 빈도와 분포는 데이터의 범위와 상관없으므로 각 응용분야의 데이터 특성에 따라 TDMA 프레임 수에 따른 네트워크 수명의 변화가 발생한다. 따라서, 응용분야에 최적화된 성능을 발휘하기 위해서 센서 노드를 배포하기 이전에 데이터의 수집 범위와 TDMA 프레임의 수를 미리 결정할 수 있다. TDMA 프레임 수와 데이터 수집 범위에 따라 2~3배 이상의 네트워크 수명 차가 발생한다.

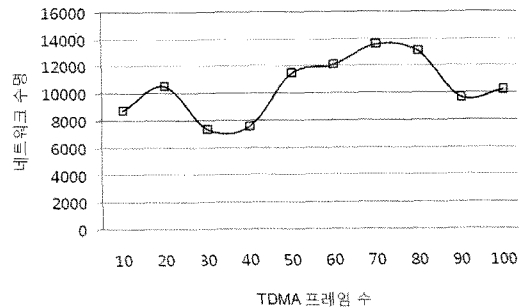


그림 13 TDMA 프레임 수에 따른 네트워크 수명

4.3 Top-k의 변동률에 따른 네트워크 수명 비교

그림 14는 Top-k의 변동률에 따른 네트워크 수명을 평가한 결과이다. 여기서 Top-k의 변동률은 Top-k 질의에 대한 결과의 변화 정도를 나타낸다. FILA는 Top-k의 변동률이 증가할 경우 필터를 갱신해야 하는 센서 노드의 수가 크게 증가한다. 이로 인해 메시지의 송/수신량이 크게 증가하여 네트워크 수명이 줄어든다. 제안하는 기법의 경우 FILA에서 발생하는 재검증과 필터 갱신을 위한 메시지 송수신이 없어 Top-k의 변동률

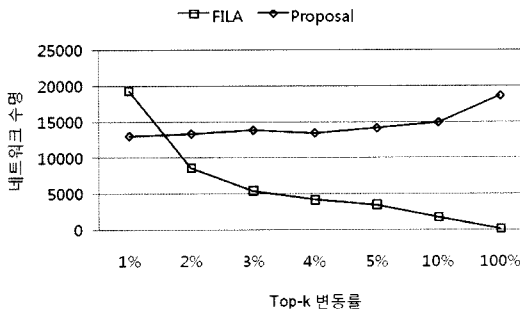


그림 14 Top-k의 변동률에 따른 네트워크 수명

이 증가할수록 우수한 성능을 보이며, 오히려 변동률이 큰 경우 Top-k에 해당하는 센서 노드의 통신 비용을 분산시켜 네트워크 수명이 증가한다.

4.4 k값에 따른 네트워크 수명 비교

그림 15는 Top-k의 개수에 따라 Top-k 변동률이 1.5%인 제안하는 기법과 FILA의 네트워크 수명을 평가한 결과이다. FILA기법은 k의 수가 증가 할수록 필터의 범위를 벗어나는 데이터가 발생할 확률이 증가하게 된다. 따라서 k의 수가 증가함에 따라 전체 네트워크의 수명이 크게 감소하는 것을 확인할 수 있다. 제안하는 기법은 필터의 갱신에 따른 부가적인 에너지 소모가 발생하지 않아 FILA에 비해 약 0.3배~2배 이상 네트워크 수명이 연장된다.

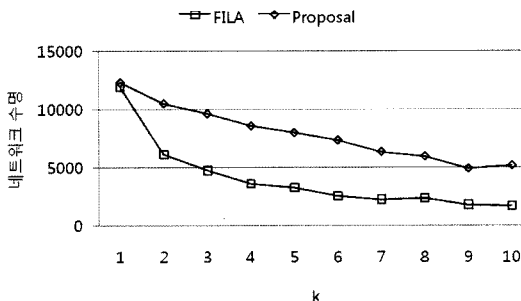


그림 15 Top-k의 수에 따른 네트워크 수명

4.5 센서 노드수에 따른 네트워크 수명 비교

그림 16은 센서 노드수에 따른 네트워크 수명의 향상 정도를 나타낸다. 적은 수의 노드로 구성된 소규모 네트워크인 경우 FILA 기법에 비해 성능이 떨어지지만, 많은 수의 노드로 구성된 대규모 네트워크일수록 좋은 성능을 보인다. 즉, 제안하는 Top-k 질의 처리 기법은 대규모 네트워크에 사용하기 적합하다.

5. 결론

본 논문은 기존 필터 기반의 Top-k 질의 처리 기법

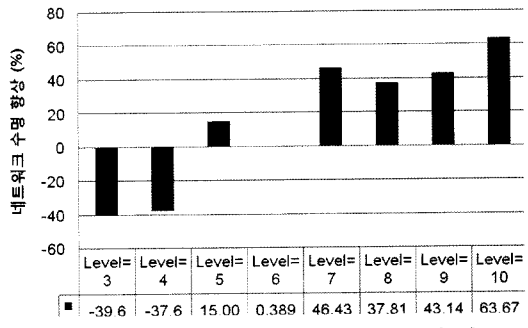


그림 16 센서 노드수에 따른 네트워크 수명 비교

의 문제점을 분석하고, 우선 순위 기반의 Top-k 질의 처리 기법을 제안하였다. 기존 기법의 경우, 데이터 정확성을 보장하기 위한 재검증 과정에서 많은 센서 데이터의 전송이 발생한다. 뿐만 아니라 변경된 필터의 정보를 갱신하기 위한 메시지 송/수신이 발생한다. 이러한 문제점을 해결하기 위해서 제안하는 기법은 측정된 데이터를 기반으로 우선순위를 부여하고, 우선순위에 따라 Top-k에 해당하는 센서 노드만 데이터 수집에 참여시킨다. 이외 센서 노드는 수면 상태를 유지하고, 데이터 수집 과정에서 추가적인 재검증이나 필터 전송을 제거하여, 센서 데이터의 전송을 현저하게 줄였다. 실험 결과, 네트워크의 규모가 커지거나 Top-k의 수가 커지는 경우 네트워크의 수명이 크게 향상되었고, Top-k의 변동률이 클수록 효율적인 질의 처리가 가능하다. 따라서, 제안하는 기법은 센서 데이터의 변동률이 크거나, 대규모 네트워크를 이용하는 응용분야에 더욱 적합하다. 제안하는 기법은 데이터 전송에 있어 계층적인 구조를 사용하기 때문에 루트 노드나 일부 센서 노드에 데이터 전송의 오버헤드가 발생한다. 향후 연구는 이러한 문제점을 해결하기 위해 데이터 전송의 오버헤드를 분산하는 기법을 제안하는 것이다.

참고 문헌

[1] R. Szweczyk, A. Mainwaring, J. Polastre, and D. Culler, "An Analysis of a Large Scale Habitat Monitoring Application," Proc. ACM Conf. Embedded Networked Sensor Systems (SenSys '04) Nov. 2004.

[2] R. Szweczyk, E. Osterweil, J. Polastre, M. Hamilton, A. Mainwaring, and D. Estrin, "Habitat Monitoring with Sensor Networks," Comm. ACM, Vol.47, No.6, pp. 34-40, Jun. 2004.

[3] J. Considine, F. Li, G. Kollios, and J. Byers, "Approximate Aggregation Techniques for Sensor Databases," Proc. IEEE Int'l Conf. Data Eng. (ICDE '04), Mar. 2004.

[4] K. Mouratidis, D. Papadias, S. Bakiras, and Y.

Tao, "A Threshold-Based Algorithm for Continuous Monitoring of k Nearest Neighbors," IEEE Trans. Knowledge and Data Eng., Vol.17, No.11, pp. 1451-1464, Nov. 2005.

[5] M.A. Sharaf, J. Beaver, A. Labrinidis, and P.K. Chrysanthis, "Balancing Energy Efficiency and Quality of Aggregate Data in Sensor Networks," VLDB J., Vol.13, No.4, pp. 374-403, Dec. 2004.

[6] N. Shrivastava, C. Buragohain, D. Agrawal, and S. Suri, "Medians and Beyond: New Aggregation Techniques for Sensor Networks," Proc. ACM Conf. Embedded Networked Sensor Systems (SenSys '04) Nov. 2004.

[7] 김진수, 박찬홍, 김종근, 강병욱, "센서 네트워크에서 계층적 필터링을 이용한 에너지 효율적인 데이터 집계 연산", 한국컴퓨터정보학회논문지, 제12권, 제1호, pp. 79-88, 2007년 3월.

[8] 김혜정, 지경복, 김창화, 김상경, 박찬정, "해양 센서 네트워크 아키텍처 중심의 질의 최적화를 위한 데이터 병합 기법", 한국정보과학회 2007 가을 학술발표 논문집, 제34권, 제2호(D), pp. 215-220, 2007년 10월.

[9] 김현태, 유태영, 정규수, 전영배, 나인호, "무선 센서 네트워크에서 데이터 수집의 효율성 및 정확성 향상을 위한 데이터 병합기법", 한국지능시스템학회 논문지(표지 및 지능), 제16권, 제5호, pp. 531-536, 2006년 10월.

[10] S. Madden, M.J. Franklin, J.M. Hellerstein, and W. Hong, "TAG: A Tiny Aggregation Service for Ad Hoc Sensor Networks," Proc. Usenix Fifth Symp. Operating Systems Design and Implementation (OSDI '02), pp. 131-146, Dec. 2002.

[11] W. Minji, X. Jianliang, T. Xueyan, L. Wang-Chien, "Top-k Monitoring in Wireless Sensor Networks," IEEE Trans. Knowledge and Data Engineering, Vol.19, No.7, pp. 962-976, Jul. 2007.

[12] W. Heinzelman, "Application-Specific Protocol Architectures for Wireless Networks", PhD dissertation, Massachusetts Inst. Of Technology, 2000.

[13] X. Tang, and J. Xu, "Extending Network Lifetime for Precision-Constrained Data Aggregation in Wireless Sensor Networks," Proc. IEEE INFOCOM, Apr. 2006.

[14] Live from Earth and Mars (LEM) Project, <http://www-k12.atmos.washington.edu/k12/grayskies/>, 2006.



성 동 욱

2005년 2월 충북대학교 정보통신공학과 공학사. 2007년 2월 충북대학교 정보통신공학과 공학석사. 2007년~현재 충북대학교 정보통신공학과 박사과정. 관심분야는 무선 센서 네트워크, 데이터베이스 시스템, LMS/LCMS, FLASH 메모리

저장 시스템.

유 재 수

정보과학회논문지 : 데이터베이스 제 36 권 제 1 호 참조



여 명 호

2004년 2월 충북대학교 정보통신공학과 공학사. 2006년 2월 충북대학교 정보통신공학과 공학석사. 2006년~현재 충북대학교 정보통신공학과 박사과정. 관심분야는 메인 메모리 기반 데이터베이스 시스템, 시공간 데이터베이스 시스템, 무선

센서 네트워크 등