

무선 센서 네트워크에서 데이터 압축을 위한 피드백 배포 기법

A Feedback-Diffusion Algorithm for Data Compression in Wireless Sensor Networks

여명호*, 성동욱*, 이석재**, 유재수*
충북대학교*, (주)애플테크**

Yeo myung-ho*, Seong dong-ook*, Lee seok-jae**,
Yoo jae-soo*
Chungbuk National University*, Appletek Inc.**

요약

네트워크 분야에서 데이터 압축은 네트워크 트래픽을 줄이기 위한 전통적이고, 효과적인 방법 중 하나이다. 센서 네트워크의 데이터는 시·공간적인 연관성을 가지고 있으며, 이러한 특성을 이용한 데이터 압축 기법들이 많이 연구되고 있다. 센서 노드는 통신 범위의 제약을 가지고 있기 때문에 공간적인 특성을 이용한 압축시 데이터의 범위로 제한된 정보를 사용한다. 하지만, 때때로 전역적인 데이터 분포와 특성은 데이터 압축 기법의 효율을 증가시킬 수 있는 기회를 제공하기도 한다. 본 논문에서는 이러한 기회를 활용하기 위한 새로운 접근의 데이터 압축 기법을 제안한다. 제안하는 기법은 기지국 혹은 슈퍼 노드에 의해 수집된 데이터를 통해 압축을 위한 정보, 즉, 피드백 정보를 생성하고 배포함으로써 네트워크 전체의 데이터 압축을 용이하게 한다. 본 논문의 우수성을 보기 위해서 시뮬레이션을 통해 성능 평가를 수행하였으며 그 결과 네트워크의 수명이 약 30% 증가하였다.

Abstract

Data compression techniques are traditional and effective to reduce the network traffic. Generally, sensor data exhibit strong correlation in both space and time. Many algorithms have been proposed to utilize these characteristics. However, each sensor just utilizes neighboring information, since its communication range is restrained. The distribution and characteristics of whole sensor data provide other opportunities to enhance the compression technique. In this paper, we propose an orthogonal approach for compression algorithm. The base station or a super node generates useful information for compression of sensor data and broadcasts it into sensor networks. Every sensor that received the information compresses their sensor data and transmits them to the base station. We define this approach as feedback-diffusion. In order to show the superiority of our approach, we compare it with the existing aggregation algorithms in terms of the lifetime of the sensor network. As a result, our experimental results show that the whole network lifetime was prolonged by about 30%.

I. 서론

최근 무선 센서 네트워크는 환경 감시, 스마트 공간, 의료 기술, 농업 기술 분야와 같은 다양한 분야의 응용으로 각광받고 있다[1]. 이러한 센서 네트워크는 온도, 습도, 진도 등과 같은 주변 환경의 유용한 정보를 수집하고, 보다 다양한 처리를 위해서 수집한 정보를 싱크 노드로 전송한다. 센서 네트워크에서 사용되는 센서 노드는 에너지 사용의 제약을 가지고 있기 때문에 데이터를 수집하거나 전송하는데 있어서 에너지를 효율적으로 사용하기 위한 방법이 필요하다.

데이터 압축은 네트워크의 트래픽을 줄이기 위한 전통적이고 효과적인 방법 중 하나이다. 센서 네트워크의 데이터는 네트워크 특성상 시간적 혹은 공간적 연관성을 지닌다[1]. 이러

한 데이터 연관성을 이용하여 센서 데이터를 압축하기 위한 많은 연구들이 진행되었다. [2]에서 제안하는 시간적인 압축 기법은 센서 데이터의 오차 허용치를 미리 정하고, 수집한 데이터의 변화가 오차 범위보다 작은 경우 데이터를 기지국으로 전송하지 않는다. 이때, 기지국은 데이터를 송신하지 않은 센서 노드의 데이터를 시간적인 유사도 관점에서 마지막으로 보고된 데이터의 오차 허용치 범위에 있다고 판단한다. [3]은 이웃 노드의 공간적인 연관성을 이용하여 데이터를 압축한다. 모든 센서 노드들은 서로 다른 시간 슬롯(slot)을 할당 받고, 시간 슬롯에 따라 자신의 데이터를 기지국으로 전송한다. 센서 노드들은 자신의 시간 슬롯을 기다리는 동안 이웃 노드들이 기지국으로 전송하는 데이터를 도청한다. 이때, 도청한 모든 값에 대한 평균값을 계산하고, 자신의 값과 이 평균값을 비교

하여 동일한 경우 데이터를 전송하지 않는다. [4]는 센서 노드들이 클러스터를 형성하고, 각 클러스터의 멤버 노드들은 자신의 클러스터 헤드 노드에게 자신의 데이터를 전송한다. 이때 클러스터 헤드 노드는 수집된 멤버 노드들의 데이터에서 중복된 데이터를 제거함으로써 데이터 압축을 수행한다. [1]은 수집된 데이터의 유사성을 기반으로 클러스터를 생성하고, 클러스터 헤드에서 중복제거와 양자화를 통해 데이터 압축을 수행한다. 기존 데이터 압축 기법의 경우, 대부분 공간적인 압축을 수행하기 위해서 통신 범위로 제한된 지역적인 데이터를 이용한다. 하지만 때때로 전역적인 데이터 분포와 특성은 데이터 압축 기법의 효율을 증가시킬 수 있는 기회를 제공한다.

본 논문에서는 기존 압축 기법과 교차 사용될 수 있는 새로운 접근의 압축 기법을 제안한다. 기지국 또는 다른 센서 노드의 데이터를 수집 가능한 센서 노드가 데이터를 수집하고, 수집된 데이터의 분포와 특성을 요약하여 다시 센서 네트워크로 배포한다. 이 정보를 수신한 센서 노드는 이 정보를 기준으로 자신의 데이터를 압축함으로써 데이터 압축의 효율이 증가된다. 이러한 일련의 과정을 피드백 배포(Feedback-Diffusion)로 정의한다.

본 논문의 구성은 다음과 같다. 제2장에서는 본 논문에서 제안하는 데이터 압축 기법의 특징과 상세한 과정을 기술한다. 제3장에서는 성능평가와 분석을 통해 제안하는 기법의 우수성을 보이고, 제4장에서 결론과 향후 연구 방향을 제시한다.

II. 제안하는 데이터 압축을 위한 피드백 배포 기법

본 장에서는 센서 데이터를 압축하기 위한 새로운 접근을 제안한다. 제안하는 압축 알고리즘은 기존의 압축 알고리즘과 교차 사용이 가능하다. 제안하는 알고리즘은 수집된 데이터를 통해 압축을 위한 정보, 즉, 피드백 정보를 생성하고 배포함으로써 네트워크 전체의 데이터 압축을 용이하게 한다. 먼저 전체 알고리즘의 특징과 피드백 배포의 개념을 설명하고, 배포 방법에 따른 분류와 피드백 생성 기법, 근사 데이터 수집 기법의 일련의 절차를 기술한다.

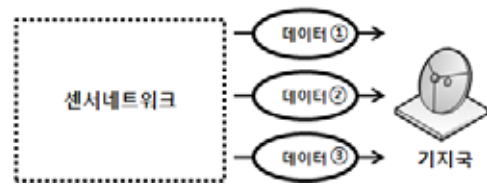
1. 특징

기존 센서 데이터의 공간적인 연관성을 이용한 데이터 압축 기법은 그림 1과 같이 기지국이 일방적으로 센서 네트워크로부터 데이터를 수집한다. 데이터는 센서 네트워크 내부에서 각 알고리즘의 특성에 따라 압축이 이루어지기도 한다. 제안하는 피드백 배포 기법은 수집된 데이터를 바탕으로 데이터 압축을 위한 피드백 정보를 생성하고, 센서 네트워크로 정보를 배포함

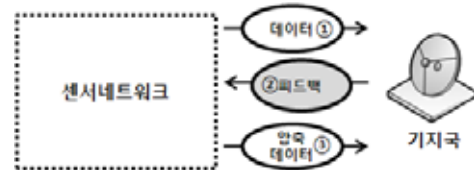
으로써 원본 데이터를 압축할 수 있다. 물론 피드백을 배포하기 위해서 에너지의 소모된다. 하지만, 적절한 피드백 정보를 생성하고 배포함으로써 압축의 효율을 높일 수 있다면 피드백 배포에 따른 에너지 손실과 압축된 데이터의 전송에 따른 에너지 이득은 상호 보완적인 관계에 있다. 따라서 어떻게 피드백 정보를 생성하고, 배포하느냐 하는 것이 본 논문의 관건이다.

피드백 배포의 장점은 다음과 같다.

- 전역적인 데이터의 분포 특성에 관한 정보를 제공한다.
- 데이터 수집 단계에서 데이터의 압축이 가능하게 한다.
- 근사 데이터 수집을 위한 전역적인 기준을 제공한다.



(a) 기존 데이터 수집 방법



(b) 피드백 정보를 이용한 데이터 수집 방법

▶▶ 그림 1. 피드백 배포

2. 피드백 정보의 생성

피드백 정보는 다양한 형태로 제공될 수 있다. 본 논문에서는 기본적으로 센서 데이터의 빈도를 기반으로 한 피드백 정보 생성 기법을 제안한다. 빈도가 높은 센서 데이터를 양자화함으로써 전체 데이터의 압축 효율을 높인다. 이 같은 방법은 허프만 코딩(Huffman Coding)과 일맥 상통한다. 하지만, 피드백 정보를 센서 네트워크로 배포하는 것은 상당한 에너지 소모를 발생시킨다. 따라서, 피드백 정보의 크기를 줄이기 위해서 센서 데이터의 오류에 관련된 몇 가지 가정을 통해 가벼운 피드백 정보를 생성한다.

가정(1). 센서 데이터는 일정 범위의 오차를 가지고 있다.

응용에 따라 센서 데이터의 오차 허용 범위를 정할 수 있으며 다음과 같은 식(1)을 이용함으로써 오차 허용 범위 내에 있는 센서 데이터를 동일한 센서 데이터로 그룹화한다. 이때, ϵ 는 오차 허용범위를 나타내며, V_{real} 과 V_{base} 는 각각 실제 센서 데이터와 기준 값을 나타낸다. 그룹화를 통해 피드백 정보에

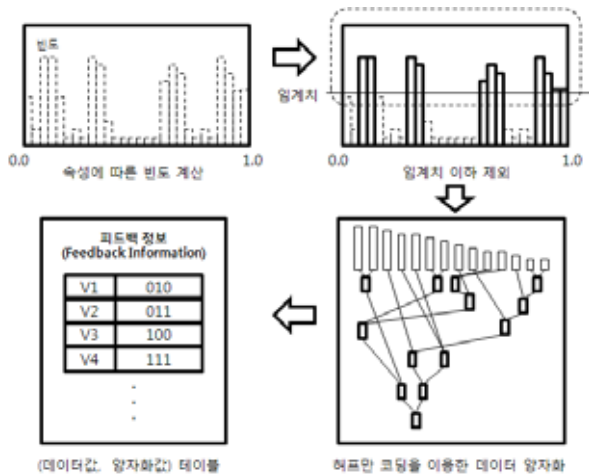
참여하는 센서 데이터의 수를 줄일 수 있다.

$$|V_{real} - V_{base}| \leq \epsilon \quad (1)$$

가정(2). 일부 소수의 센서 노드는 완전히 잘못된 값을 센싱한다.

센서 노드는 자원의 제약으로 인해 센서 자체의 오류가 발생하거나 배포 환경의 상황에 따라 오류가 발생하기 쉬운 환경에 배포된다. 따라서 다른 센서 데이터의 경향성을 따르지 않는 Outlier 데이터를 보고하기도 한다. 따라서 빈도가 임계치 E 보다 작은 센서 데이터는 제거함으로써 피드백 정보에 참여하는 센서 데이터의 수를 줄일 수 있다.

그림 2는 피드백 정보를 생성하는 과정을 나타낸다. 속성에 따른 정렬을 통해 오차 허용 범위 ϵ 에 포함되는 센서 데이터를 그룹화하고, 센서 데이터의 빈도가 임계치 E 이하인 센서 데이터 그룹은 제거한다. 그 다음, 허프만 코딩을 이용하여 양자화하고, {측정값, 양자화값}으로 구성된 피드백 정보 테이블(Feedback Information Table)을 생성한다.

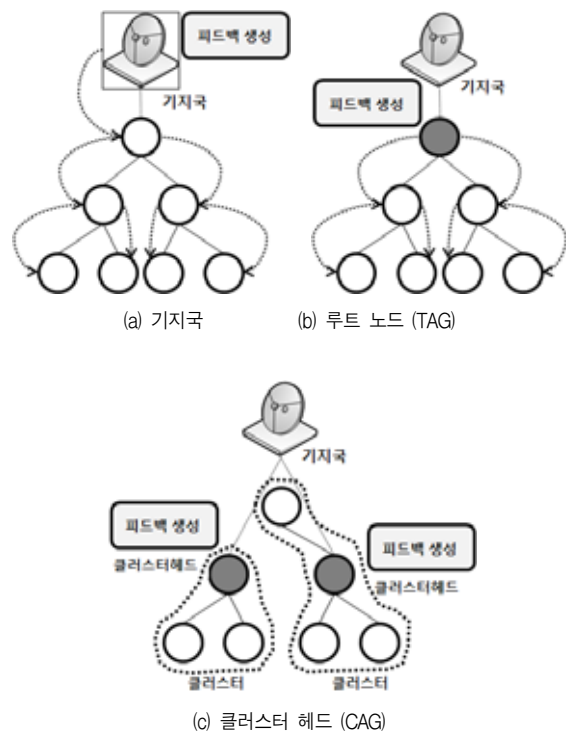


▶▶ 그림 2. 피드백 정보 생성 과정

3. 피드백 배포

피드백의 배포는 크게 네트워크 내부에서 이루어지는 경우와 외부, 즉, 기지국으로부터 이루어지는 경우로 나누어진다. 그림과 같이 피드백을 생성하는 주체에 따라 그림 3과 같이 몇 가지로 구분할 수 있다. 첫 번째, 기지국에서 수집한 데이터를 통해 피드백을 생성하는 경우, 기지국으로부터 배포가 시작된다. 이 경우, 기지국은 일반적으로 센서 노드보다 좋은 계산 성능을 가지고 있으므로 수집된 데이터의 복잡한 처리가 가능하다. 또한, 센서 네트워크의 모든 데이터를 수집함으로써 전역적인 데이터 특성을 완벽하게 피드백 정보에 반영할 수 있다. 하지만, 기지국이 센서 네트워크와 거리가 먼 경우 피드백 정보 생성을 위한 데이터 수집 시 많은 에너지를 소모한다는 단점

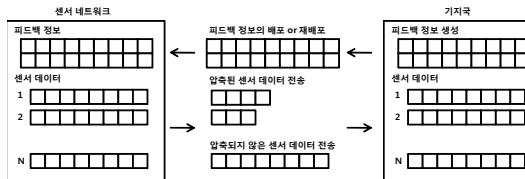
이 있다. 두 번째, TAG를 기반으로 하는 데이터 수집 알고리즘의 경우, 라우팅 트리의 상위 노드 혹은 루트 노드가 피드백 정보를 생성할 수 있다. 이 경우, 전역 혹은 부분적인 데이터의 특성이 반영될 수 있으며, 센서 네트워크 내부에서 피드백 정보가 배포됨으로써 피드백 정보의 라우팅 비용을 줄일 수 있다는 장점을 가지고 있다. 세 번째, CAG를 기반으로 하는 데이터 수집 알고리즘의 경우, 지역적으로 클러스터를 구성하고 클러스터 헤드가 피드백 정보를 생성한 후 배포한다. 클러스터 구성 방법에 따라 피드백 정보 생성 방법이 변할 수 있으며, 압축의 효율을 높일 수 있다.



▶▶ 그림 3. 피드백 생성 주체에 따른 분류

4. 압축 데이터 수집

피드백 배포를 통해 각 센서 노드로 전달된 피드백 정보는 각 센서 노드가 센서 데이터를 양자화하는데 활용된다. 그림 4는 피드백 정보를 배포하고 압축 데이터를 수집하는 과정을 나타낸다. 기지국에서 피드백정보를 생성하고, 센서 네트워크로 피드백 정보를 배포한다. 각 센서 노드는 이 정보를 저장하고 있으며, 자신이 측정된 센서 데이터가 피드백 정보의 오차 허용 범위 ϵ 에 포함되는 경우, 압축 정보를 기지국으로 전송한다. 만약 피드백 정보에 포함되어 있지 않다면, 가정(2)에 의해 데이터 전송을 포기한다. 하지만, 응용 프로그램의 특성에 따라 압축되지 않은 정보를 그대로 전송함으로써 수집 데이터의 정확도를 향상시킬 수 있다.



▶▶ 그림 4. 피드백 정보 배포와 압축 데이터 수집 과정

5. 피드백 정보의 갱신

매주기마다 피드백 정보를 배포하는 것은 에너지를 크게 소모한다. 반대로 피드백 정보를 갱신하지 않으면, 압축율이 저하되는 문제점을 가지고 있다. 제안하는 피드백 배포 기법은 적절한 갱신 시점을 찾기 위해서 기지국에서 수집되는 데이터의 수를 계산하고, 데이터의 수가 일정 크기보다 적은 경우 피드백 정보를 갱신하기 위해서 새로운 피드백 정보를 배포한다. 이때, 양자화된 데이터의 수집하는 정도를 "Hit율"이라고 정의한다. 기지국은 수집 데이터의 Hit율이 센서 배포 전에 설정된 최소 Hit율보다 낮은 경우 피드백 정보를 다시 배포하고, 각 센서 노드의 피드백 정보는 갱신된다.

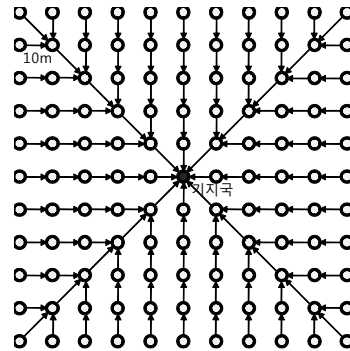
III. 성능 평가 및 분석

1. 실험 환경

제안하는 피드백 배포 기법의 우수성을 보이기 위해 TAG 기반의 일반적인 데이터 수집 기법과 시뮬레이션을 통해 성능을 비교평가 하였다. 표 1은 시뮬레이션에 사용된 환경 변수이다. 시뮬레이션에서 사용된 데이터는 미국 워싱턴주의 온도 데이터를 활용하였다[5]. 공간적인 연관성을 만들기 위해서 데이터 모델의 위상차를 두어 각 센서 노드가 서로 다른 데이터를 수집하도록 설정하였으며, 센서 노드는 그림 5와 같은 형태의 네트워크 토폴로지를 형성한다. 센서 노드의 메시지 전송에 소모되는 에너지 모델은 {메시지 크기}x{전송 비용}+{증폭 비용}x{거리}이며, 전송 비용은 50nJ/b, 증폭 비용은 100pJ/b/m²로 설정하였다. 메시지 수신에 소모되는 에너지 모델은 {메시지 크기}x{수신 비용}이며, 수신 비용은 50nJ/b로 설정하였다. 수집된 센서 데이터의 정확성을 보장하기 위해서 피드백 정보를 이용할 수 없는 데이터의 경우 양자화되지 않은 원본 데이터를 그대로 수집하는 정책을 사용하였다.

[표 1] 성능 평가 환경

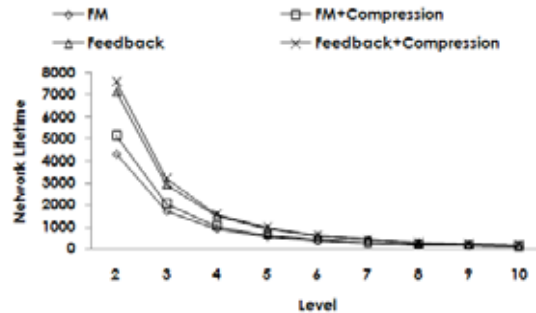
파라미터	변화	기본값
라우팅 트리의 레벨	2-10레벨	3레벨
센서 식별자	-	4byte
센서 데이터의 크기	-	4byte
오차 허용 범위 (ε)	0-4%	1%
피드백 정보의 Hit율	10-95%	80%
발생 빈도에 따른 임계치 (E)	1-16%	4%



▶▶ 그림 5. TAG 기반의 네트워크 토폴로지

2. 실험 결과

그림 6은 TAG 기반의 FM(Flooding Method)과 제안하는 피드백 배포 기법의 네트워크 수명을 비교한 결과이다. 기존 압축 기법의 교차 사용을 평가하기 위해서 [2]에서 사용된 압축 기법을 적용하였다. x축의 값은 라우팅 트리의 레벨을 의미한다. 실험결과 FM에 비해 제안하는 기법이 약 30%의 네트워크 수명 향상을 보였으며, 각 기법에 압축기법을 적용한 경우 추가적인 에너지 사용의 효율이 발생하였다. 이는 제안하는 기법의 특성상 기존 압축 알고리즘과 교차사용이 가능하다는 것을 의미하며, 전체적으로 피드백 배포를 통한 압축을 수행함으로써 데이터 전송에 소모되는 에너지 소모가 크게 줄었다는 것을 보여준다.



▶▶ 그림 6. 네트워크 수명 비교

IV. 결론

본 논문은 기존 압축 기법과 교차 사용 가능한 피드백 배포 기법을 제안한다. 공간적인 특성을 이용한 대부분의 압축 기법은 통신의 제약으로 인해 인접한 센서 데이터의 특성만을 이용하는 단점이 있다. 제안하는 기법에서 기지국 혹은 슈퍼 노드에 의해 생성된 피드백 정보는 센서데이터의 전역적인 분포 특성을 반영하고 있기 때문에 효율적인 양자화를 수행한다. 시뮬레이션을 통한 실험 결과, 기존 데이터 수집 기법에 비해 네

트위크의 수명이 약 30% 향상되었고, 기존 압축 기법과 교차 사용함으로써 에너지 사용의 효율을 더욱 높일 수 있었다. 또한, 센서 네트워크의 오류와 센서 데이터의 오차 범위를 이용하여 에너지 효율적인 데이터 수집이 가능함을 확인하였다. 향후 연구는 피드백 정보를 효율적으로 압축하는 기법과 내부 네트워크 처리를 통한 배포 기법을 제안하는 것이다.

■ 참고 문헌 ■

- [1] 여명호, 이미숙, 박종국, 이석재, 유재수 "무선 센서 네트워크에서 네트워크 트래픽 감소를 위한 데이터 중심 클러스터링 알고리즘", 정보과학회논문지: 정보통신, 제35권, 제2호, pp.139-148, 2008년 4월.
- [2] M. Sharaf, J. Beaver, A. Labrinidis and P. Chryanthi, "Tina: A scheme for temporal coherency-aware in-network aggregation", In Proceedings of the 2003 ACM Workshop on Data Engineering for Wireless and mobile Access, Sept. 2003.
- [3] X. Meng, L. Li, T. Nandagopal and S. Lu, "Event contour: An efficient and robust mechanism for tasks in sensor networks", In Proceedings of Technical report, 2004.
- [4] S. Patten, B. Krishnamachari and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks", In Proceedings of International Conference on Information Processing in Sensor Networks, 2004.
- [5] Live from Earth and Mars (LEM) Project, <http://www-k12.atmos.washington.edu/k12/grayskies/>, 2006.