

센서네트워크 환경에서 효율적인 데이터 퓨전 기법

최경*, 박경란*, 채기준*, 박종준**, 주성순**

*이화여자대학교 컴퓨터공학과

**한국전자통신연구원

e-mail : cbk0907@ewhain.net, pkr0226@ewhain.net, kjchae@ewha.ac.kr

juny@etri.re.kr, ssjoo@etri.re.kr

An Efficient Data Fusion Mechanism on Wireless Sensor Networks

Kyung Choi*, Kyung-Ran Park*, Ki-Joon Chae*

Jong-Jun Park**, Seong-Soon Joo**

*Dept. of Computer Science and Engineering, Ewha Womans University

**Electronics and Telecommunications Research Institute

요 약

제한된 배터리를 가지는 센서 네트워크의 수명을 길게 하기 위하여 에너지 소비를 줄이고 효율성을 높이는 다양한 방법이 제안되었다. 센서 노드에서의 데이터 전송은 가장 큰 에너지 소비 활동이 되기 때문에, 데이터가 전송되는 양을 줄여서 에너지 소비를 줄일 수 있는 방법이 한 부분으로 연구되고 있다. 본 논문에서는 이와 같은 특징을 반영하고, 시간적 연관성만을 고려한 데이터 퓨전 기법인 TiNA(Temporal coherency-aware in-Network Aggregation)를 기반으로 시간, 공간적 연관성을 동시에 고려하여 데이터 퓨전을 하는 효율적인 데이터 퓨전(Data Fusion) 기법을 제안하였다.

1. 서론

유비쿼터스 컴퓨팅 환경의 기본이 되는 기반 기술로서 센서 네트워킹 기술의 중요성 또한 강조되고 있다. 센서 네트워크는 유무선 네트워크 인프라에 다양한 센서를 설치하고, 이를 통해 데이터를 감지하며, 감지된 데이터를 응용서비스 서버와 연동하는 기술이다. 이러한 센서 네트워크는 군사, 환경 모니터링, 의료, 농업, 날씨 분석, 위치 정보 시스템 등 다양한 응용 분야에서 활용될 수 있다. 센서 네트워크는 작은 크기의 센서에서 오는 제약점 때문에 메모리와 처리 능력, 에너지에 제한을 갖는다. 그러므로 센서 네트워크 상에서의 보안 기술 연구에 있어서 센서가 갖는 무선 통신의 기본 취약점뿐만 아니라 센서 노드의 크기 자체에서 오는 제약점 또한 고려하여야 하는 것이다.

본 논문에서는 이러한 무선 센서 네트워크의 특징과 최신의 데이터 퓨전 기법들에 대해서 분석하여 이에 적합한 효율적인 데이터 퓨전 기법을 제안하였다. 이를 위해 시간, 공간적인 연관성을 동시에 고려하여 데이터 퓨전을 진행함으로써 효율성과 퓨전 후 데이터의 정확성을 고려하였다.

센서 노드에서의 데이터 전송은 가장 큰 에너지 소비 활동이 되기 때문에 데이터가 전송되는 양을 줄

여서 에너지 소비를 줄일 수 있는 방법이 한 부분으로 연구되고 있으며, 데이터 애그리게이션 및 퓨전이 하나의 요구 사항이 되었다.

데이터 애그리게이션은 애그리게이션 노드가 같은 종류의 센싱 데이터를 모은 후 정해놓은 함수(Sum, Avg, Min 등)를 통해 하나의 데이터로 재생성하여 전송함으로써 에너지 효율성을 증대시킨다. 데이터 퓨전은 퓨전 노드가 같거나 다른 종류의 센싱 데이터를 모은 후, 연관성 정보(시간, 공간, 데이터) 및 정해놓은 기준에 따라 새로운 데이터를 생성하여 전송하는 기법을 일컫는다. 또한, 퓨전 후 생성된 데이터에 대한 정확성을 높이기 위한 연구와 퓨전 결과에 대한 신뢰성을 높이기 위한 연구가 함께 진행되고 있다. 주로 데이터 퓨전과 데이터 애그리게이션은 혼용하여 사용되고 있으며, 본 연구에서는 연관성 정보를 토대로 데이터의 양을 줄여서 전송하는 기법을 제안하였기에 데이터 퓨전으로 사용하기로 한다.

2. 관련연구

데이터 애그리게이션 기법은 공간, 시간, 데이터에 따른 연관성을 하나 혹은 그 이상 고려하여 이 정보를 토대로 애그리게이션을 진행하는 것으로 분류해 볼 수 있다.

공간적인 연관성을 고려한 기법은 가까운 곳에 있는 노드는 비슷한 값을 가지게 될 확률이 높다는 가정 아래 MFST, AFST, Dfuse, CAC, CODA, CAG, PREMON 과 같은 여러 연구가 제안되었다[1, 2, 3, 4, 5,

1) 본 연구는 지식경제부의 IT 원천기술개발사업의 일환으로 수행하였음. [2008-F-052, QoS 및 확장성지원(S-MoRe) 센서네트워크 고도화 기술개발]

6, 7].

데이터 간 연관성을 고려하여 데이터 애그리게이션을 제안한 Distributed DF[8]는 Alexei Makarenko 와 Hugh Durrant-Whyte 가 제안한 것으로 센싱된 데이터가 넓은 곳에 퍼진 분산된 데이터 퓨전을 실행할 때에 유연성이 있는 시스템 아키텍처를 위하여 제안된 데이터 퓨전 방법이며, 시간적 연관성을 고려하여 데이터 애그리케이션 기법을 제안한 연구들로 MFS, TiNA, CAG, PREMON 등이 대표적이다[9, 10, 6, 7].

TiNA(Temporal coherency-aware in-Network Aggregation)[10]에서는 시간적 연관성을 고려하여 새로운 센서 리딩 값의 전송 여부를 결정함으로써 전송량을 줄일 수 있었다. 네트워크 내부에서의 애그리케이션을 통해 에너지 소비를 줄임과 동시에 애그리케이션 후에 생성된 데이터의 질을 높일 수 있는 방법을 제안하였고 시뮬레이션을 통해 성능을 분석하였다.

리프 노드(Leaf node)는 부모 노드에게 성공적으로 전송된 마지막 리딩 값(V_{old})을 저장하고 있고, 중간 노드들은 각각의 자식 노드로부터 받은 마지막 데이터와 그 노드에서 리딩된 마지막 데이터를 가지고 있다. 리프 노드가 새로운 값(V_{new})을 읽었을 때 마지막 리딩 값과 비교하여, 그 값이 미리 정해진 값(tct)보다 클 경우에만($(V_{new}-V_{old}) / V_{old} > tct$) 새로운 데이터를 전송하며, 부모 노드는 자식 노드로부터 온 데이터가 없을 때 그 노드로부터 받은 마지막 데이터로 애그리케이션하여 전송한다.

밀집된 네트워크 상에서 트래픽 양이 많은 경우와 쿼리의 응답 요구 시간이 짧을 경우 충돌이나 재전송으로 인하여 각 노드에서 리딩된 값이 손실된 채로 애그리케이션이 진행된다면, 애그리케이션 후의 결과가 부정확해 질 수 있으나, TiNA 를 사용하여 전송을 줄임으로써 이런 손실을 막을 수 있어서 애그리케이션 후에 생성된 데이터의 정확성을 높일 수 있다.

CAG(The Clustered AGgregation)[6]는 정해진 임계치(threshold) 내에 있는 값을 가진 노드로 클러스터를 구성함으로써 공간적인 연관성을 고려하였고, 센서가 리딩한 값이 임계치 내에 있는 동안 기존의 클러스터를 유지하기 때문에 그 시간 동안의 연관성을 고려하였다. 센서 모듈에서 측정된 데이터와 모델을 통해 생성된 데이터를 이용하여 공간적인 연관성뿐 아니라 시간적인 연관성을 고려하여 CAG 의 효율성을 나타냈고, 인터랙티브 모드와 스트리밍 모드로 설계하여 다른 목적과 환경에 따라 다양하게 쓰일 수 있게 하였다.

3. 효율적인 데이터 퓨전 기법 제안

3.1 시간적 연관성

시간적 연관성을 고려한 TiNA 를 기반으로 시간, 공간적 연관성을 동시에 고려하여 데이터를 퓨전함에 따라 전송량을 줄이고 정확성은 유지할 수 있는 데이터 퓨전 기법을 제안한다.

시간적 연관성을 고려한 기법은 데이터 전송 흐름에서 이전 시간과의 연관성을 고려하여 새로운 센싱

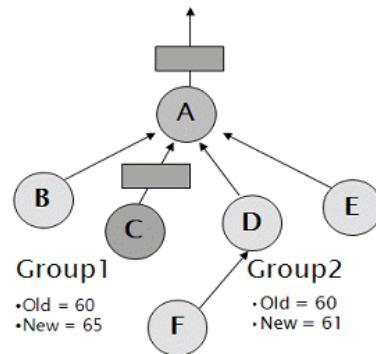
데이터의 전송 여부를 결정하며, 이를 위한 식은 다음과 같다.

$$|V_n - V_{n-1}| > \alpha \cdot \tau$$

리프 노드는 부모 노드에게 성공적으로 전송된 마지막 센싱 데이터(V_{n-1})를 저장하고 있고, 중간 노드들은 각각의 자식 노드로부터 받은 마지막 데이터와 그 노드에서 센싱된 마지막 데이터를 가지고 있다.

리프 노드가 새로운 데이터(V_n)를 읽었을 때 마지막 센싱 데이터와 비교하여 그 값이 사용자가 미리 정한 값(τ)과 α 의 곱보다 클 경우에만 새로운 데이터를 전송하며, 부모 노드는 자식 노드로부터 온 데이터가 없을 때 그 노드로부터 받은 마지막 데이터로 퓨전하여 전송한다.

α 는 디폴트 값으로 1을 갖는다. 하지만 0과 1 사이의 계수로 데이터 셋을 통한 실험을 통해 전송량과 전체를 보냈을 때와 비교했을 때, 줄어든 전송량에 비해 보다 정확한 결과를 낼 수 있는 최적의 값을 도출하여 사용자가 원하는 정확도 내에서 효율적인 데이터 퓨전을 진행할 수 있게 한다. 주기적으로 비슷한 데이터의 흐름이 있을 경우, 다음 주기에 사용자가 정한 한계치 내에서 효율적인 α 값을 적용시킬 수 있다.



[그림 1] 시간적 연관성을 고려한 데이터 퓨전

[그림 1]과 같이 τ 가 2 라면, Group 1 의 노드 C에서는 새로운 데이터를 전송하는 반면 Group 2 의 노드 D에서의 새로운 데이터는 전송되지 않는다. 노드 A에서는 노드 C의 새로운 데이터와 노드 B, D, E의 이전 데이터와 함께 자신의 센싱 데이터를 함께 퓨전하여 전송한다.

TiNA 에서 tct 라는 비율적인 한계값을 통해 센싱된 데이터의 전송 여부를 결정하여 데이터의 크기에 따라 전송 여부가 달라지고, 그에 따르는 결과의 정확도도 달라진다. 제안하는 기법에서는 τ 라는 이전 데이터와 새로운 데이터와의 차이값을 통해 전송 여부를 결정한다. 예를 들어 TiNA에서는 같은 tct 값에 대해 센싱 데이터 20 과 60 에서 적용되는 한계치가 달랐으나 제안된 기법에서는 차이값을 적용함으로써 센싱 데이터의 크기에 관계없이 한계치를 적용할 수 있다.

3.2 공간적 연관성

가까운 노드는 연관성 정도가 클 확률이 높기 때문에 자신의 이웃 노드와의 연관성을 고려하여 새로 센싱된 데이터의 전송 여부를 결정한다. 이를 위한 식은 다음과 같다.

$$|V_n - V_{n-1}| > \beta \cdot \tau$$

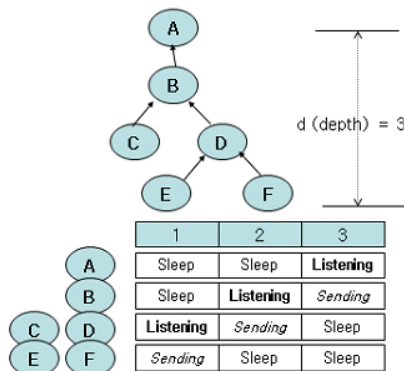
센서 노드는 한 홉 내의 주변 노드가 마지막으로 보낸 값을 오버헤어링하고, 한 홉으로 연결된 이웃 노드의 마지막 오버헤어링 데이터(V_1)와 새로운 센싱 데이터(V_2)를 비교 후 전송 여부를 결정한다.

리프 노드가 새로운 데이터(V_2)를 읽었을 때 최근에 주변 노드에서 오버헤어링한 마지막 센싱 데이터와 비교하여, 그 값이 사용자가 미리 정한 값(τ)과 β 의 곱보다 클 경우에만 새로운 데이터를 전송하며, 부모 노드는 자식 노드로부터 온 데이터가 없을 때 그 노드로부터 받은 마지막 데이터로 퓨전하여 전송한다.

β 는 디폴트 값으로 1을 갖는다. 하지만 0과 1 사이의 계수를 적용시킨 후 데이터 셋을 통한 실험으로 전송량과 전체를 보냈을 때와 비교했을 때 줄어든 전송량에 비해, 보다 정확한 결과를 낼 수 있는 최적의 값을 도출하여 사용자가 원하는 한계치 내에서 효율적인 데이터 퓨전을 진행할 수 있게 한다.

3.3 데이터 퓨전을 위한 동기화

TAG(Tiny AGgregation)[11]에서 센서 노드 간의 동기화를 위해 사용한 기법을 토대로 연관성을 고려한 데이터 퓨전 노드 간의 동기화를 진행한다.



[그림 2] 노드 간 동기화

각 커뮤니케이션 슬롯은 주어진 데이터 전송 주기(duration)를 클러스터 내부의 트리 깊이(depth)로 나눈다.

이와 같이 생성된 커뮤니케이션 슬롯 동안 각 노드의 상태는 [그림 2]와 같다. 라우팅 트리의 깊이가 3일 때 총 세 개의 슬롯이 생성되게 되고, 첫 번째 슬롯에서 노드 C, D, E, F는 각각 데이터를 전송하고 받기 위해 깨어 있게 된다. 이 때 노드 E와 F에서 받은 데이터는 노드 D에서 퓨전하여 다음 레벨로 전송하고 위의 레벨에 있는 노드 A와 B는 각각 수면(Sleep) 상태를 유지한다. 두 번째 슬롯에서는 노드 C와 D가 데이터를 전송하고 노드 B는 이 데이터를 받기 위해 깨어있으며, 세 번째 슬롯도 이와 같다.

동기화를 위해 데이터를 전송하는 레벨의 노드와 데이터를 받는 레벨의 노드는 주어진 주기 동안 항상 깨어 있어야 하므로, 공간적 연관성을 추가로 고려시 발생하는 오버헤어링을 위해 깨어야 하는 오버헤드가 추가로 발생하지 않음을 알 수 있다.

4. 실험 및 결과 분석

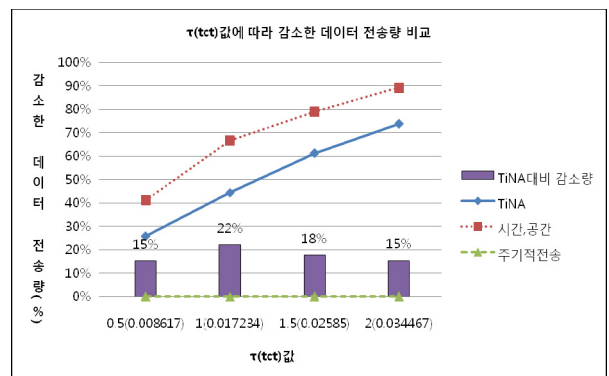
센서 네트워크 구조에서 시간, 공간적 연관성에 따른 데이터 퓨전의 효율성을 증명하기 위하여 시뮬레이터 Qualnet을 이용하여 전송량과 퓨전 결과의 정확성을 측정하는 실험을 진행하였다.

Qualnet은 가상의 네트워크 기반에서 다양한 프로토콜 설계 분석 검증과 네트워크 어플리케이션 등을 실제로 구축하기 전에 가상의 공간에서 구축하여 문제점을 분석하고 예측할 수 있는 소프트웨어이다. WLAN를 비롯하여, 센서 네트워크, 모바일 애드혹 네트워크, Wired LAN, WAN, Cellular networks, Satellite Network 등에 대한 기본적인 라이브러리를 가지고 있으며 사용자가 요구에 따라 프로토콜을 수정하여 사용할 수 있다. 본 연구에서는 Zigbee 프로토콜을 사용하고 있는 Qualnet 4.5 버전의 센서 네트워크 라이브러리를 사용하였고, 프로토콜 추가 및 수정을 통해 실험하였다.

제한한 시간, 공간적 연관성을 동시에 고려하여 퓨전한 경우와 TiNA를 통해 퓨전한 경우를 비교하여 줄어든 데이터의 전송량을 통한 효율성 및 퓨전 결과의 정확성을 증명하도록 하였다. 실험은 15개 노드가 트리(tree) 구조로 구성된 클러스터로, 하위 노드에서 클러스터 헤더 노드까지 세 홉일 때로 진행되었다.

실험에 사용된 데이터는 온도 데이터로 단위는 화씨(Fahrenheit)이다. 온도 데이터는 비슷한 시간에 측정할 때와 가까운 곳에 있을수록 연관성을 많이 띄기 때문에 이를 통한 실험은 효율적일 것이다.

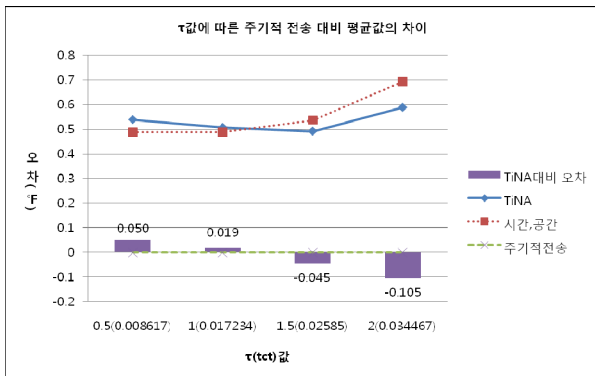
실측 데이터 값을 평균으로 하고 분산값 1인, 가우시안 분포를 따르는 공간적 연관성을 고려한 데이터 셋을 형성하여 클러스터 내부의 노드 센싱 데이터로 사용하였다. 데이터 퓨전을 위한 함수는 평균을 이용하였으며, 1분 간격으로 총 600번의 패킷을 전송하였다. 4바이트씩 총 600번의 패킷 전송으로 주기적 전송량은 총 36,000바이트이며, 이때 줄어든 전송량을 실험을 통해 결과 도출해 보았다.



[그림 3] τ (tct)값에 따라 감소한 데이터 전송량 비교

[그림 3]은 τ (tct)값에 따라 줄어드는 전송량을 보여주고 있다. 여기서 tct 가 0.008617 일 때와 τ 값이 0.5 일 때, 주기적 전송 시 평균값에서 비슷한 정도의 차이를 보이기 때문에 함께 쓰고, 이를 통해 TiNA 와 제안 기법을 비교하였다.

τ (tct)값이 증가함에 따라 감소한 데이터 전송량이 비슷하게 증가함을 보이지만, 시간·공간적 연관성을 고려한 제안안이 최대 22% 더 효율적인 감소량을 보임을 나타낸다.



[그림 4] τ (tct)값에 따른 주기적 전송 대비 평균값의 차이

[그림 4]에서는 평균 온도 값의 변화 추이에 따라 제안한 기법과 TiNA 를 비교하면, τ 가 0.5 와 1 인 경우는 오차값이 TiNA 보다 적음을 보이고 1.5 와 2 인 경우엔 TiNA 가 조금 더 적은 오차를 보임을 나타내지만, TiNA 와 제안한 안의 오차값 차이가 최대 0.1 도를 넘지 않고, 미리 정한 한계값 내의 정확성을 유지할 수 있음을 보였다.

5. 결론 및 향후 연구

본 논문에서는 센서 네트워크 환경의 특성에 따라 시간, 공간적 연관성을 고려한 효율적이고 정확성을 유지할 수 있는 데이터 퓨전 기법을 제안하였다.

센서 네트워크는 노드 자원의 제약과 제한된 에너지 및 메모리 공간, 낮은 연산 능력 등 여러 제약 사항을 고려하여야 한다. 이와 같은 특징을 고려하여 데이터가 전송되는 양을 줄여서 에너지 소비를 줄일 수 있는 데이터 퓨전 기법을 제안하였고, 좀 더 효율적이고 정확한 퓨전을 위해 시간과 공간적 연관성을 동시에 고려하였다.

이전의 센싱 데이터를 고려하는 시간적 연관성 및 이웃 노드에서 마지막으로 전송한 센싱 데이터를 고려하는 공간적 연관성을 토대로 새로 센싱된 데이터의 전송 여부를 결정하였다. 이와 동시에 줄어드는 데이터 간의 퓨전으로도 정확성을 유지할 수 있게 하기 위하여 사용자가 정한 요구 정도에 따라 데이터의 전송 여부 결정 및 결과의 정확성을 유지하는 기법을 제안하였다.

이들은 실측 데이터를 이용하여 생성된 가우시안 분포를 따르는 데이터 셋과 Qualnet 시뮬레이터를 통

해 실험 및 분석하였다. 이를 통해 데이터가 전송되는 양이 감소하는 사실과 사용자의 요구에 따른 정확성을 유지함을 확인할 수 있었다.

앞으로 향후 연구로는 온도 데이터뿐만 아니라 여러 데이터 셋을 토대로 제안한 데이터 퓨전 기법을 보다 개선시키고 상세히 분석할 것이다. 정확성 분석 및 연관성 정보를 좀 더 효율적으로 활용하여 데이터의 전송량 및 오버헤드를 최소화하여 에너지 소모량을 줄이고, 연관성 정도에 따라 α , β 결정 여부를 개선하는 등 퓨전 후 결과의 정확성을 최대한 유지할 수 있는 기법에 대한 연구가 필요하다.

참고문헌

- [1] Hong Luo, Yonghe Liu, Sajal K. Das, "Routing Correlated Data with Fusion Cost in Wireless Sensor Networks," IEEE Transaction on Mobile Computing Vol. 5, No. 11, pp. 1620-1632, Nov. 2006.
- [2] Hong Luo, Yonghe Liu, Sajal K. Das, Jun Luo, "Adaptive Data Fusion for Energy Efficient Routing in Wireless Sensor Networks," IEEE Transaction on Mobile Computing Vol. 55, No. 10, Oct. 2006.
- [3] Rajnish Kumar, Matthew Wolenetz, Bikash Agarwalla, "DFuse: A Framework for Distributed Data Fusion," ACM SenSys, 2003.
- [4] Guang-yao Jin, Myoung-Soon Park, "CAC: Context Adaptive Clustering for Efficient Data Aggregation in Wireless Sensor Networks," NETWORKING 2006, LNCS.
- [5] SangHak Lee, JuneJae Yoo, TaeChoong Chung, "Distance-based Energy Efficient Clustering for Wireless Sensor Networks," IEEE International Conference on Local Computer Networks LCN, 2004.
- [6] Sunhee Yoon, Cyrus Shahabi, "The Clustered AGgregation (CAG) Technique Leveraging Spatial and Temporal Correlations in Wireless Sensor Networks," ACM Transactions on Sensor Network, May 2007.
- [7] Samir Goel, Tomasz Imielinski, "Prediction-based Monitoring in Sensor Network : Taking Lessons from MPEG," ACM Computer Comm. Review, 2001.
- [8] Alexei Makarenko, Hugh Durrant-Whyte, "Decentralized Data Fusion and Control in Active Sensor Networks," Seventh International Conference on Information Fusion, 2004.
- [9] Wei Yuan, Srikanth V. Krishnamurthy, Satish K. Tripathi, "Synchronization of Multiple Levels of Data Fusion in Wireless Sensor Networks," Global Telecommunications Conference, GLOBECOM, Dec. 2003.
- [10] Mohamed A. Sharaf, Jonathan Beaver, Alexandros Labrinidis, Panos K. Chrysanthis, "TiNA: A Scheme for Temporal Coherency-Aware in-Network Aggregation," ACM MobiDE, 2003.
- [11] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein, Wei Hong, "TAG: a Tiny AGgregation Service for Ad-Hoc Sensor Networks," 5th Annual Symposium on Operating Systems Design and Implementation (OSDI), Dec. 2002.