

센서 모니터링을 위한 칼만필터 기반의 효율적인 적응적 샘플링 기법

김민기[†] · 민준기^{**}

요 약

센서 네트워크 환경에서는 각 센서는 정의된 샘플링 주기에 따라서 외부 환경을 측정하고 측정된 값을 기지국으로 전송한다. 따라서, 샘플링 주기는 대역폭, 전력량 등 센서들의 중요 자원의 소비에 지대한 영향을 끼친다. 본 논문에서는 측정값 특성에 따라서 센서의 샘플링 주기를 조절하는 새로운 적응적 샘플링 기법을 제안한다. 제안하는 기법은 KF (Kalman-Filter)에 기반하여 미래의 측정값을 예측한다. 그리고, 실측값과 예측값의 차이에 따라서 센서 측정값들의 중요도를 파악하고 이에 따라서 샘플링 주기를 변화시킨다. 실험에서 제안하는 기법의 효과성을 보였다.

키워드 : 무선 센서 네트워크, 적응적 샘플링, 칼만필터

An Efficient Adaptive Sampling Technique based on the Kalman Filter for Sensor Monitoring

Min-Kee Kim[†] · Jun-Ki Min^{**}

ABSTRACT

In sensor network environments, each sensor measures the physical environments according to the sampling period, and transmits a sensor reading to the base station. Thus, the sample period influences against importance resources such as a network bandwidth, and a battery power. In this paper, we propose new adaptive sampling technique that adjusts the sampling period of a sensor with respect to the features of sensor readings. The proposed technique predicts a future readings based on KF (Kalman Filter). By using the differences of actual readings and estimated reading, we identify the importance of sensor readings, and then, we adjust the sampling period according to the importance. In our experiments, we demonstrate the effectiveness of our technique.

Keywords : WSN, Adaptive Sampling, Kalman Filter

1. 서 론

무선 센서 네트워크 (Wireless Sensor Network: WSN)는 제한된 전원을 지닌 수 백에서 수 천개의 센서들로 이루어져 있다. 최근 마이크로 센서 기술과 무선 통신 기술의 발전으로 대량 저비용 센서들로 구성된 무선 센서 네트워크는 군사 보안, 환경 모니터링 등의 다양한 분야에 적용가능하게 되었다. WSN의 규모가 커지면서 기존의 대역폭, 배터리 등의 문제가 점점 더 대두되고 있는 시점이다. 제한된 대역

폭은 데이터 특성, 질의의 조건, 가능한 자원(통신, 전력, 처리장치) 등을 고려하여 효율적으로 할당되어야 한다[1, 2].

최근 무선 센서 네트워크의 비약적인 발전에 힘입어 센서 데이터 모니터링에 관한 연구가 활발히 진행 중이다. 센서 데이터 모니터링은 특정지역에 분포되어 있는 센서들을 이용하여 해당 지역의 정보를 샘플링하여 기지국 (base station)으로 데이터를 전송하여 지속적으로 관찰하는 작업을 말한다. 모니터링을 위한 무선 센서 네트워크는 지속적으로 측정값을 기지국으로 전송함으로써 네트워크 자원을 많이 소비한다. 따라서, 이러한 센서 네트워크의 자원 소비를 최소화하기 위한 많은 연구들이 활발히 진행되고 있다 [3-6, 15, 16].

본 논문에서는 위에 기술한 문제점을 해결하기 위하여 각 센서의 샘플링 대상의 특성에 따라 샘플링 주기를 적응적으

* 이 논문은 2010년 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2010-0016165).

[†] 준 회원 : 한국기술교육대학교 컴퓨터공학과 박사과정

^{**} 종신회원 : 한국기술교육대학교 컴퓨터공학부 조교수

논문접수 : 2009년 9월 8일
수정일 : 1차 2010년 2월 4일
심사완료 : 2010년 2월 26일

로 조정하는 기법을 제안한다. 즉, 샘플링 대상에 특성에 기반하여 데이터 수집 횟수를 조정하여 제한된 네트워크 자원을 효율적으로 사용하도록 하였다.

측정값의 변화가 큰 데이터는 그렇지 못한 데이터 보다 많은 자원을 할당받게 된다. 따라서, 전체적인 네트워크 측면에서 볼 때 중요한 데이터의 수집은 늘리고 정적인 데이터의 수집을 줄임으로써 자원 낭비가 줄어들기 때문에 전력, 대역폭 등을 효과적으로 사용할 수 있다. 추가적으로 DKF (Dual Kalman Filter)[10]를 이용한 근사 데이터 수집 기법과 결합하여 전송 횟수를 감소시키도록 하였다.

또한, 본 논문에서 제안한 기법은 센서 별로 서로 다른 샘플링 주기를 유지함으로써 데이터 전송 시점을 분산시키게 한다. 이에 따라서, 네트워크 대역폭 사용량을 시간에 따라 적절히 분배하는 것과 같은 결과를 가져와 특정 시간에 집중되는 네트워크 대역폭 사용량을 감소시킬 수 있고 네트워크 혼잡 상태가 유발되어 손실되는 데이터의 양을 줄일 수 있다.

본 논문의 특징을 정리하면 다음과 같다.

- 센서에서 샘플링되는 데이터의 특성을 분석하여 중요도에 따라 데이터 우선순위를 결정.
- 데이터 우선순위에 따른 센서의 샘플링 주기를 조정
- DKF기법과 결합한 데이터의 전송 횟수 감소.
- 센서별 샘플링 주기 유지를 통한 대역폭 사용량의 분산

2. 관련연구

스트림 데이터 환경 및 센서 네트워크 환경에서 효율적인 데이터의 전송을 위한 다양한 기법들이 제안되었다 [7-13, 16, 17].

Tatbul[7] 등은 스트림 데이터 환경에서 전송되는 스트림 데이터의 양이 네트워크 대역폭보다 커질 경우, 이를 해결하는 Load shedding 기법을 제안하였다. Load shedding 기법은 스트림 소스에서 데이터를 전송할 때 탈락(drop)되는 횟수 측정하고 측정된 탈락 횟수에 따라 네트워크 상태를 파악한다. 데이터의 탈락 횟수가 많아지면 네트워크 상태가 혼잡하다고 판단한다. 네트워크 혼잡 시 각 스트림 소스들의 데이터 전송 횟수를 같은 비율로 감소시켜 데이터 탈락 횟수를 감소시키고 네트워크 상태를 회복하는 기법이다. Load shedding 기법을 통하여 효과적으로 네트워크 혼잡 상황을 해결 할 수 있었다. 하지만 이 기법은 전체적인 네트워크의 데이터 전송 성공률을 증가 시켰지만 네트워크 상태만을 고려하였기 때문에 임의 탈락(Randomly Drop) 현상과 같은 문제가 발생된다. 또한 Load shedding 기법은 네트워크의 혼잡 상황 발생을 미연에 방지하지 못하고 이미 네트워크 혼잡상황이 발생한 후 동작되는 단점이 있다.

Adaptive Filtering 기법[8]은 스트림 데이터 환경에서 사용되는 기법으로서, 스트림 소스에 precision bound를 적용한다. precision bound는 스트림 소스에서 발생하는 데이터

의 예상 범위를 나타낸다. 스트림 소스는 precision bound를 설정한 후 새로 발생된 데이터 값이 precision bound 안에 있다면 데이터를 전송하지 않는다. 스트림 처리기는 이전의 전송받은 스트림 데이터를 근사데이터로 간주한다. precision bound 밖의 데이터 발생 시에는 스트림 처리기로 데이터를 전송하고 precision bound 재설정 하는 방식이다. 스트림 소스에서 발생하는 데이터가 precision bound를 적절히 만족한다면 전송되는 데이터의 양의 상당량 감소하는 장점이 있다. 그러나, Adaptive Filtering 기법에서는 스트림 데이터 전송 주기에 대한 내용이 고려되어 있지 않다. 또한 데이터 값이 지속적으로 변동할 경우 precision bound 재설정과 데이터 전송 횟수가 증가하는 단점이 있다.

센서 네트워크에서 네트워크 비용을 줄이기 위하여 사용되는 단순한 방법은 시간 축약 기법(Temporal suppression)[11]이다. 시간 축약 기법은 현재의 측정값과 가장 최근에 전송한 측정값을 비교하여 측정값을 전송 여부를 결정한다. 이때, 센서가 기지국으로 데이터를 전송하지 않는 경우 이전과 변화가 없다고 간주할 수 있다. [11]에서 제안하는 시간 축약 기법은 센서에서 취득되는 데이터의 오차 범위를 미리 정하고, 수집한 데이터의 변화가 오차 범위보다 큰 경우만 데이터를 기지국으로 전송한다. 시간 축약 기법은 센서에서 샘플링되어 취득되는 데이터의 변화가 적을 경우에서 전송되는 데이터의 양을 상당히 줄일 수 있지만, 취득되는 데이터의 변화가 크게 되면 기존의 일반적인 샘플링 기법과 비슷한 성능을 낸다는 단점이 있다.

센서 네트워크에서는 측정값의 전송을 최소화하기 위한 방법으로 확률 모델을 이용한 근사 근사 데이터 취득 기법들 [9, 10]이 제안되었다. 이러한 기법들은 예측 모델을 사용하여 데이터를 예측하고 실제 측정값과 예측값의 차이가 크지 않다면 센서는 데이터를 전송하지 않으며, 기지국은 모델을 이용한 예측값을 근사 데이터로 간주하는 방식이다.

PAQ [9]는 통계에서 많이 쓰이는 AR(Autoregressive) 기법에 기반한 예측 모델을 사용하여 센서의 값을 예측한다. 센서에 예측 범위를 설정한 후 범위를 만족 할 때와 만족할지 못할 때의 경우로 구분한다. 측정값이 예측 범위를 만족하지 못할 경우에, 기지국으로 데이터를 전송하는 방식이다. 예측이 잘 이루어지고 있는 상황에서는 데이터 전송비용을 크게 감소시킬 수 있다. 그러나 PAQ는 AR 기법의 예측 모델을 생성하기 위하여 긴 시간의 학습단계가 필요하다.

Jain 등은 예측 모델로써 Kalman Filter(KF)를 사용하는 기법을 제안하였다 [10]. KF는 현재의 시스템 변수와 가장 최근의 측정값만을 이용하여 미래의 값을 예측할 수 있기 때문에 [7]과 비교하여 학습과 메모리 비용 적다.

일반적으로 센서의 샘플링 주기는 데이터 전송 주기와 동일하다. 따라서 본 논문에서 제안하고자 하는 기법과 같이 샘플링 주기를 조정함으로써 샘플링 비용 및 네트워크 전송 비용을 감소시키고자 하는 연구들[12, 13]이 있었다.

[12]은 선형 예측 모델을 사용하여 미래의 값을 예측한다. [12]에서 센서는 예측 모델을 사용하여 예측한 값과 실제 측

정한 값의 차이인 오차를 구한다. 오차가 큰 경우는 예측하지 못한 데이터의 발생으로 좀 더 많은 데이터의 수집이 필요하다고 간주한다. 따라서 오차가 큰 경우에는 샘플링 주기를 감소시키며, 오차가 작을 경우에는 샘플링 주기를 증가시킨다. 측정되는 데이터의 형태가 선형과 유사하다면 측정값을 적절히 반영하여 다음의 값을 적절히 측정할 수 있다.

Data centric modeling [13]에서는 센서의 Sleep 주기를 조정하는 방식을 사용한다. 센서는 현재 측정된 값과 가장 최근에 측정된 값의 오차를 이용하여 오차율을 계산한다. 사용자는 오차율 허용치를 설정하여 오차율이 오차 허용치보다 크거나 작은 경우로 구분한다. 오차율이 오차 허용치보다 크면 센서의 기존의 Sleep 주기를 1/2로 감소시키고 오차율이 오차 허용치보다 작으면 기존의 Sleep 주기를 2배 증가시킨다. Data centric modeling 기법은 Sleep 주기를 조절을 통해 샘플링 주기를 조절하는 것과 동일한 효과를 얻게 된다. 또한 에너지 소비를 크게 감소시켜 에너지 효율성을 높였다. Data centric modeling에서는 Sleep 주기의 변화의 제한폭이 제시되지 않아서, Sleep 주기가 급격하게 늘어나 데이터 수집량이 크게 줄어, 수집한 데이터의 신뢰성이 떨어지는 단점이 있다.

3. 배경지식

본 절에서는 본 연구의 기반이 되는 KF와 이를 센서 네트워크 환경에 적용한 DKF에 대하여 간단히 설명하도록 하겠다.

KF는 연속 데이터의 선형 필터링 문제의 재귀적 해법으로 1960년에 R. E. Kalman [14]의하여 소개되었다. KF는 예측/수정의 두 개의 패러다임(paradigm)의 기반을 둔 내부 상태(Internal State) 예측 선형 알고리즘이다. KF에서 사용되는 기호들에 대한 설명은 다음의 <표 1>에 요약하였다.

<표 1> 수식의 기호

기호	의미
K	시간 인덱스(Index)
x_k	처리 상태 벡터(Vector)
A	x_k 에서 x_{k+1} 으로의 상태 전위 매트릭스(Matrix)
w_k	시스템 모델 노이즈(Noise)
z_k	실제 측정 값 벡터
H	시스템 상태와 측정 벡터 관련 매트릭스
v_k	실제 측정값의 노이즈
\hat{x}_k^-	k시간 이전에 예측한 내부 변수
\hat{x}_k^+	k시간 이후에 예측을 위해 만든 내부 변수
θ_k	k시간의 예측 가감변수
K	Kalman Gain(각각의 측정치에 가중을 준다)

KF는 다음과 같은 수학 수식식으로 (1), (2)로 구성되어 있다. 수식 (1)은 실제 측정 값 z_k 와 KF 시스템 상태 벡터 x_k 와의 관계를 나타내는 것으로 H는 시스템 상태 벡터와 측정값의 연관 매트릭스이다. 수식 (2)는 시간의 따른 KF의 시스템 상태 벡터들의 상관관계를 나타낸 것이다.

$$z_k = Hx_k + v_k \tag{1}$$

$$x_{k+1} = Ax_k + w_k \tag{2}$$

일반적인 KF의 예측과 수정의 두 패러다임의 재귀적 구조는 (그림 1)에 나타나있다.

측정값과 예측값의 오차 e_k 는 식(3)과 같이 나타낼 수 있다. 이 e_k 를 최소화하기 위한 칼만 이득 (Kalman Gain) K는 수식 (4), (5)을 이용하여 구할 수 있다. 식 (4), (5)의 P와 R은 오차 공분산과 측정 값 노이즈의 공분산 매트릭스들이다.

$$e_k = |z_k - Hx_k| \tag{3}$$

$$K = P \cdot H^T (HP \cdot H^T + R)^{-1} \tag{4}$$

$$P = (I - KH)P \tag{5}$$

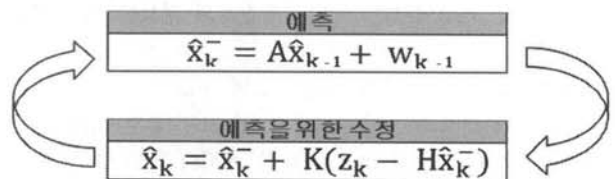
DKF [5]는 KF를 센서 네트워크 환경에 적용한 근사 데이터 모니터링 기법이다. DKF에서 기지국은 각 센서에 대응하는 KF모델들을 갖고 센서는 자신의 KF를 갖는다.

각 센서는 사용자 오류 범위인 δ 를 이용하여 실측값과 예측값의 차이가 δ 보다 작다면 예측값을 실측값의 근사값으로 간주하고 실측값을 기지국에 전송하지 않는다. 기지국에서는 해당 센서의 KF 모델을 유지하고 있으므로 예측값을 데이터의 전송 없이도 획득할 수 있다. 아래의 (그림 2)는 기지국과 센서와의 근사 데이터 수집 방법을 도식화한 것이다.

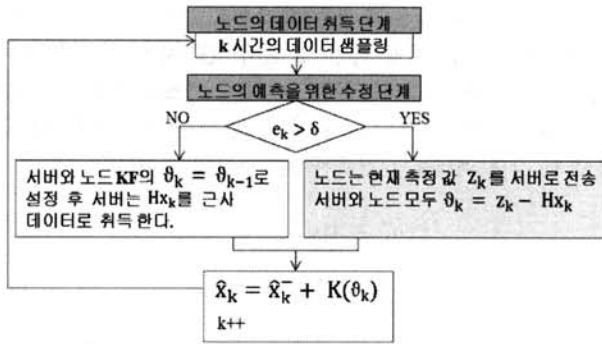
(그림 2)에서 센서에서의 $e_k > \delta$ 일 때를 제외하고는 센서는 기지국에 측정값을 전송하지 않는다. 기지국은 데이터가 수신되지 않으면 해당 센서의 측정값이 예측 범위 안에 있다고 판단하여 센서와의 데이터 통신 없이 Hx_k 를 근사 데이터로 취득할 수 있다. 센서의 오차 e_k 가 사용자 파라미터인 δ 보다 크게 되면 기지국에 실제 측정값 z_k 를 전송하여 각각의 KF를 다시 동기화 한다.

본 논문에서는 DKF와 같이 기본적으로는 각 센서에서 취득되는 데이터의 변화를 이용하여 현재의 시간에서 다음에 취득되는 데이터를 예측하기 위하여 KF를 사용한다.

이 후 적응적 샘플링에서는 예측된 데이터의 오차들을 이



(그림 1) KF의 재귀적 구조



(그림 2) 기지국-센서의 근사 데이터 수집 방법

용하여 각 센서에서 취득되는 데이터를 분석하여 우선순위를 파악하고 이에 따라 적응적으로 샘플링 주기를 조정하기 위함이다. 적응적 샘플링에 대한 자세한 설명은 4장에서 다루도록 한다.

4. 적응적 샘플링

본 연구는 DKF와 적응적 샘플링 기법을 통합하여 센서의 에너지 소비를 줄이는 기법을 제안한다. 적응적 샘플링이란 대상의 중요도에 따라 샘플링 주기 (Sampling Interval : IS) 를 조정함으로써 센서의 샘플링 횟수를 적절히 조정하는 것이다. 일반적으로, 센서의 데이터 전송 비용이 가장 큰 비중을 차지하지만 센서의 샘플링 비용도 고려하여야 한다. 또한, 무선 센서 네트워크에서 센서의 샘플링 주기는 곧 데이터 전송 주기이다.

즉, 센서가 데이터를 취득하게 되면 측정값을 전송 경로 (routing path) 를 따라서 기지국으로 전송하게 되는데 샘플링 주기를 대상에 중요도에 따라 적절히 조절할 수 있다면 불필요하게 빈번히 샘플링하여 전송되는 데이터양을 줄일 수 있다. 따라서, 적응적 샘플링을 이용하면 네트워크 전송 비용 및 샘플링 비용을 줄일 수 있는 장점이 있다. 부과적으로 센서는 각기 다른 샘플링 주기를 가지게 되고 데이터를 전송하는 시점이 분산되게 된다. 이로써 특정 시간에 집중하여 소모되는 네트워크 대역폭을 시간에 따라 적절히 분산시킬 수 있다.

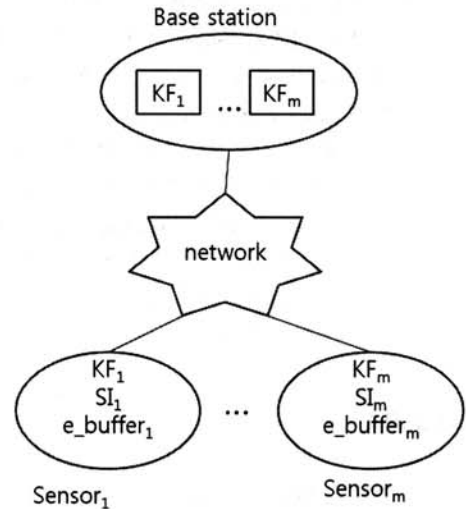
본 논문에서 제안하려고 하는 적응적 샘플링 기법은 기존의 기법들과 다음과 같은 차별성을 가진다. 우선, 기존의 Load-shedding 기법 [7]과 달리 제안된 기법은 예측 모델을 이용 네트워크 혼잡 상태를 미연에 방지한다. 또한 기존의 적응적 샘플링 기법들은 각 센서에서 현시점의 측정값과 예측값의 오차만을 이용하여 샘플링 주기를 판단하는데 반하여 제안하는 기법에서는 과거 측정값들의 정보를 이용한다.

4.1 샘플링 주기 변화를 위한 시스템 구조

센서 데이터에서 모니터링에서 화재나, 홍수, 침입 등이 발생한 지역은 기타의 다른 지역의 데이터 보다 좀 더 빈번히 외부환경을 측정할 필요가 있다. 이러한 지역들의 측정

값은 다른 지역에 비하여 측정값이 급속히 변하는 특징을 지닌다.

본 논문에서 제한하는 시스템의 구조는 (그림 3)과 같다. (그림 3)에 나타나 있는 시스템 구조는 DKF 기법과 적응적 샘플링 기법을 고려하여 설계된 것이다. (그림 3)에 나타난 바와 같이 각 센서들의 Kalman Filter 모델 (KF_i)을 지니고 있으며 기지국은 모든 센서들의 Kalman Filter 모델들을 유지하고 있다. 또한 각 센서 노드들은 자신의 샘플링 주기 (SI_i)를 가지고 있으면 이 SI를 관리하기 위하여 과거 측정값과 예측값의 차이들을 e_buffer에 유지하고 있다.

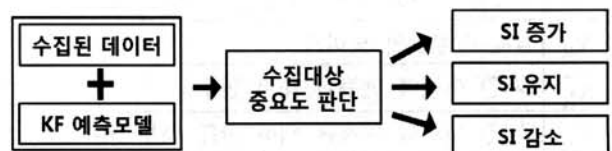


(그림 3) 시스템 구조

4.2 샘플링 주기 변화

(그림 4)는 적응적 샘플링을 도식화한 모델이다. (그림 4)에서 보듯이 수집된 데이터에 Kalman Filter 를 이용한 예측값을 이용하여 수집대상의 중요도를 판단하고 이를 기준으로 샘플링 주기 (SI)를 증가, 감소 또는 유지한다.

본 논문에서 제안하는 적응적 샘플링 기법은 (그림 3)의 구조에서 각 센서들이 독립적으로 SI(Sampling Interval)를 유지하도록 하였다. 이를 위하여 센서는 오차 버퍼(e_buffer)을 유지해야한다. e_buffer의 목적은 샘플링 대상의 중요도를 판단하기 위해서다. 데이터의 중요도를 판단하는데 있어서, 기존의 적응적 샘플링 기법들 [12, 13]과 같이 데이터의 실측값과 예측값의 차이를 기준으로 한다. 즉, 실측값과 Kalman Filter를 이용한 예측값의 차이가 크다면 해당 센서



(그림 4) 적응적 샘플링 모델

에서 중요한 데이터들이 많이 발생하고 있는 것으로 간주하여 해당 센서의 샘플링 주기 (SI)를 감소시킨다. 이와 반대로 실측값과 예측값이 유사하다면 앞으로도 Kalman Filter를 이용한 예측값이 실측값을 잘 반영할 것으로 간주하여 해당 센서의 샘플링 주기를 증가시킨다.

시간 k에서의 실측값 z_k 와 예측값 Hx_k 의 차이는 3장의 식 (3)의 부터 e_k 로 표현할 수 있다. 위에서 언급한 바와 같이, 본 연구는 DKF에 기반하고 있으므로 센서는 e_k 가 오류 범위 δ 보다 클 경우 실측값 z_k 를 기지국으로 전송한다.

즉, e_k 가 어떤 값이던지 간에 δ 보다 클 경우에는 항상 데이터 전송이 발생한다는 것이다. 따라서, e_buffer 에서는 최근 n 개의 $\min(e_k, \delta)$ 를 관리하고 이를 이용하여 측정값의 중요도를 판단한다.

측정값들의 중요도를 판단하기 위하여 e_buffer 에 기록되어 있는 n개의 오차값 ($=\min(e_k, \delta)$) 들을 수식 (6)에 적용하여 e_{state} 를 구한다.

e_{state} 를 구하는데 있어서 가장 최근의 오차 e_k 가 그 이전의 오차 e_{k-1} 보다 더 큰 가중치를 가져야만 최근의 데이터 변화를 정확하게 반영할 수 있다. 따라서, 수식(6)에서 e_buffer 에 저장되어 최근의 e_k 부터 가장 오래된 e_{k-n+1} 까지를 순차적으로 1/2의 지수 승을 함으로써 최근 값에 더 큰 가중치를 부여하여 오차 변화율인 e_{state} 를 구하였다.

수식(6)에서 오차 범위 δ 로 나누는 이유는 e_{state} 를 0과 1 사이의 상대적인 값으로 나타내기 위해서이다. 즉, 최근 n 시간 동안 KF를 이용한 예측값이 정확하여 실측값과 동일할 경우에는 $e_k, e_{k-1}, \dots, e_{k-n+1}$ 은 모두 0임으로 e_{state} 는 0이 된다. 이에 반하여 현재 사용중인 KF 모델에 의한 예측이 계속 맞지 않을 경우 (즉 $\min(e_k, \delta) = \delta$) 에는 다음 수식 (7)에서와 같이 e_{state} 가 1에 근접하게 된다.

위의 수식 (7)에서 n을 무한대로 보냈으나 실제 센서는 메모리의 한계가 존재하게 된다. 본 연구에서는 e_buffer 의 크기 n을 10으로 설정하였다. 이 경우에 e_{state} 의 최대값은 약 0.999 ($= 1 - 1/2^{10}$) 로 1에 가깝다.

다시 말하면, 수식 (6)에 의하여 얻어진 e_{state} 값이 1에 가까워지면 측정값과 예측값의 차이가 심하다는 것이고, 0에 가까워지면 예측모델이 실제 환경을 정확히 예측하고 있다는 것이다. 샘플링 대상의 e_{state} 의 값이 크다는 것은 해당 센서가 위치한 곳에 외부 환경이 급격히 변화하고 있다는 것으로 중요도가 높은 것으로 인지한다. 이 경우에는 샘플링 주기를 감소시켜 측정값을 지속적으로 측정하고 이를 바탕으로 KF의 모델을 계속 갱신하여야 한다. 이와 반대의 경우에는 현재 환경이 KF를 기반으로 예측 가능한 상태에서 발생되고 있는 것임으로 샘플링 주기를 증가시켜 센서의 에너지 소비를 감소시킨다.

실질적으로 e_{state} 은 사용자 정의 파라미터 ρ (SI 증가 기준), λ (SI 감소 기준)를 기준으로 샘플링 대상의 중요도를 낮음, 적당함, 높음으로 판단하고 그에 따라 (그림 5)같이 SI를 조정한다.

본 연구에서 샘플링 주기 SI는 $SI^{min} \leq SI \leq SI^{max}$ 의 범위 안에서 조정이 되고 SI^{max} 는 샘플링 주기 최대값이고 SI^{min} 는 샘플링 주기 최소값이다. 이는 샘플링 주기 SI를 변경하는데 있어서, 샘플링 주기는 무한히 길어진다거나 너무 빈번히 샘플링되는 것을 방지하기 위한 것이다.

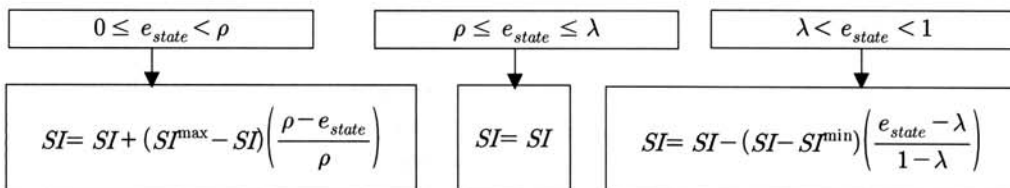
(그림 5)에서 e_{state} 값이 사용자 정의 파라미터 ρ 보다 작을 경우, 샘플링 대상의 중요도가 작은 것으로 인지하고 (즉, 예측 모델이 정확하다) 샘플링 주기를 증가 시킨다. 이 경우, 예측 모델이 정확할수록 ($\rho - e_{state}$)의 값이 커짐으로, 샘플링 주기 SI의 증가폭이 커지게 된다. 또한 위에서 언급한 바와 같이 SI는 최대 SI^{max} 까지만 증가하도록 하였다. 즉 SI가 SI^{max} 일 경우, $SI^{max} - SI$ 은 0이 되어 샘플링 주기 SI가 더 이상 증가되지 못하도록 하였다.

또한 e_{state} 의 값이 ρ 와 λ 사이에 존재할 경우에는 SI가 적절한 것으로 보고 현재 SI를 유지한다. e_{state} 의 값이 λ 보다 클 경우, SI를 감소시킨다. 이 경우 위의 SI를 증가시키는 경우와 유사하게 감소시킨다.

결론적으로 본 연구에서는 과거의 실측값들과 예측값들의

$$e_{state} = \frac{(\min(e_k, \delta) \times \frac{1}{2}) + (\min(e_{k-1}, \delta) \times \frac{1}{2^2}) + \dots + (\min(e_{k-n+1}, \delta) \times \frac{1}{2^n})}{\delta} \tag{6}$$

$$\lim_{n \rightarrow \infty} e_{state} = \lim_{n \rightarrow \infty} \frac{\delta \times \frac{1}{2} + \delta \times \frac{1}{2^2} + \dots + \delta \times \frac{1}{2^n}}{\delta} = \lim_{n \rightarrow \infty} \frac{\delta(\frac{1}{2} + \dots + \frac{1}{2^n})}{\delta} = 1 \tag{7}$$



(그림 5) 샘플링 주기 변경 알고리즘

차이의 변화량을 0과 1 사이의 상대적인 값인 e_{state} 로 계산하고 이를 기반으로 하여 샘플링 주기를 적용적으로 변환하는 알고리즘을 제시하였다. 이를 통하여 샘플링 대상의 변화가 큰 경우에는 샘플링 주기를 단축하여 빈번히 측정, 전송하도록 하고 변화가 없는 경우에는 샘플링 주기를 증가시키도록 하였다. 이를 통하여 센서 네트워크의 자원을 효율적으로 분배하여 전체적인 전력 소비 및 네트워크 대역폭을 효과적으로 사용하도록 하였다.

본 논문에서 제안하는 적응적 샘플링 기법은 기존의 기법들과 큰 차이를 가지고 있다. 기본적으로 <표 2>에 기존의 기법과 제안기법의 주요 차이점에 대하여 간략하게 정리하였다. 제안하는 기법은 기존 기법과 달리 예측 모델을 이용함으로써 실제 측정값과 예측 값 사이의 예측오차를 통하여 샘플링 대상의 변화율을 판단할 수 있다. 예측 오차를 이용하기 때문에 샘플링 대상의 변화를 보다 정확하게 판단할 수 있다. 따라서, 이를 이용하여 샘플링 주기를 적용적으로 조절한다. 또한 센서네트워크 환경에 따라 파라미터를 설정하여 오차율과 샘플링 변화비율을 적절하게 보장한다.

<표 2> 제안기법과 기존 기법들의 차이점

	DKF	DCM	Proposed Technique
샘플링 주기변화	X	O	O
예측모델 사용	O	X	O
파라미터 사용	X	X	O
오차율 보장	O	X	O

5. 실험

이절에서는 본 논문에서 제안하는 적응적 샘플링 기법에 대한 효율성을 보인다. 이를 위하여 고정 샘플링 주기를 지닌 DKF[10]과 기존의 적응적 샘플링 기법인 Data Centric Modeling[13] (DCM으로 표현)을 구현하고 본 논문에서 제안한 기법과 성능 비교를 하였다.

5.1 실험 환경

본 연구의 성능평가를 아래의 <표 3>의 환경에서 시뮬레이션을 하였고 센서 데이터는 Intel Lab에서 제공하는 실제 측정 센서데이터(온도 데이터)를 사용하였다¹⁾. Intel Lab 데이터는 Intel Berkeley 연구실에 배치된 54개의 센서들로부터 얻어진 데이터들로 구성되어 있다.

성능 평가를 위하여 센서들의 에너지 소비 모델은 free space channel model [15]을 이용하였고 아래 <표 3>을 이용하여 수식(8),(9)으로 나타 낼 수 있다.

수식 (8)은 수신 시 발생하는 에너지 소비를 식으로 나타낸 것이다. 수식 (8)에서 k는 비트의 수, $E_{Rx-elec}(k)$ 는 k비트를 수신하는데 드는 비용이다. E_{elec} 의 값은 50nJ/bit이다. 수식 (9)은 송신 시 발생하는 에너지 소비를 나타낸 것이다.

수식 (9)에서 d는 전송 거리를 나타내고 $E_{Tx-elec}(k)$ 는 k비트를 전송하는데 드는 비용이고, $E_{Tx-amp}(k, d)$ 는 이를 d 거리만큼 전송하기 위한 증폭비용이다. 단위 증폭 비용 E_{amp} 는 값은 100pJ/bit/m²이다.

$$E_{Rx}(k) = E_{Rx-elec}(k) = E_{elec} * k \tag{8}$$

$$E_{Tx}(k, d) = E_{Tx-elec}(k) + E_{Tx-amp}(k, d) = E_{elec} * k + G_{amp} * k * d^2 \tag{9}$$

본 실험을 위하여 <표 4>에 나타났는 파라미터들을 사용하였다.

우선, 위의 파라미터들 중 제안하는 기법의 샘플링 주기의 변화 기준이 되는 ρ 와 λ 를 변화시켜 성능을 측정하였다. 이를 통하여 오차가 가장 적은 ρ 와 λ 값을 구하였다. 또한 제안 기법과 다른 기법들을 비교하기 위하여 시간별 전송비용, 시간에 따른 누적 전송 수 그리고 시간에 따른 오차율에 대한 성능을 평가하였다.

<표 3> 통신비용 관련 파라미터

작업	비용
Transmitter Electronics ($E_{Tx-elec}$) Receiver Electronics ($E_{Rx-elec}$) ($E_{Tx-elec} = E_{Rx-elec} = E_{elec}$)	50nJ/bit
Transmit Amplifier(G_{amp})	100pJ/bit/m ²

<표 4> 파라미터들

기호	기본값	설명
SI^{max}	17	최대 샘플링 주기
SI^{min}	7	최소 샘플링 주기
δ	0.09	사용자 오류 범위
ρ	0.35	SI 증가 기준
λ	0.65	SI 감소 기준

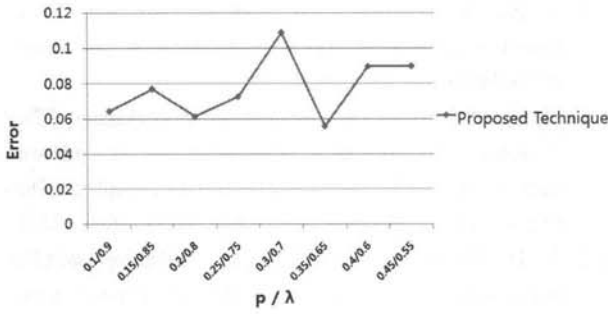
5.2 실험 결과

우선 본 논문에서 제안하는 적응적 샘플링 기법은 ρ 와 λ 값을 기준으로 샘플링 주기를 조정한다. 따라서, ρ 와 λ 의 값을 변화시키면 제안 기법으로 얻을 수 있는 근사값의 오차를 파악하였다. (그림 6)은 이에 대한 결과를 나타낸다.

기본적으로 ρ 와 λ 가 클수록 샘플링 주기가 증가됨으로 센서의 에너지 소비가 적게 된다. 따라서 단순히 에너지 소비만을 고려할 경우에는 ρ 와 λ 에 큰 값을 지정하면 되는 것이다. 하지만 이러한 경우 오차율의 증가를 유발하게 된다. 따라서, 본 실험에서는 (그림 6)에서 가장 적은 오차율을 나타내는 $\rho(=0.35)$ 와 $\lambda(=0.65)$ 를 선택하여 실험을 진행하였다.

DKF의 샘플링 주기가 고정되어 있으며 기본값은 12이다. 이러한 고정 주기 샘플링은 센서의 데이터 전송이 특정 시간에 집중되는 결과를 초래한다. 반면에 제안하는 기법과 DCM은 샘플링 주기가 가변적이다. 이러한 각 센서들의 샘플링

1) <http://db.csail.mit.edu/labdata/labdata.html>에서 얻을 수 있음

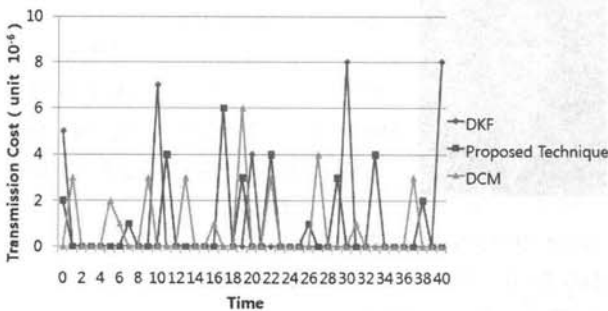


(그림 6) 최적 사용자 변수

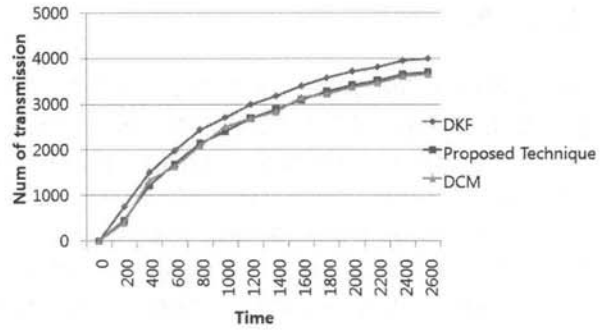
플링 주기의 차이는 센서들의 데이터 전송 시점을 분산시키는 결과를 낸다. 따라서 대역폭이 제한되는 무선 센서 네트워크 환경에서 가변 주기 샘플링이 보다 효율적이다. (그림 7)은 각 기법에 따른 시간별 전송비용을 나타낸다. 여기서 전송비용은 수식 (8), (9)를 이용한 전체 센서 네트워크에서 측정값을 전송하기 위하여 소모된 총 전송비용을 말한다. 실험 결과에서 보듯이 가변 샘플링 주기를 (제안기법과 DCM) 갖는 기법들의 데이터 전송이 분산되는 것을 볼 수 있다. (분산값을 구한 결과 DKF가 4.707, 제안기법이 2.196, DCM이 2.001 나왔다)

(그림 8)은 시간의 흐름에 따른 누적 전송횟수를 보여준다. DKF는 고정된 샘플링 주기로 데이터를 전송한다. 반면 제안하는 알고리즘은 4장에서 설명하였듯이 데이터의 실제 측정값과 예측값의 차이에 따라 적응적으로 샘플링 주기를 조절한다. 따라서 제안하는 기법은 DKF보다 전송 횟수가 적다. 이에 반하여 DCM은 2장에서 설명하였듯이 가장최근에 값과 현재의 값의 차가 크지 않을 경우에 샘플링 주기가 2배 증가 한다. 따라서 제안하는 기법보다 평균적으로 긴 샘플링 주기를 갖게 된다. 따라서 센서가 시간의 흐름에 따라 전송횟수가 감소한다. 동일 시간의 누적 전송횟수를 비교하였을 때 DKF, 제안하는 기법, DCM 순으로 전송횟수가 적음을 알 수 있다.

에너지 효율 측면에서 DCM이 제안하는 기법보다 조금 좋은 에너지 효율을 보인다. 그러나 DCM은 가장 최근의 오차만을 기준으로 샘플링 주기를 2배씩 증가 시킨다. 따라서 (그림 9)에서 나타나듯이 제안하는 기법과 DKF보다는 월등히 높은 오차를 보인다. 평균적으로 제안하는 기법의 오차율은 0.12이고 DCM은 0.57이다. 특히 본 논문에서 제안하는



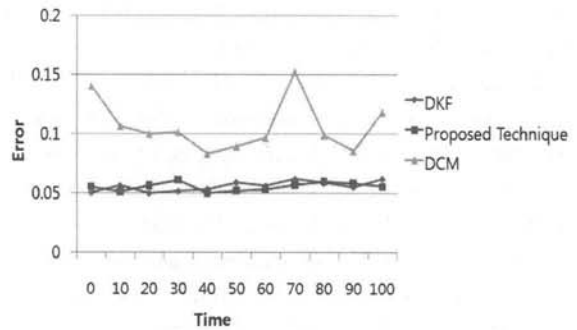
(그림 7) 시간별 전송 비용



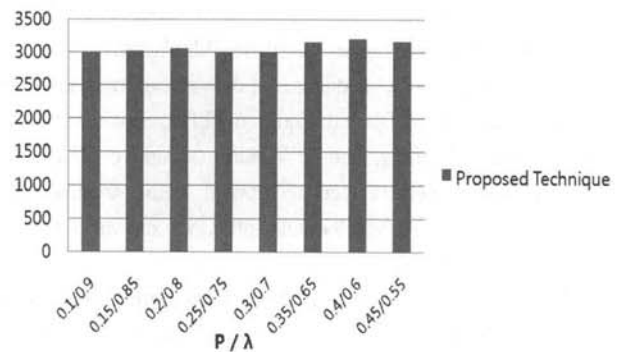
(그림 8) 시간에 따른 누적 전송횟수

기법은 고정 샘플링 주기를 이용하는 DKF와 유사한 오차율을 보인다. 결론적으로 제안 기법은 고정 샘플링 주기의 기법인 DKF과 같은 정확도를 유지하면서도 센서 노드의 에너지 소비를 감소시킨다.

본 논문에서 제안하는 기법은 파라미터를 ρ ($= 0.35$)와 λ ($= 0.65$)를 고정하여 실험 하였다. 따라서 파라미터 값의 변화가 실험에 어떤 영향을 주는지를 알아보기 위하여 파라미터 ρ 와 λ 의 값을 변화시켜가며 누적 전송횟수를 측정하였다. (그림 10)은 실험시간을 1600으로 제한하고 다른 실험환경은 이전과 같게 하여 측정된 결과이다. (그림 10)의 막대 그래프를 보면 알 수 있듯이 파라미터 ρ 와 λ 의 값을 변화시켜도 누적 전송횟수 수에는 큰 변화가 없을 수 있다. 위에서 기술하였듯이 파라미터 ρ 와 λ 의 값을 고정되게 사용한 이유는 적은 오차를 보장하기 위함이다.



(그림 9) 시간에 따른 오차율



(그림 10) 파라미터 변화에 따른 누적 전송횟수

6. 결 론

본 논문에서는 DKF구조에 기반한 적응적 샘플링 기법을 제안하였다. 제안하는 적응적 샘플링 기법은 센서로부터 샘플링되어 취득되는 데이터의 특성을 파악하고 이에 따라 센서별로 샘플링 주기를 조정/유지함으로써 효율적인 네트워크 자원을 이용할 수 있었다. 제안하는 적응적인 샘플링 기법은 전력소비를 줄였고 센서 별 서로 다른 SI 주기를 유지함으로써 데이터 취득과 전송 시점 분산하여 네트워크 대역폭 사용의 효율성을 증대시켰다.

일반적으로 고정 샘플링 기법은 가변 주기 샘플링 기법보다 오차가 작다. 그러나 데이터 전송 측면에서는 적응적 샘플링 기법이 고정 샘플링 기법보다 효율적이다. 본 논문에서는 실험을 통하여 제안하는 기법이 고정 샘플링 기법인 DKF와 유사한 정확도를 보이면서도 기존의 적응적 샘플링 기법인 DCM 기법에 다소 떨어지지만 유사하게 에너지를 소비함을 보였다. 추후 연구로는 데이터의 특성의 변화에 따른 샘플링 주기의 변화를 조정하는 파라미터의 자동 설정에 대하여 연구하고자 한다.

참 고 문 헌

- [1] I. Lazaridis, Q. Han, X. Yu, S. Mehrotra, N. Venkatasubramanian, D. V. Kalashnikov, and W. Yang, "QUASAR: Quality aware sensing architecture," ACM SIGMOD, Vol.33, No.1, pp.26-31, March. 2004.
- [2] P. Bonnet, J. E. Gehrke, and P. Seshadri, "Towards sensor database systems," In proceedings of Second Intl. Conf. on Mobile Data Management, Jan. 2001.
- [3] S. R. Madden, M. J. Franklin and J. M. Hellerstein, "TinyDB: An Acquisitional Query Processing System for Sensor Networks," ACM TODS, Vol.30, No.1, pp.122-173, 2005.
- [4] A. Bharathidasan and V. A. S. Ponduru, "Sensor networks: an overview," IEEE, Vol.22, pp.20 - 23, 2003.
- [5] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "Survey on sensornetworks," IEEE Communications Magazine, Vol.40, No.8, pp.102-116, August. 2002
- [6] S. Goel, T. Imielinski, "Precision based monitoring in sensor networks: Taking lessons form MPEG Computer Communication Review," Vol.40, No 5, pp.82-95, 2001.
- [7] N. Tatbul, U. Cetintemel, S. Zdonik, M. Cherniack, and M. Stonebraker, "Load shedding in a data stream manager," In proceedings of VLDB Intl. Conf. on VLDB, September 2003.
- [8] C. Olston, J. Jiang, and J. Widom, "Adaptive filters for continuous queries over distributed data streams," In proceedings of ACM SIGMOD Intl. Conf. on Management of Data, June 2003.
- [9] D. Tulone, and S. Madden, "PAQ: Time Series Forecasting For Approximate Query Answering," In proceedings of LNCS, Vol.3868, pp.21-37, 2006
- [10] A. Jain, E. Y. Chang, and Y. Wang "Adaptive Stream Resource anagement Using Kalman Filters," In proceedings of SIGMOD Intl. Conf. June. 2004.
- [11] M. Sharaf, J. Beaver, A. Labrinidis and P. Chryanthris, "Tina: A scheme for temporal coherencyaware in- network aggregation," In Proceedings of ACM Intl. Conf. on Data Engineering for Wireless and mobile Access, Sept. 2003.
- [12] A. D. Marbini and L. E. Sacks. "Adaptive sampling mechanisms in sensor networks," In London Communications Symposium, 2003.
- [13] R, Dantu, K. Abbas, M. O'Neill II and A. Mikler "Data centric modeling of environmental sensor networks," Global Telecommunications Conf. 2004.
- [14] G. Welch and G. Bishop, "An Introduction to the Kalman Filter," ACM SIGGRAPH Intl. Conf. on Computer Graphics and Interactive Techniques, August 2001.
- [15] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," In proceedings of IEEE Intl. Conf. June 2000.
- [16] S. Gandhi, S. Nath, S. Suri, and J. Liu. GAMPS "Compressing multi sensor data by grouping and amplitude scaling," ACM SIGMOD Intl. Conf. on Management of Data, pp.171-182, 2009.
- [17] C. Siyao, J. Li. "Sampling based (δ , ϵ)-approximate aggregation algorithm for sensor networks," In Intl. Conf. on Distributed Computing Systems (ICDCS), 2009.



김민기

e-mail : kintelk@kut.ac.kr

2008년 한국기술교육대학교 인터넷미디어 공학부 정보보호전공(학사)

2010년 한국기술교육대학교 정보미디어공학과(석사)

2010년~현 재 한국기술교육대학교 컴퓨터공학과 박사과정

관심분야 : 센서네트워크, 스트림 데이터 등



민준기

e-mail : jkmin@kut.ac.kr

1995년 숭실대학교 전자계산학과(학사)

1997년 한국과학기술원 전산학과(석사)

2002년 한국과학기술원 전산학전공(박사)

2003년~2004년 한국과학기술원 ost-Doc 및 초빙교수

2004년 한국전자통신연구원 선임연구원

2005년~현 재 한국기술교육대학교 컴퓨터공학부 조교수

관심분야 : XML, 시공간DB, 스트림 데이터, 센서네트워크