

적응적 샘플링에 기반한 효율적인 센서 모니터링 기법

김민기, 민준기

한국기술교육대학교 정보미디어공학과

e-mail : kintelk@kut.ac.kr, jkmin@kut.ac.kr

An Efficient Sensor Monitoring Technique based on Adaptive Sampling

Min-kee Kim, Jun-Ki Min

Dept. of Information Media Engineering,

Korea University of Technology and Education

요 약

센서 네트워크 구조에서는 많은 수의 센서 노드들이 지속적으로 센서 데이터를 베이스 스테이션(Base Station)으로 전송한다. 각 노드의 샘플링 주기는 베이스 스테이션으로 전송되는 길목의 네트워크 자원인 대역폭, 계산 비용 등에 지대한 영향을 끼친다. 본 논문에서는 샘플링 대상의 스트림 데이터 특성에 따른 각 노드의 샘플링 주기에 관련된 새로운 적응적 샘플링 기법을 제안한다. 본 논문에서는 KF (Kalman-Filter) 에 기반을 둔 예측 기법을 사용하였다. 이는 각 노드는 KF의 예측값과 실측값의 차를 사용하여 허용된 범위 안에서 자동적으로 샘플링 주기를 조정하는 방식이다. 따라서, 우리는 샘플링 대상의 데이터 특성에 따른 우선순위에 기반 네트워크 자원을 효과적으로 사용하고 Dual Kalman Filter(DKF) 기법과 결합하여 센서 네트워크에서의 가장 큰 문제 중 하나인 에너지 소비의 최소화하면서 효과적으로 근사 데이터 전송하도록 하여 에너지 소비량을 줄였다.

1. 서론

무선 센서 네트워크 (Wireless Sensor Network)의 규모가 커지면서 기존의 대역폭, 배터리 등의 문제가 점점 더 대두되고 있는 시점이다. 무선 센서 네트워크 환경에서는 제한된 대역폭을 이용하여 최대한의 정보를 얻어야 한다는 목적을 두고 있고 제한된 대역폭은 데이터 특성, 질의 조건, 가능한 자원 (통신, 전력, 처리장치), 센서 우선순위 (하나의 센서에서 전송하는 데이터는 다른 센서의 데이터 보다 중요 할 수 있다.) 등을 고려하여 효율적으로 할당되어야 한다[1,2]. 최근 센서 네트워크의 비약적인 발전에 힘입어 모니터링에 관한 연구가 활발히 진행 중이다. 모니터링을 위한 무선 센서 네트워크 구조에서는 기존의 무선 센서 네트워크의 구조 보다 더 많은 센서 데이터의 전송과, 계산을 필요로 하기 때문에 네트워크 자원이 더 많이 부족한 실정이다. 이러한 문제들을 해결하기 위한 연구들이 활발히 진행되고 있다[3,4,5].

본 논문에서는 위에 기술한 문제점들을 해결하기 위하여 무선 센서 네트워크를 구성하고 있는 각 노드의 샘플링 대상의 특성에 따라 샘플링 주기를 적응적으로 조정하는 기법을 제안하려고 한다. 즉 샘플링 대상에 특성에 기반 한 우선순위를 통하여 데이터 수집횟수를 조정하여 제한된 네트워크 자원을 효율적으로 사용하도록 하였다. 우선순위가 높은 데이터는 그렇지 못한 데이터 보다 많은 자원을 할당받게 된다. 따라서, 전체적인 네트워크 측면에서 불 때 중요한 데이터의 수집은 늘리고 불필요한 데이터의 수집을 줄임으로써 자원 낭비가 줄어들기 때문에 전력, 대역폭, 계산 비용 등을 현저히 줄일 수 있다. 또한

DKF [5]를 이용한 근사 데이터 수집과 결합하여 전송 횟수를 감소시키도록 하였다.

2. 관련연구

데이터 스트림 분야에서의 네트워크 자원 관리의 주된 연구 분야에서는 Data Filtering[5,6] 기법이 연구 중이다. 이 기법은 네트워크 자원 관리를 위하여 Adaptive Precision bounds[6]를 사용하여 스트림 데이터에서 특이한 경향이 발생되면 (데이터가 adaptive precision bound 를 만족하지 못할 때) 적은 비용으로 서버에 업데이트 (Update)를 하였다. 하지만 형식화된 샘플링 때문에 제한된 대역폭에서 효율적으로 최대한의 유용한 정보를 수집하지 못하였다.

이전에 제안된 적응적 샘플링 (Adaptive Sampling)[3,9] 기법은 오직 네트워크 상태만을 활용하여 샘플링 주기를 조정하였다. 네트워크에 데이터를 전송하기 전에 네트워크의 상태를 파악하여 네트워크의 데이터 튜플 (tuple)의 손실 비율이 높은 경우에만 샘플링 주기 (Sampling Interval)를 증가 시켰다. 이 기법은 전체적인 네트워크의 데이터 전송 성공률을 증가 시켰지만 샘플링 대상의 중요도에 따른 것이 아니라 네트워크 상태만을 고려하였기 때문에 임의 드롭(Randomly Drop) 현상과 같은 문제로 인하여 데이터 우선순위에 따른 효율적 자원 활용을 보장하지 못하였다.

본 논문에서 제안하려고 하는 적응적 샘플링 기법은 기존의 기법들[5,6,7]과 다음과 같은 차별성을 가진다. 첫째로 Load-shedding[7] 기법은 예측 모델을 사용하지 않는다.

둘째, 우리의 기법은 예측 모델을 이용 네트워크 혼잡 상태를 미연에 방지함에 있지만 Load-shedding 기법은 네트워크 상태가 이미 나빠져 제 기능을 수행 하지 못하였을 때 임의의 데이터 튜플을 드롭 시킨다. 따라서, 기존의 기법들은 효율적인 네트워크 자원 이용을 하지 못하며 데이터 우선순위 또한 보장 할 수 없다.

3. 배경지식

본 절에서는 본 연구의 기반이 되는 Kalman Filter와 DKF(Dual Kalman Filter)에 대하여 간단히 설명하도록 하겠다.

Kaman Filter (KF)는 연속 데이터의 선형 필터링 문제의 재귀적 해법으로 1960년에 R.E.Kalman [8]의하여 소개 되었고 이후 예측처리, 이동체 추적(Object Tracking) 등에 사용되었다. KF는 예측/수정의 두 개의 패러다임(paradigm)의 기반을 둔 내부 상태(Internal State) 예측 선형 알고리즘이다. 3장에서 나오는 기호들에 대한 설명은 다음의 <표 1>에 요약하였다.

<표 1>수식의 기호

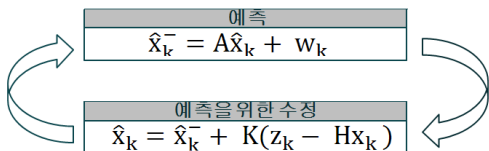
기호	의미
K	시간 인덱스(Index)
X_k	처리 상태 벡터(Vector)
A	X_k 에서 X_{k+1} 으로의 상태 전이 매트릭스(Matrix)
W_k	시스템 모델 노이즈(Noise)
Z_k	실제 측정 값 벡터
H	시스템 상태와 측정 벡터 관련 매트릭스
v_k	실제 측정값의 노이즈
\hat{X}_k^-	k시간 이전에 예측한 내부 변수
\hat{X}_k^+	k시간 이후에 예측을 위해 만든 내부 변수
θ_k	k시간의 예측 가감변수
K	Kalman Gain(각각의 측정치에 가중을 준다)

KF는 다음과 같은 수학적 함수로 구성이 되고 시스템 모델은 아래의 수학적식으로 표현 된다.

$$X_{k+1} = AX_k + W_k \tag{1}$$

$$Z_k = HX_k + v_k \tag{2}$$

일반적인 KF의 예측과 수정의 재귀적 구조는 아래의 <그림 1>과 같다.



(그림 1) KF 예측/수정의 재귀적 구조

KF는 시스템은 예측과 수정의 두 단계로 구성이 되어 있다. 측정값과 예측값의 오차 e_k 는 식(3)과 같이 나타낼 수 있다.

$$e_k = |Z_k - HX_k| \tag{3}$$

이 e_k 를 최소화하기 위한 Kalman Gain K는 함수식 (4),(5)을 이용하여 구할 수 있다.

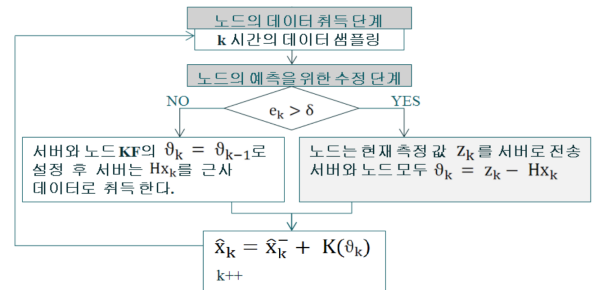
$$K = P \cdot H^T (HP \cdot H^T + R)^{-1} \tag{4}$$

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

식 (4), (5)의 P와 R은 오차 공분산과 측정 값 노이즈의 공분산 매트릭스들이다.

본 연구의 모니터링을 목적으로 최소한의 자원을 이용 근사 데이터를 수집하는데 있다. 즉 우리의 연구 분야에서는 정확한 실측값 보다 KF를 이용한 근사 데이터 취득이 자원 소비 측면에서 보다 우수한 성능을 발휘할 수 있다.

DKF[5]는 KF를 센서 네트워크 환경에 적용한 근사 데이터 모니터링 기법이다. DKF에서 서버는 각각의 노드에 해당하는 KF모델들을 갖고 노드는 자신의 KF를 갖는다. 각 센서 노드는 사용자 파라미터인 예측범위 δ 를 이용하여 예측 범위 안의 데이터를 근사 데이터로써 수용한다는 방식이다. 아래의 (그림 2)는 서버와 하나의 노드와의 근사 데이터 수집 방법을 도식화 한 것이다.



(그림 2) 서버-노드의 근사 데이터 수집 방법

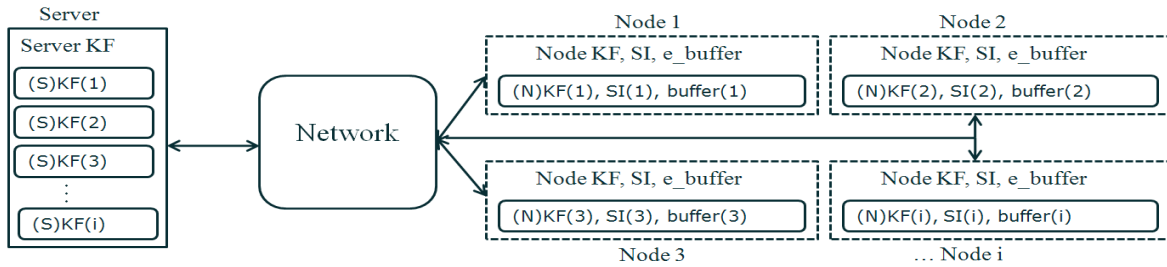
(그림 2)에서 노드에서의 $e_k > \delta$ 일 때를 제외하고는 노드는 서버에 데이터를 전송하지 않는다. 서버는 데이터가 수신되지 않으면 해당 노드의 측정값이 예측 범위 안에 있다고 판단하여 노드와의 데이터 통신 없이 HX_k 를 근사 데이터로 취득 할 수 있다. 노드의 오차 e_k 가 사용자 파라미터인 δ 보다 크게 되면 서버에 실제 측정값 z_k 를 전송하여 각각의 KF를 다시 동기화 한다.

4. 적응적 샘플링

서론에서 기술했듯이 본 연구는 DKF와 적응적 샘플링 기법을 통합하여 센서의 에너지 소비를 줄이는 기법을 제안한다. 본 논문에서 제한하는 시스템의 구조는 (그림 3)과 같다. (그림 3)의 구조는 적응적 샘플링과의 병합 구조를 고려하여 설계된 것이다.

적응적 샘플링이란 대상의 중요도에 따라 샘플링 주기(센서가 데이터를 한번 샘플링하고 다음 번 샘플링 할 때 까지의 기간 (SI: Sampling Interval)을 조정함으로써 센서의 샘플링 횟수를 적절히 조정하는 것이다.

센서의 데이터 전송 비용이 가장 큰 비중을 차지하지만

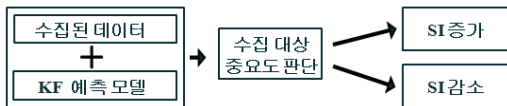


(그림 3) 제안 시스템의 구조

노드의 샘플링 비용도 고려하여야 한다. 일반적으로, 무선 센서 네트워크에서 센서의 샘플링 주기는 곧 데이터 전송 주기이다. 즉 센서가 데이터를 취득하게 되면 데이터를 상위 노드로 전송 하게 되는데 샘플링 주기를 대상에 중요도에 따라 적절히 조정할 수 있다면 빈번하게 발생하는 불필요한 샘플링/전송 데이터를 줄일 수 있다. 따라서, 적응적 샘플링을 이용하면 네트워크 전송 비용 및 샘플링 비용을 줄일 수 있는 장점이 있다.

4.1 샘플링 주기 변화

센서 네트워크 상에서의 데이터 우선순위로 수명만큼 중요한 문제이다. 센서 네트워크에서 허용 가능한 대역폭을 모두 사용하고 있을 때 네트워크에서는 임의로 데이터 드롭 현상이 발생하게 된다. 이는 우선시 되어야 할 데이터가 있을 때 중요도에 상관없이 같은 확률로 데이터를 잃게 되는 현상이다. 따라서, 적응적 샘플링을 사용하면 샘플링 주기에 따라 센서들의 전송 횟수도 달라져 대역폭을 중요도에 따른 우선순위로 사용할 수 있게 된다. (그림 4)는 적응적 샘플링 도식화 모델이다.



(그림 4) 적응적 샘플링 모델

본 논문에서 제안하는 적응적 샘플링 기법은 (그림 3)의 구조에서 각 노드들이 독립적으로 SI(Sampling Interval)를 유지하도록 하였다. 이를 위하여 노드는 (그림 3)과 같이 별도의 오차 버퍼(e_buffer)를 유지해야 한다. e_buffer의 목적은 샘플링 대상의 중요도를 판단하기 위해서다.

본 연구에 있어서, 데이터의 중요도를 판단하는데 있어서, 데이터의 실측값과 예측값의 차이를 기준으로 한다. 즉, 실측값과 Kalman Filter를 이용한 예측값의 차이가 크다면 해당 센서에서 중요한 데이터들이 많이 발생하고 있는 것으로 간주하여 해당 센서의 샘플링 주기 (SI)를 감소시킨다. 이와 반대로 실측값과 예측값이 유사하다면 앞으로 Kalman Filter를 이용한 예측값이 실측값을 잘 반영할 것으로 간주하여 해당 센서의 샘플링 주기를 증가시킨다.

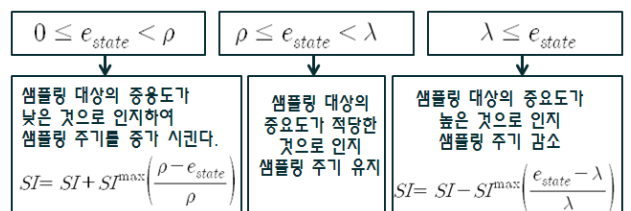
실측값과 예측값의 차이는 2장의 식(3)의 e_k 로 표현 할 수 있다. 각 센서 노드들은 최근 n 개의 e_k 들을 e_buffer에 관리하고 이를 기반으로 각 노드의 중요도의 판단 기준인 e_{state} 를 구한다. e_{state} 는 노드의 오차 변화율이다. 오차 변화율을 구하는데 있어서 가장 최근의 오차 e_k 가 그 이전의 오차 e_{k-1} 보다 더 큰 가중치를 가져야만 최근의 데이터 변화를 정확하게 반영할 수 있다.

(7)

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

수식(7)은 e_buffer에 저장되어 최근의 e_k 부터 가장 오래된 e_{k-n+1} 까지 순차적인 1/2의 지수승을 함으로써 최근 값에 가중을 주어 오차 변화율인 e_{state} 를 구하였다. 또한 식(7)에서 δ 는 DKF에서 사용한 오차 허용 범위로서, 식(7)을 δ 로 나누는 이유는 예측이 적절하게 수행되고 있을 때 e_{state} 를 0과 1 사이의 상대적 값으로 나타내기 위해서고 예측이 범위를 벗어나는 상황($e_k > \delta$)이 아니면 e_{state} 가 1을 넘지 않는다. (e_{state} 가 1을 초과하여도 무방하며 단 순히 e_{state} 를 상대값으로 나타내기 위함이다.)

기본적으로 e_{state} 값이 1을 넘거나 1에 가까워지면 변화가 심해지고 있고, 0에 가까워지면 변화가 적어진다고 판단한다. 샘플링 대상의 오차 변화율이 클수록(e_{state} 의 값이 크다) 중요도가 높은 것으로 인지 한다. (중요도가 높으면 SI감소, 낮으면 SI증가한다.) 실질적으로 e_{state} 오차 변화율은 사용자 정의 파라미터 ρ (SI 증가 기준), λ (SI 감소 기준)를 기준으로 샘플링 대상의 중요도를 낮춤, 적당함, 높음으로 판단하고 그에 따라 (그림 5)같이 SI를 조정한다.

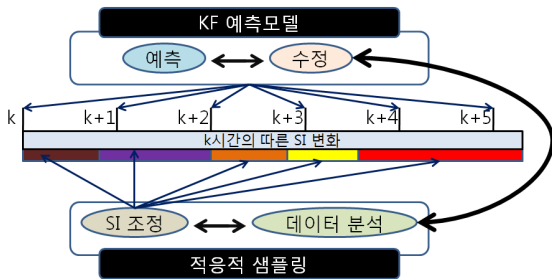


(그림 5) 샘플링 대상 중요도 판단 알고리즘

SI는 $SI^{\min} \leq SI \leq SI^{\max}$ 의 범위 안에서 조정이 되고 SI^{\max} 는 샘플링 주기 최대 값이고 SI^{\min} 샘플링 주기 최소 값이다. 본 알고리즘을 통하여 샘플링 대상의 우선순위 및 효율적 자원 분배를 보장하였고 노드의 샘플링을 통한 전력 소비를 감소 시켰다.

4.2. 적응적 샘플링과 근사 데이터 수집의 통합

적응적 샘플링과 근사 데이터 수집기법은 독립적으로 수행이 된다. 즉 서로의 처리과정에 있어서 서로에게 영향을 주지 않는다. 다만 적응적 샘플링 기법을 위하여 노드의 KF의 수정단계의 e_k 를 e_buffer에 저장하여 사용할 뿐이다. 이를 제외하고는 별도의 연관 작업을 수행하지 않는다. (그림 6)은 노드의 KF가 일정한 주기로 예측을 하는 것과 별개로 SI 변화 하는 것을 나타낸다. 이는 실제로 KF를 이용한 예측 주기와 적응적 샘플링 주기 SI의 필연성이 없다는 것을 말한다. 결과적으로 별도의 연관 작업 및 상호 영향 없이 노드의 KF의 수정 단계와 적응적 샘플링의 데이터 분석단계를 연동하여 DKF에 적응적 샘플링 기법을 적용할 수 있었다.



(그림 6) 예측 주기와 SI와의 관계

5. 성능평가

본 연구의 성능평가를 아래의 <표 2>의 환경에서 30개의 노드에서 총 10만개의 센서 데이터를 발생 시켜 실험을 하였다. 실험 항목은 총 데이터 전송 횟수, 노드의 평균 잔여 전력, 총 데이터 드롭 횟수(할당된 대역폭을 모두 사용하고 있을 때 네트워크의 데이터 손실 횟수), 평균 대역폭 사용률의 항목에 대하여 성능을 평가하였다.

<표 2> 실험의 파라미터 설명

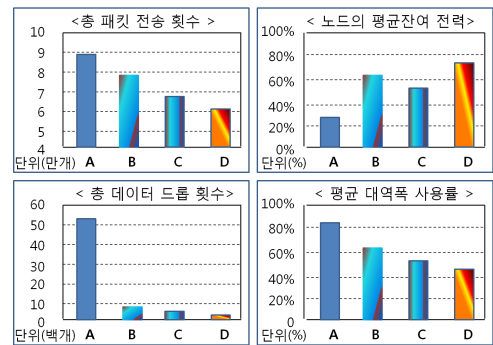
Parameter	Value	Parameter	Value
노드의 수	30개	SI^{\min}	0.2
패킷 크기	500byte	SI^{\max}	2
대역폭	5Mbyte	ρ	0.3
총 발생 데이터 수	10만개	λ	0.7
		δ	0.32

(그림 7)은 각 항목에 대한 결과를 막대 그래프로 표시 하였다. 그림에서 A는 일반적인 모니터링 기법, B는 적응적 샘플링 기법, C는 DKF, D는 적응적 샘플링과 DKF의 통합 방법을 나타낸다. 성능평가 결과 B, C를 각

각 사용 하였을 때 A 보다 각 항목에서 현저하게 우수한 성능을 입증하였고, B와 C를 비교하였을 때는 C의 성능이 좀 더 우수 하였다. B 와 C 같이 사용한 D는 각각을 독립적으로 사용한 것 보다 우수한 성능을 보였다. 이로써 DKF를 이용한 근사 데이터 수집에 적응적 샘플링 기법을 적용하였고 또한 기존의 기법보다 우수한 성능을 보인다는 것을 증명 하였다.

6. 결론

DKF구조의 근사 데이터 수집을 통한 데이터 전송량을 상당량 줄일 수 있다. 그러나 본 논문에서 제안 기법인 또한 DKF 구조에 적응적 샘플링 기법의 적용을 통하여 데이터의 중요도에 따른 네트워크 자원 분배를 통해 효율적으로 네트워크 자원 사용 할 수 있었다. 또한 샘플링 주기를 조절함으로써 노드의 샘플링을 통한 전력 소비를 줄일 수 있었다.



(그림 7) 성능평가 결과

참고문헌

[1] I. Lazaridis, Q. Han, X. Yu, S. Mehrotra, N. Venkatasubramanian, D. V. Kalashnikov, and W. Yang. QUASAR: Quality aware sensing architecture. Appeared in ACM SIGMOD Record, vol. 33(1), March 2004.
 [2] P. Bonnet, J. E. Gehrke, and P. Seshadri. Towards sensor database systems. In Second Intl. Conf. on Mobile Data Management, Hong Kong, January 2001.
 [3] Ankur Jain, Edward Y. Chang Adaptive Sampling for Sensor Networks Proceedings of the First Workshop on Data Management for Sensor Networks (DMSN 2004), Toronto, Canada, August 30th, 2004.
 [4] Daniela Tulone, Samuel Madden PAQ: "Time Series Forecasting For Approximate Query Answering In Sensor Networks" Work supported by the National Science Foundation under Grant No. IIS-0448124.
 [5] Ankur Jain, Edward Y. Chang, YuanFang Wang Adaptive Stream Resource anagement Using Kalman Filters In Proc.of SIGMOD 2004, June 1318, 2004, Paris, France.
 [6] C. Olston, J. Jiang, and J. Widom. Adaptive filters for continuous queries over distributed data streams. In Proc.of ACM SIGMOD Intl. Conf. on Management of Data, San Diego, California, USA, June 2003.
 [7] N. Tatbul, U. Cetintemel, S. Zdonik, M. Cherniack, and M. Stonebraker. Load shedding in data streams. In 29th Intl. Conf. on Very Large Data Bases (VLDB), pages 309 -- 320, Berlin, Germany, September 2003.
 [8] Greg Welch, Gary Bishop An Introduction to the Kalman Filter http://www.cs.unc.edu/~welch
 [9] A. D. Marbini and L. E. Sacks. Adaptive sampling mechanisms in sensor networks. In London Communications Symposium, London, UK, 2003.