
무선 센서 네트워크 모니터링 시스템을 위한 데이터 통계 분석 기반 데이터 필터링 기법

Data Statical Analysis based Data Filtering Scheme for Monitoring System on Wireless Sensor Network

이현조, 최영호, 장재우
전북대학교 컴퓨터공학과

Hyun-Jo Lee(o2near@chonbuk.ac.kr), Young-Ho Choi(yhchoi@dblab.chonbuk.ac.kr)
Jae-Woo Chang(jwchang@chonbuk.ac.kr)

요약

최근 무선 센서 네트워크를 활용한 다양한 모니터링 시스템 구축이 활발히 이루어지고 있다. 무선 센서 네트워크 기반 모니터링 시스템 구축을 위해서, 세 가지 사항을 고려해야 한다. 첫째, 지속적인 모니터링을 위해서, 노드 실패 감지 기법이 요구된다. 둘째, 센서노드는 제한된 배터리 용량을 지니기 때문에, 에너지 소모량 감소를 위한 효율적인 데이터 필터링 기법이 요구된다. 마지막으로 데이터 필터링 수행 시, 계산 오버헤드를 감소시키는 기법이 필요하다. 기존 칼만 데이터 필터링 기법은 우수한 필터링 성능을 나타내는 반면, 데이터 예측값 계산과정이 복잡하여 센서 노드에서의 계산 오버헤드가 증가하는 단점이 존재한다. 이러한 문제점들을 해결하기 위하여 본 논문에서는 데이터 통계 분석 기반 데이터 필터링 기법을 제안한다. 제안하는 기법은 첫째, 노드 실패 감지를 지원하기 위해, 주기적으로 노드 생존 메시지를 수집한다. 둘째, 불필요한 전송 메시지 수 감소를 위하여 샘플 데이터를 노드 생존 메시지에 포함하여 전송하고, 수집된 샘플 데이터 집합을 바탕으로 데이터 필터링을 수행한다. 마지막으로, 서버에서 데이터 통계 분석을 이용한 데이터 필터링 범위를 계산하기 때문에, 센서에서는 단순 비교연산만을 수행함으로써 센서 노드에서의 계산 오버헤드를 감소시킨다. 아울러 성능 분석을 통해 제안하는 기법이 기존 칼만 필터링 기법보다 전송 메시지 수 측면에서 성능이 우수함을 보인다.

■ 중심어 : | 데이터 필터링 | 무선 센서 네트워크 | 모니터링 시스템 |

Abstract

Recently, various monitoring systems are implemented actively by using wireless sensor networks(WSN). When implementing WSN-based monitoring system, there are three important issues to consider. At First, we need to consider a sensor node failure detection method to support the ongoing monitoring. Secondly, because sensor nodes use limited battery power, we need an efficient data filtering method to reduce energy consumption. At Last, a reducing processing overhead method is necessary. The existing Kalman filtering scheme has good performance on data filtering, but it causes too much processing overhead to estimate sensed data. To solve these problems, we, in this paper, propose a new data filtering scheme based on data statical analysis. First, the proposed scheme periodically aggregates node survival massages to support a node failure detection. Secondly, to reduce energy consumption, it sends the sample data with a node survival message and do data filtering based on those messages. Finally, it analyzes the sample data to estimate filtering range in a server. As a result, each sensor node can use only simple compare operation for filtering data. In addition, we show from our performance analysis that the proposed scheme outperforms the Kalman filtering scheme in terms of the number of sending messages.

■ keyword : | Data Filtering | Wireless Sensor Network | Monitoring System |

* 본 연구는 산업자원부와 한국산업기술재단의 지역혁신인력양성사업으로 수행된 연구결과이며, 아울러 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. 2010-0000202)

접수번호 : #091118-002

심사완료일 : 2010년 02월 12일

접수일자 : 2009년 11월 18일

교신저자 : 장재우, e-mail : jwchang@chonbuk.ac.kr

I. 서론

최근 무선 통신과 전자 기기의 눈부신 발전으로 초소형, 고성능, 저비용의 무선 센서 노드 개발이 가능하다. 무선 센서 노드들은 건물, 도로, 강 등 다양한 공간에 배치 가능하여, 이를 활용한 무선 센서 네트워크(Wireless Sensor Network: WSN) 기술이 많은 관심을 받고 있다. 무선 센서 네트워크는 센서 노드 주변의 온도, 습도, 조도 등의 환경 정보를 지속적으로 감지하고, 무선 통신을 이용하여 감지된 데이터를 신속하게 서버에 전송할 수 있기 때문에, 시설 재배, 건물, 교량, 산불 감시 등의 분야에서 모니터링 시스템으로 활용되고 있다[1][6].

이러한 무선 센서 네트워크 기반의 모니터링 시스템을 구축하기 위해서는, 센서 노드의 자원 제약을 고려하는 것이 필수적이다. 일반적으로 무선 센서 노드란 배터리에 의해 전력공급을 받으며 통신기능이 있는 상대적으로 낮은 성능을 갖는 컴퓨터이다. 센서 노드는 전력을 비롯한, 계산 능력, 저장공간 등의 자원이 제한적이다. 특히, 센서 네트워크가 설치되는 장소가 쉽게 접근할 수 없는 장소인 경우에는 소모된 전력을 재충전하거나 배터리를 교체하는 것이 매우 어렵거나 불가능하다. 따라서 각 센서의 전력 소모량을 최소화하여, 센서 네트워크의 수명(lifetime)을 증가시키는 것이 필수적이다. 센서 내 에너지는 센서 내에서 데이터 처리를 수행할 때, 혹은 다른 센서와 데이터 통신을 수행할 때 소모된다. 이에 따라 데이터 송수신 시 전송되는 데이터의 크기를 감소시키거나 혹은 데이터의 전송 횟수를 감소시켜, 센서의 에너지 소모량을 감소시키기 위한 연구가 활발히 진행되고 있다. 특히, 무선 데이터 수신과 전송에 요구되는 에너지 소비가 매우 크기 때문에 [4][5], 데이터의 전송 횟수를 최대한 줄이는 것이 센서의 배터리 소모를 줄이기 위한 가장 좋은 방법이다 [2][5-8][10][11].

이에 따라 본 논문에서는, 효율적인 센서 노드 에너지 소모량 감소를 위한 새로운 데이터 필터링 기법을 제안한다. 제안하는 기법은 수집된 데이터를 분석하여 데이터 필터링의 범위를 설정하고, 이를 전체 센서 노

드에 적용함으로써 예측 가능한 데이터를 효율적으로 필터링하는 기법이다. 아울러, 제안하는 기법은 추가적인 컴포넌트 구성없이, 생존 메시지를 활용하여 센서 노드의 고장과 같은 노드 실패 감지를 지원한다.

논문의 구성은 다음과 같다. 2장에서는 관련 연구에 대해 소개한다. 3장에서는 본 논문에서 제안한 무선 센서 네트워크 기반 모니터링 시스템을 위한 효율적인 데이터 필터링에 대해 설명하고, 4장에서는 제안한 기법의 우수성을 입증하기 위한 성능평가를 수행한다. 마지막 5장에서는 본 논문의 결론 및 향후 연구 방향을 제시한다.

II. 관련 연구

1. 무선 센서 네트워크 기반 모니터링 시스템

무선 센서 네트워크 기술을 이용한 모니터링 시스템은 환경 감시, 의료, 홈 네트워크, 농작물 재배 및 산불 감시 등 다양한 분야에 적용되고 있다. 첫째, 이기욱의 연구[16]에서는 무선 센서 네트워크를 냉동 컨테이너 등의 운송 수단에 적용한 모니터링 시스템을 구축하였다. 구축된 모니터링 시스템은, 냉동 컨테이너를 통해 운송되는 식품의 품질을 감시하고, 이에 대한 관리 서비스를 지원한다. 둘째, 임화정의 연구[17]에서는 무선 센서 네트워크를 도로, 교각 등의 구조물에 적용한 구조물 안전 모니터링 시스템을 구축하였다. 이를 통해 도로에 걸리는 하중, 진동, 외부 충격 등의 부하를 감지하여 실시간으로 사용자에게 알려준다. 특히 기존의 유선 센서 기반의 도로 모니터링 시스템의 경우, 시스템 구축 혹은 확장 시 인프라 설치를 위한 도로 공사에 많은 시간과 비용이 소모되는 문제점이 존재한다. 이를 무선 센서 네트워크를 활용함으로써, 소모되는 시간 및 자원을 절감하였다.

이러한 무선 센서 네트워크 기반 모니터링 시스템은 첫째, 시스템 구축이 용이하고, 둘째, 설치 비용을 감소시키며, 마지막으로 확장이 용이한 장점이 존재한다.

2. 데이터 필터링

데이터의 전송 횟수를 감소시키는 연구에는 이전 데이터와의 오차 범위를 고려하여 데이터의 전송 여부를 결정하는 데이터 필터링(data filtering) 기법이 대표적이다[8-13]. 센서 데이터에 기반하는 많은 응용들은, 근사 결과(approximate answer)를 오차 한도 내에서 허용한다. 예를 들어, 실제 온도가 10.5°C라고 할 때, 동일 기종의 센서 A와 B를 통해 온도를 측정한다고 가정하자. 이때, 센서 노드의 물리적 특성상 오차가 발생하기 때문에, A와 B는 동일 기종임에도 각각 10.4°C, 10.6°C를 온도 값으로 측정할 수 있다. 그러나 일반적으로 10.4°C와 10.6°C는 크게 차이가 없으며, 사용자는 센서의 물리적 특성을 고려하고 있기 때문에, 실제 온도가 10.4°C일 때 온도 데이터 값이 10.6°C라고 해서 크게 문제가 되지는 않는다. 이와 같은 특징을 이용해서 센서 노드에서의 메시지 전송 횟수를 감소시키는 방법이 데이터 필터링이다. 데이터 필터링 기법은 어떤 센서가 새롭게 측정한 데이터 값과 이전까지의 데이터를 기반으로 예측한 데이터 값과의 차이가 미리 정한 에러 범위 안에 있을 경우, 새로 측정된 데이터를 전송하지 않는다. 이때 해당 노드의 부모 노드는 그 자식 노드가 이전에 보냈던 데이터를 저장하고 있으며, 이를 이용하여 필요한 연산 처리를 수행한다.

한편, 데이터 필터링의 효율성을 향상시키기 위해서는, 보다 높은 정확도를 지닌 데이터 예측값을 측정해야 한다. 이를 위해 선형 필터 이론에 기반한 다양한 적응적 데이터 필터링 기법들이 제안되었다[10-13]. 선형 필터 기반의 적응적 필터링 기법의 대표적인 선형 필터로는 칼만(Kalman) 필터가 존재한다. 칼만 필터의 데이터 예측은 다음 식 (1) 및 (2)에 의해 계산된다.

$$avg_{update} = \frac{avg_{prev} * SensingError^2 + Data * (sd_{prev})^2}{sd_{prev}^2 + SensingError^2} \quad (1)$$

$$sd_{update} = \frac{sd_{prev}^2 + SensingError^2}{sd_{prev}^2 * SensingError^2} \quad (2)$$

여기서, avg_update는 현재 측정값을 이용하여 예측

한 데이터의 평균값, avg_prev는 이전에 예측한 데이터 평균값, 그리고 Data는 현재 감지한 센서값을 의미한다. 또한 SensingError는 실제 센서에서 데이터 측정시 발생하는 에러를 의미하며, Sd_prev는 이전에 예측한 데이터의 표준편차를 의미한다. 예를 들어, avg_prev = 40, sd_prev = 1, Data = 41, SensingError = 0.5, 데이터 필터링 한계 오차율 = 3% 라고 가정하면, 식 (1)과 (2)에 의해 예측값 avg_update = 40.8, sd_update = 5가 된다. 따라서 예측값 40.8과 현재 측정값 41 사이의 예측 오차는 (41-40.8)/40.8 = 0.0049, 즉 약 0.5% 정도의 오차율을 나타낸다. 오차율이 데이터 필터링 한계 오차율 3%보다 작기 때문에, 현재 데이터는 전송하지 않는다. 칼만 필터의 높은 예측 정확도 때문에, 데이터 필터링이 필요한 다양한 응용 분야에서 활용되고 있다.

한편, 장수민의 연구[18]에서는 무선 센서 네트워크에서 스카이라인 질의처리를 위한 데이터 필터링 기법을 제안하였다. 스카이라인 질의는 전체 객체 집합에서 대상 객체의 여러 속성을 다른 객체가 지배(Dominate)하지 않는 관심 있는 객체 집합을 검색하는 질의이다. 예를 들어, “해변에서 가깝고 숙박료가 싼 호텔들을 검색하라.”와 같은 질의가 스카이라인 질의이다. 질의의 대상이 되는 각각의 호텔들은, 호텔과 해변과 사이의 거리와 호텔의 숙박료라는 두 가지 속성을 갖고 있다. 이때 주어진 여러 개의 호텔들 간에 상호 비교를 통하여, 특정호텔에 비하여 해변에서 멀면서 숙박료도 비싼 호텔들을 제거한 후, 남아있는 호텔들만을 구하는 것이 스카이라인 질의 처리이다. 센서 네트워크에서 스카이라인 질의처리 수행시, 불필요한 False Positive 데이터의 전송에 의한 데이터 전송량이 증가한다. 따라서 이를 해결하기 위해서, 클러스터 기반의 데이터 필터링 기법인 ECF(Excellent Cluster Filter)기법을 제안하였다. ECF 기법은 첫째, 전체적인 센서들의 데이터를 비교하여 우수 클러스터를 선정한다. 둘째, 우수 클러스터에 해당하는 일부분의 센서들만 탐지하여 최적화된 필터를 생성한다. 마지막으로 생성된 필터를 전체 영역의 센서들에게 적용함으로써 불필요한 센서 데이터가 싱크노드(BS)에 전송되는 것을 최소화한다. ECF 기법은 다른 클러스터에 비해 필터링 영역이 큰 우수 클러스터

를 기준으로 데이터 필터링을 생성하기 때문에, 불필요한 False Positive 데이터의 전송량을 크게 감소시킨다.

III. 무선 센서 네트워크 기반 모니터링 시스템을 위한 효율적인 데이터 필터링 기법

1. 기존 데이터 필터링 기법의 문제점

무선 센서 네트워크 기반의 모니터링 시스템을 구축하기 위해서는, 다음과 같은 사항들을 고려해야 한다. 첫째, 센서 노드는 배터리와 같은 제한적인 에너지 자원을 이용하기 때문에, 효율적으로 에너지 소모량을 감소시킬 수 있어야 한다. 센서에서 가장 많은 에너지를 소모하는 작업은 데이터 전송 작업이다. 따라서 예측 가능한 데이터 데이터의 전송 횟수를 감소시킬 수 있는 효율적인 데이터 필터링 기법이 필수적이다. 둘째, 고장이나 배터리 소모 등에 의해 노드 실패 발생시, 이에 대한 확인을 지원해야 한다. 모니터링 시스템은 설치된 지역의 환경 정보를 지속적으로 수집하는 것이 주목적이다. 그러나 센서 노드 실패시 해당 노드 담당 지역에서 환경 정보를 수집할 수 없기 때문에, 센서 노드 실패를 사용자에게 알려줌으로써 시스템의 유지 및 보수를 용이하게 해야 한다. 따라서 일정 주기마다 생존 메시지를 전송하여 무선 센서 노드의 실패 여부를 확인하는 것이 필수적이다. 따라서 이러한 고려사항에 근거하여, 기존 데이터 필터링 기법을 분석하여 문제점을 살펴본다.

기존 데이터 필터링 기법 중 대표적인 칼만 필터의 경우, 센서 노드에서 자체적인 필터링 범위 조정이 가능하기 때문에, 필터링 범위 조절을 위한 추가적인 메시지 전송을 감소시킬 수 있다. 또한 데이터 예측 정확도가 높기 때문에, 전송되는 데이터의 수가 매우 적다. 그러나 칼만 필터의 경우, 다음과 같은 단점들이 존재한다.

첫째, 데이터 예측에 필요한 계산 오버헤드가 크다. 데이터 예측을 위한 오버헤드는 row data 변환을 위한 계산 오버헤드, 데이터 예측값을 위한 계산 오버헤드로 구성된다. 먼저 실제 센서 노드에서 감지되는 데이터는 row data 형태이기 때문에, 데이터 예측의 정확도를 향상시키기 위해서는, 매 주기마다 row data를 real 값으

로 변환해야 한다. row data 변환식은 사용되는 센서의 종류에 의해 결정되기 때문에, 이기종 센서 기반의 모니터링 시스템 구축 시 센서 노드 종류에 따른 row data 변환 코드 구성이 필요하다. 따라서 각각의 센서 노드는 자신의 row data 변환 코드 이외에, 불필요한 다른 센서 노드의 row data 변환 코드를 저장해야하는 단점이 존재한다. 다음, 데이터 예측값을 계산하기 위해서는 식 (1) 과 (2)를 통해 소수점 이하의 값에 대한 곱셈/ 나눗셈 연산이 요구된다. 센서 노드가 일반적인 PC 에 비해 제한된 CPU 자원을 지닌 점을 고려하면, 보다 적은 자원을 소모하는 다른 연산으로 대체하는 것이 바람직하다.

둘째, 칼만 필터는 이기종 센서 사용시 시스템 구축 및 확장이 어려운 단점이 존재한다. 예를 들어 MTS 420 센서보드(Crossbow)의 온도 센서는 0.5, 습도 센서는 0.03의 오차율을 지니는 반면, TelosB는 온도 센서 및 습도 센서 모두 0.1의 오차율을 지니기 때문에, 센서 종류에 따른 SensingError 값의 변경이 요구되며, 이는 센서 노드에 저장되는 컴포넌트의 크기를 증가시킨다. 따라서 데이터 필터링 기법이 실제 모니터링 시스템에서 활용되기 위해서는, 이러한 문제점들을 해결해야 한다.

2. 제안하는 데이터 필터링 기법

기존 칼만 데이터 필터링 기법의 문제점을 해결하기 위해, 본 연구에서는 새로운 데이터 필터링 기법을 제안한다. 제안하는 데이터 필터링 기법은 센서 노드 생존 여부 확인을 위해 주기적으로 서버에 전송하고, 집계된 데이터 집합의 특성을 분석하여, 필터링 범위를 설정한다. 즉, 필터링 범위 설정을 위한 데이터 집계 및 분석, 분석된 데이터 필터링 범위에 대한 real data의 row data 변환, 필터링 범위 전송은 모두 자원의 제약이 없는 서버에서 이루어진다. 따라서 필터링 범위 설정을 통한 데이터 필터링은 다음과 같은 장점이 존재한다. 첫째, 데이터 예측값 계산을 위한 연산 오버헤드가 감소된다. 즉, 서버에서 데이터의 통계 분석 및 데이터 필터링 범위 설정 작업을 수행하기 때문에, 센서 노드에서는 전송된 데이터 필터링 범위를 바탕으로 단순 비

교 연산을 수행하여 데이터의 필터링 여부를 확인할 수 있다. 따라서 계산 오버헤드가 감소한다. 둘째, 센서 종류에 의해 결정되는 변수들을 서버에서 처리하기 때문에 이기종 센서 노드 활용이 보다 용이하다. 기존 칼만 필터 기법의 경우, 센서 종류에 따라 결정되는 row data 변환식, SensorError 등을 센서 노드에서 전부 처리하기 때문에, 필터링 컴포넌트의 크기가 증가하는 단점이 존재한다. 또한 이기종의 센서 노드가 추가될 경우, 필터링 컴포넌트 수정 및 수정된 컴포넌트의 재전송 작업이 요구된다. 그러나 제안한 기법은 해당 작업들을 서버에서 처리하기 때문에, 센서 노드 추가에 따른 컴포넌트 재전송 작업을 최소화 할 수 있다.

본 연구에서는 효율적인 데이터 필터링을 위한 필터링 범위 설정을 위해 데이터의 평균 및 표준 편차를 활용한다. 데이터 필터링을 위한 수행 과정은 필터링 범위 초기화(1번째 줄 ~ 7번째 줄), 데이터 필터링 수행(8번째 줄 ~ 11번째 줄), 그리고 필터링 범위 재설정(12번째 줄 ~ 16번째 줄)의 세 단계로 이루어진다. 전체적인 알고리즘은 [그림 1]과 같다.

```

Data Statical Analysis(
//In each sensor
01) while(I_init){
02)  Data = Sensing(Type of Sensor);
03)  Send(Data);
//In Server
04) while(I_init){
05)  Sample_Data_list = list.add(Data); }
06) F_Range = S_Analysis(Sample_Data_list);
07) Broadcast(F_Range);
//In each sensor
08) Set(F_Range);
09) for every Sensing period {
10) Data = Sensing(Type of Sensor);
11) if(!ID_filter(F_Range)) then Send(Data); }
//In Server
12) for every Survival_MSG_period(I_Reset)
13) do Step from (01) to (03);
14) New_F_Range =
    Re_Analysis(New_sample_list);
15) Broadcast(New_F_Range);
//In each sensor
16) do Step from (08) to (11)
    
```

그림 1. 데이터 통계 분석 기반 필터링 알고리즘

첫째, 필터링 범위 초기화 단계에서는 충분한 양의 표본 데이터를 수집하고(1번째 줄~3번째 줄), 이를 분석하여 전체 데이터의 분포도를 예측한다(4번째 줄~6번째 줄). 이 후 전체 데이터 분포도를 통해 필터링 범위 F_Range를 설정하고, 브로드캐스트를 통해 모든 센서노드에 필터링 범위를 전송한다. 필터링 범위 초기화를 위해 표본 데이터의 수집 및 분석을 위한 매개변수를 설정해야 한다. [표 1]은 이를 위한 기호를 나타낸다.

표 1. 표본 데이터 수집 및 분석 관련 표기

기호	의미
I_{init}	표본 데이터 수집 주기
μ_{init}	표본 데이터 평균
σ_{init}	표본 데이터 표준편차
dn_{init}	표본 데이터 수
N	전체 노드 수
$Data_i$	i 번째 수집된 데이터

$$dn_{init} = I_{init} * N \tag{3}$$

$$\mu_{init} = \frac{1}{dn_{init}} * \sum_{i=0}^{dn_{init}} Data_i \tag{4}$$

$$\sigma_{init} = \sqrt{\frac{1}{dn_{init}} * \sum_{i=0}^{dn_{init}} (Data_i - \mu_{init})^2} \tag{5}$$

식 (3), (4), (5)에 의해 수집된 표본 데이터의 평균 및 표준 편차를 계산할 수 있다. 이 때, 만약 표본 데이터 수 dn_{init} 이 충분히 큰 수라면, 표본 데이터의 분포는 정규분포를 따른다고 할 수 있다. 반면, dn_{init} 이 작을 경우, 표본 데이터의 분포는 정규 분포에 비해 넓게 분포된 T-분포를 따른다고 할 수 있다. 표본 데이터 수집 주기 I_{init} 이 후에 수집되는 데이터들이 현재 분석된 표본 데이터와 유사한 경향을 나타낼 경우, 이는 예측가능한 데이터이므로 필터링이 가능하다. 이를 기반으로 필터링 범위를 초기화할 수 있다. [표 2]는 필터링 범위 초기화에 필요한 기호를 나타낸다.

표 2. 초기 필터링 범위 설정 관련 표기

기호	의미
CI	데이터 필터링 예측 범위
Z	정규분포 확률변수
T	T-분포 확률 변수

$$\mu_{init} - T^* \frac{\sigma_{init}}{\sqrt{dn_{init}}} \leq Data \leq \mu_{init} + T^* \frac{\sigma_{init}}{\sqrt{dn_{init}}} \quad (6)$$

(where, $dn_{init} \leq 30$)

$$\mu_{init} - Z^* \frac{\sigma_{init}}{\sqrt{dn_{init}}} \leq Data \leq \mu_{init} + Z^* \frac{\sigma_{init}}{\sqrt{dn_{init}}} \quad (7)$$

(where, $dn_{init} > 30$)

식 (6)와 (7)는 초기 필터링 범위를 나타낸다. 일반적으로 $dn_{init} > 30$ 일 경우 정규분포를 사용해도 무방하기 때문에, 범위를 결정하는데 정규분포 확률변수 Z를 사용한다. 확률변수 Z나 T 는 CI에 의해서 결정된다. CI(Confidential Interval)는 수집된 데이터가 초기 표본 데이터를 기반으로 설정된 범위 내에 포함될 확률에 대한 예측값이다. 만약 CI = 0.8일 경우, linit 이후에 수집되는 데이터들은 약 80% 확률로 식 (4)나 (5)의 범위에 속할 것으로 예측된다는 의미이다.

둘째, 필터링 수행 단계에서는 주어진 필터링 범위를 설정하고(8번째 줄), 필터링 범위 밖에 존재하는 데이터를 감지할 경우, 이를 상위 노드로 전송한다(9번째 줄 ~ 10번째 줄). 이때, 기존 칼만 필터링과 달리 곱셈/나눗셈 연산을 수행하지 않고 단순 비교연산만을 수행하기 때문에, 센서 노드에서의 계산 오버헤드가 감소한다.

셋째, 필터링 범위 초기화 이후, 센서 노드 데이터의 변화에 의해 설정된 필터링 범위 이외의 데이터가 빈번하게 발생할 수 있다. 이는 예측한 데이터의 특성이 변경되었음을 의미하며, 따라서 변경된 데이터의 특성을 반영하여 필터링 범위를 재설정할 필요가 있다. 이를 위해 필터링 범위 재설정 단계에서는, 생존 메시지 전송 주기 L_{reset} 마다 N개의 노드로부터 N개의 데이터를 수집하여(12번째 줄~13번째 줄), 새로운 필터링 범위를 측정한다(14번째 줄). 다음, 측정된 새로운 필터링 범위를 모든 센서 노드에 브로드캐스팅한다(15번째 줄). 새로운 필터링 범위 측정을 위해 재설정 표본 평균

μ_{reset} 및 재설정 표본 표준편차 σ_{reset} 를 계산하고, 수정된 전체 표본()의 평균을 식(8)을 사용하여 계산한다.

$$\mu' = \frac{\mu_{init} * dn_{init} + \mu_{reset} * N}{dn_{init} + N}, \quad (8)$$

(after calculate μ' , $dn_{init} = dn_{init} + N$, $\mu_{init} = \mu'$)

식 (8)은 재설정 주기에 따라 데이터를 수집한 후, 수정된 표본 평균을 구하는 식이다. 예를 들어 $dn_{init} = 100$, $\mu_{init} = 15$, $N = 50$, $\mu_{reset} = 17$ 이라 하면, 식 (8)에 의해, 새로운 표본 평균 $\mu' = 15.7$ 이 된다.

$$\sigma' = \sqrt{\frac{\sigma_{init}^2}{dn_{init}} + \frac{\sigma_{reset}^2}{N}} \quad (where, dn_{init} > 30, N > 30) \quad (9)$$

$$\sigma' = \frac{(dn_{init} - 1)\sigma_{init}^2 + (n - 1)\sigma_{reset}^2}{dn_{init} + N - 2} * \sqrt{\frac{1}{dn_{init}} + \frac{1}{N}} \quad (10)$$

식 (9), (10)은 재설정 주기에 따라 데이터를 수집한 후, 수정된 표준 편차를 구하는 식이다.

$$\mu' - T^* \frac{\sigma'}{\sqrt{dn_{init} + N}} \leq Data \leq \mu' + T^* \frac{\sigma'}{\sqrt{dn_{init} + N}} \quad (where, dn_{init} + N \leq 30) \quad (11)$$

$$\mu' - Z^* \frac{\sigma'}{\sqrt{dn_{init} + N}} \leq Data \leq \mu' + Z^* \frac{\sigma'}{\sqrt{dn_{init} + N}} \quad (where, dn_{init} + N > 30) \quad (12)$$

마지막으로 앞에서 계산한 μ' 및 σ' 를 식 (11), (12)에 대입하여, 재설정된 필터링 범위를 측정한다.

예를 들어, 초기 필터링 범위 설정 시, $dn_{init} = 100$, CI = 0.8, 표본 평균 $\mu_{init} = 15$, 표본 표준편차 $\sigma_{init} = 15$ 라고 가정한다. dn_{init} 이 충분히 크므로 정규분포를 따른다고 할 수 있으며, CI에 의해 $Z = 1.28$ 이다. 따라서 식 (5)에 의해 초기 필터링 범위의 최소값 = 13.5, 최대

값 = 16.5 이며, 초기 데이터 필터링 범위는 다음과 같다.

$$13.5 \leq \text{Data} \leq 16.5$$

만약 어떠한 센서 노드에서 측정된 데이터가 11일 경우, 이는 필터링 범위에 속하지 않기 때문에, 상위 노드로 전송된다. 다음, 일정 주기 L_{reset} 이 지난 후, 새롭게 필터링 범위를 재설정한다. 이때, 새로 수집하는 데이터의 수, $N = 50$ 이고, N 개의 데이터에 대한 표본 평균 및 표준 편차는 각각 $\mu_{\text{reset}} = 17, \sigma_{\text{reset}} = 10$ 이라 가정하자. $dn_{\text{init}} > 30, N > 30$ 이므로, 식 (8)과 (9)에 의해 $\mu' = 15.7, \sigma' = 0.592$ 이다. 따라서 식 (12)에 의해 재설정되는 필터링 범위는 다음과 같다.

$$15.64 \leq \text{Data} \leq 15.76$$

즉, 필터링 범위 재설정 이후, 데이터의 약 80%가 위의 범위에 포함되어 필터링될 것으로 예측된다.

IV. 성능 평가 및 고찰

제안된 데이터 분석 기반 데이터 필터링 기법(OFM : 성능평가시 Our Filtering Method 이라 명명)을 기존의 적응적 필터링 기법 중, 가장 우수한 칼만 필터링 기법과 성능 평가를 수행한다. ECF의 경우, 처리하는 질의 종류가 다르기 때문에 비교대상에서 제외하였다. 칼만 필터의 데이터 오차율 설정을 위해, 실제 센서를 활용하여 환경 데이터를 수집하고 분석하였으며, 그 결과 최적의 데이터 오차율을 3%로 설정하였다. 또한 칼만 필터는, 주기적인 노드 생존 메시지 전송을 하지 않기 때문에, 노드 생존 여부를 확인할 수 없다. 그러나 실제 모니터링 시스템 구축시 노드 생존 여부 확인에 따른 신속한 유지/보수가 필수적이기 때문에, 칼만 필터에도 주기적인 생존 메시지 전송 모듈을 구현하였다. 실험은 TOSSIM 시뮬레이터를 이용하여 수행되었으며, TinyOS 기반의 MintRoute를 사용하여 집계 트리를 구성하였다. 센서 노드는 매 주기마다 0.5%의 확률로 작동을 정지하도록 설정하였으며, 총 200주기 동안 데이터를 생성한다. 모든 실험은 100회 수행하여 그 평균을 결과값으로 측정하였다. 한편, 성능평가를 위해, 이전 데이터를 기준으로 주어진 변화율 범위 내에서 증

가 혹은 감소하는 선형적 데이터 생성기를 구현하여 사용하였다. 예를 들어, 첫번째 데이터가 10, 데이터 변화율이 10%의 경우, 두번째 데이터는 9~11 사이의 범위에서 생성된다. 만약 두번째 생성된 데이터가 9일 경우, 세 번째 데이터는 8.1~9.9 사이에서 생성된다. 단, 데이터 생성기를 통해 생성된 모든 데이터는 real data가 아닌 row data이다. 따라서 칼만 필터의 경우, 생성된 row data를 real data로 변환시키기 위한 row data 변환식 부분을 추가하였다.

1. 성능 평가

[그림 2]는 제안하는 OFM 필터링 기법에서 $CI=0.95$, 노드 수 = 20, 데이터 변화율 = 4%일 때, 노드 생존 메시지 전송 주기 변화에 따른 데이터 전송율을 나타낸다. 이때 전체 데이터 전송횟수는 주기별 노드 생존 메시지 전송횟수와 필터링되지 않은 데이터의 전송횟수를 더한 값이다. 전체 생성된 데이터 량은 노드수*실험 주기 이므로, $20*200 = 4,000$ 개의 데이터가 생성되었다. 아울러 상위 노드에서의 라우팅을 위한 단순 메시지 전송횟수를 고려할 때, 필터링 미적용시 전체 데이터 전송횟수는 전체 생성된 데이터 량 * 트리 깊이이다. 따라서 트리의 평균 깊이가 3일 때, 필터링 기법 미적용시 데이터 전송량은, $3*4,000=12,000$ 이다. [그림 1]에서 전체 노드별 생성 데이터 수 데이터는 데이터 전송횟수는 주기가 증가함에 따라 감소하는 경향을 나타낸다. 주기가 짧을수록 데이터 분석에 의한 필터링 범위 설정이 정확하기 때문에 필터링 되는 데이터의 량이 증가한다. 반대로 주기가 길수록 필터링 범위 설정의 정확도가 감소하기 때문에, 필터링 되는 데이터의 량이 감소한다. 따라서 노드 생존 메시지 전송 주기 변화에 큰 영향없이, 평균 데이터 전송횟수가 유사하다. 이를 통해 최적의 노드 생존 메시지 전송 주기를 40주기로 설정한다.

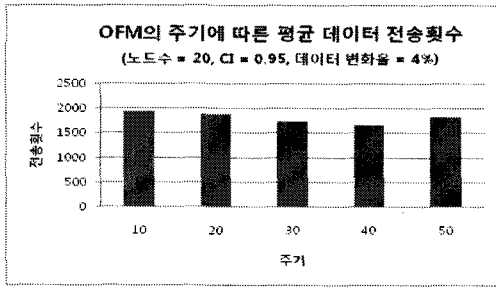


그림 2. 주기에 따른 평균 데이터 전송횟수

[그림 3]은 OFM에서 노드 생존 메시지 전송 주기 = 40, 노드 수 = 20, 데이터 변화율 = 4%일 때, CI 변화에 따른 데이터 전송율을 나타낸다. 단, 전체 데이터 전송 횟수가 통계적으로 충분히 큰 수이기 때문에, 데이터의 분포는 정규분포를 따른다고 가정한다. 실험 결과, CI가 증가할수록 데이터 전송 횟수가 감소한다. 이는 CI가 증가할수록 데이터 필터링 범위가 커지고, 따라서 필터링되는 데이터의 양이 증가하기 때문이다. 이를 통해 OFM을 위한 최적의 CI를 0.95로 설정한다.

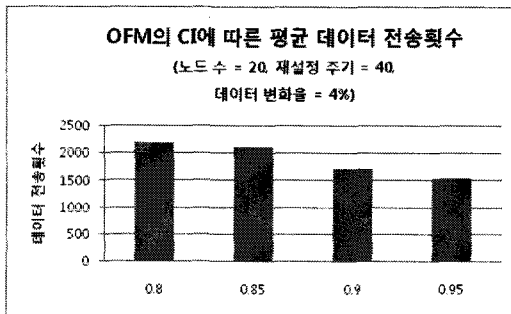


그림 3. CI에 따른 평균 데이터 전송횟수

[그림 4]는 노드 생존 메시지 전송 주기 = 40, 데이터 변화율 = 4% 일 때, 노드 수 변화에 따른 평균 데이터 전송횟수를 나타낸다. 또한 OFM의 CI=0.95이다. 칼만 필터링의 경우, 노드 수 증가에 따라 2262, 2371, 2665 회의 평균 데이터 전송 횟수를 나타내며, OFM의 경우, 1931, 2164, 2511의 평균 데이터 전송횟수를 나타낸다. 실험결과 노드 수 증가에 따라 데이터 전송횟수가 증가하는데, 이는 노드 수 증가에 의해 트리의 깊이가 증가하기 때문이다. 아울러, OFM의 데이터 전송 횟수가 칼

만 필터링에 비해 약 6%~15% 정도 감소하여, 필터링 성능이 향상되었음을 알 수 있다. 이는 칼만 필터링의 경우, row data를 real data로 변환시키는 과정 및 예측 데이터 계산 과정에서 오차가 발생하여 데이터 필터링 성능이 저하되었기 때문이다.

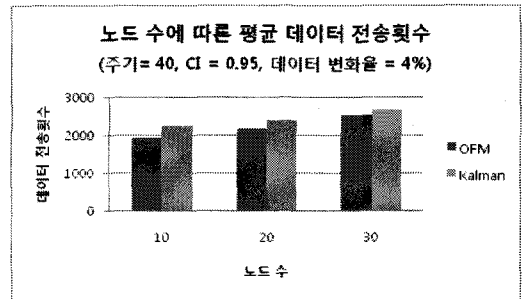


그림 4. 노드 수 증가에 따른 평균 데이터 전송횟수

[그림 5]는 노드 수 = 20, 노드 생존 메시지 전송 주기 = 40, CI = 0.95 일 때, OFM 및 칼만 필터링에서의 데이터 변화율에 의한 평균 데이터 전송횟수를 나타낸다. OFM과 칼만 필터링 기법 모두, 데이터 변화율 증가시 데이터 전송 횟수가 증가하였다. 이는 데이터 변화율 증가시, 분석된 데이터 필터링 범위(OFM) 또는 예측된 데이터 값(칼만)의 정확도가 감소하기 때문이다. 한편, OFM의 경우 칼만 필터에 비해 약 5%~10% 정도 데이터 전송횟수가 감소되었다. 이는 칼만 필터링 기법의 경우, 데이터 예측시 오차가 발생하여 필터링되는 데이터 양이 감소하기 때문이다.

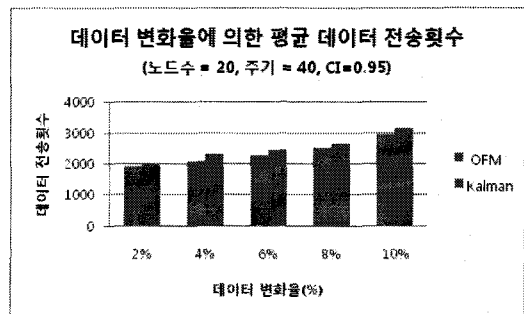


그림 5. 데이터 변화율에 따른 평균 데이터 전송횟수

2. 성능 고찰

성능 평가 결과, 제안한 기법은 칼만 필터 기법에 비해, 센서 노드 수 증가에 의한 평균 메시지 전송량을 약 6%~15% 정도 감소시킨다. 아울러 데이터 변화에 따른 평균 메시지 전송량을 약 5%~10% 정도 감소시킨다. 칼만 필터 기법의 경우, row data 변환 및 데이터 예측 값 계산작업을 제한된 CPU 자원을 지니는 센서 노드에서 수행하여 많은 오차가 발생하기 때문에, 필터링 성능이 저하된다. 그러나 제안한 기법은 많은 연산을 요구하는 작업들을 서버에서 수행하고, 센서 노드에서는 단순 비교연산만을 이용하여 데이터 필터링을 수행하기 때문에, 계산 오버헤드 및 오차를 감소시켜 데이터 필터링 성능을 향상시킨다.

한편, 제안하는 기법은 전체 센서 노드에서 수집한 샘플 데이터를 기반으로 필터링 범위를 설정하기 때문에, 지역성을 고려하지 못하는 문제점을 지닌다. 예를 들어, N주기 동안 A, B 지역의 평균온도는 각각 23도, 20도이며, 두 지역 모두 온도의 표준편차는 10이고, 이때 필터링 범위 설정을 위한 $CI = 0.8$ 이라고 가정하자. 지역적인 특성을 고려할 경우, A지역의 필터링 범위는 $22 < Data(A) < 24$, B지역의 필터링 범위는 $19 < Data(B) < 21$ 이다. 그러나 제안한 필터링 기법은 지역적인 특성을 고려하지 않고 있기 때문에, 전체 필터링 범위를 $20.5 < Data < 22.5$ 로 설정하여, 필터링 결과가 다르게 나타난다. 이를 해결하기 위해서, 제안하는 기법과 클러스터링 기법을 결합한, 클러스터별 데이터 필터링 범위 설정 기법에 대한 연구가 요구된다. 이를 통해 지역적 특성을 고려함으로써, 필터링 효율성 향상을 기대할 수 있다. 또한 데이터 필터링 예측범위 CI를 자동적으로 변화시킴으로써, 효과적인 데이터 필터링 범위를 설정하는 기법이 필요하다.

V. 결론 및 향후 연구

본 논문에서는 센서의 효율적인 에너지 소모량 감소를 위한 새로운 데이터 필터링 기법을 제안하였다. 제안한 기법은 데이터 집계 처리와 데이터 필터링을 결합

한 방법으로써, 표본 데이터 및 노드 생존 메시지를 통해 전체 데이터의 특성을 분석하고, 분석된 데이터 특성을 바탕으로 필터링 범위를 설정하여 데이터 전송 여부를 결정한다. 또한 데이터의 분석 및 범위 계산을 서버에서 수행함으로써 센서노드의 데이터 처리 오버헤드를 감소시킨다. 성능 평가 결과, 제안한 데이터 분석 기반 데이터 필터링 기법은 기존 칼만 필터링에 비해 데이터 전송 횟수를 약 5%~10% 감소시켜, 전체적인 성능을 향상시켰다. 아울러 무선 센서 네트워크 기반 모니터링 시스템에서 노드의 생존 여부 확인을 지원하기 위해, 주기적으로 샘플 데이터를 포함한 노드 생존 메시지를 전송하고, 수집된 샘플 데이터를 통해 필터링 범위 설정을 수행한다. 이를 통해 추가적인 컴포넌트의 구성없이 노드 실패 확인을 지원한다. 제안한 데이터 필터링 기법은 데이터 분석을 이용한 필터링 범위 설정을 통해 효율적으로 메시지 전송 횟수를 감소시킨다. 제안한 기법은 기존 칼만 필터에 비해 전체적인 필터링 성능이 향상되었으며, 추가적인 컴포넌트 구성없이 주기적인 노드 생존 메시지 전송을 이용하여, 노드 생존 여부 확인을 지원한다. 따라서 제안한 필터링 기법은 실제 센서 네트워크 모니터링 시스템에서 활용이 가능한 우수한 기법이라 할 수 있다.

향후 연구로는 제안하는 기법과 클러스터링 기법을 결합하여, 클러스터별 데이터 특성을 고려한 필터링 범위 설정 기법에 대한 연구를 수행하는 것이다.

참고 문헌

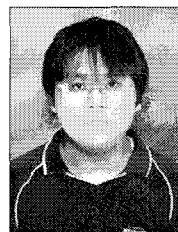
- [1] J. Kahn, "Next century challenges: Mobile networking for 'smart dust,'" In Proceedings of MOBICOM, pp.271-278, 1999.
- [2] A. Chandrakasan, "Power Aware Wireless Microsensor Systems," In Proceedings of ESSCIRC, Florence, Italy, 2002.
- [3] A. Cerpa, "Hibitat Monitoring: Application Driver for Wireless Communications Technology," In Proceedings of SIGCOMM,

- 2001.
- [4] V. Shnyder, "Simulating the Power Consumption of Large-Scale Sensor Network Applications," In Proceedings of ACM SenSys, pp.188-200, 2004.
- [5] S. Madden, "Tag: A Tiny Aggregation Service for ad hoc Sensor Networks," In Proceedings of OSDI, pp.131-146, 2002.
- [6] C. Intanagonwiwat, "Impact of Network Density on Data Aggregatio in Wireless Sensor Networks," In Proceedings of ICDCS, p.457, 2002.
- [7] Y. Yao and J. Gehrke, "The Cougar Approach to In-Network Query Processing in Sensor Networks," SIGMOD Record, Vol.31, No.3, pp.9-18, 2002.
- [8] Antonios Deligiannakis, "Hierarchical In-Network Data Aggregation with Quality Guarantees," In Proceedings of EDBT, pp.658-675, 2004.
- [9] Xingbo Yu, "Approximate Monitoring by Aggregation-Oriented Clustering in Wireless Sensor Networks," ICDE, submitted, Bangalore, India, 2003.
- [10] S. Madden, "The Design of an Acquisitional Query processor for Sensor Networks," In Proceedins of SIGMOD, pp.491-502, San Diego, America, 2003.
- [11] A. Jain, "Adaptive stream resource management using Kalman Filters," In Proceedings of the ACM SIGMOD/PODS Conference (SIGMOD '04), pp.11-22, 2004.
- [12] S. Santini, "An Adaptive Strategy for Quality-Based Data Reduction in Wireless Sensor Networks," Proceedings of the 3rd International Conference on Networked, 2006.
- [13] K. Elleithy, "Decentralized Kalman Filter in Wireless Sensor Networks - Case Studies. Advances in Computer," Information and Systems Sciences and Engineering. pp.61-68, 2005.
- [14] R. Min, "Low-power wireless sensor networks," In Proceedings of VLSI Design, Bangalore, India, 2001.
- [15] K. Kalpakis, "Efficient Algorithms for Maximum Lifetime Data Gathering and Aggregation in Wireless Sensor Networks," UMBC CS TR-02-13, 2002.
- [16] 이기욱, 김정이, "무선 센서 네트워크를 이용한 냉동 컨테이너 모니터링 시스템 설계", 한국컴퓨터정보학회논문지, 제12권, 제5호, pp.312-326, 2007.
- [17] 임화정, 이좌형, 박충명, 정인범, "무선 센서 네트워크 기반의 구조물 안전 감시 시스템", 한국해양정보통신학회논문지, 제12권, 제2호, pp.391-400, 2007.
- [18] 장수민, 강광구, 유재수, "센서데이터의 연속적인 스카이라인 질의 처리를 위한 효율적인 필터링기법", 정보과학회논문지, 제15권, 제12호, pp.938-942, 2009.

저자 소개

이 현 조(Hyun-Jo Lee)

정희원

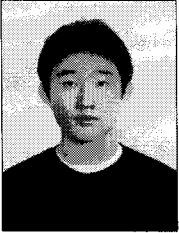


- 2006년 : 전북대학교 컴퓨터공학과(공학사)
- 2008년 : 전북대학교 대학원 컴퓨터공학과(공학석사)
- 2008년 ~ 현재 : 전북대학교 대학원 컴퓨터공학과 박사과정

<관심분야> : 데이터 마이닝, 공간 데이터베이스, 고차원 색인 구조

최 영 호(Young-Ho Choi)

준회원



- 2009년 : 전북대학교 컴퓨터공학과(공학사)
- 2008년 ~ 현재 : 전북대학교 대학원 컴퓨터공학과 석사과정

<관심분야> : 센서 네트워크, LBS, 공간 데이터베이스

장 재 우(Jae-Woo Chang)

정회원



- 1984년 : 서울대학교 전자계산기공학과(공학사)
- 1986년 : 한국과학기술원 전산학과(공학석사)
- 1991년 : 한국과학기술원 전산학과(공학박사)

- 1996년 ~ 1997년 : Univ. of Minnesota, Visiting Scholar
- 2003년 ~ 2004년 : Penn State Univ., Visiting Scholar.
- 1991년 ~ 현재 : 전북대학교 컴퓨터공학과 교수

<관심분야> : 공간 네트워크 데이터베이스, 상황인식, 하부저장구조