센서 네트워크에서 데이터 압축을 위한 피드백 배포 기법

(A Feedback Diffusion Algorithm for Compression of Sensor Data in Sensor Networks)

여 명 호 † 성 동 욱 † 조 용 준 † 유 재 수 ††
(Myung Ho Yeo) (Dong Ook Seong) (Yong Jun Cho) (Jae Soo Yoo)

요 약 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 제한된 범위내의 통신이 가능하며, 자신의 통신 반경 내의 데이터만을 활용한다. 만약 네트워크의 전체 데이터 분포 특성을 활용할 수 있다면, 데이터 압축의 효율을 증가시킬 수 있다. 본 논문에서는 네트워크 전체 데이터 분포 특성을 활용하기 위한 새로운 접근의 피드백 배포 기법을 통한 데이터 압축 기법을 제안한다. 제안하는 기법은 기지국 혹은 슈퍼 노드에 의해 수집된 데이터의 빈도를 이용하여 허프만 코드를 생성하고, 배포함으로써 네트워크 전체의 데이터 압축을 용이하게 한다. 본 논문의 우수성을 보이기 위해서 시뮬레이션을 통해 성능 평가를 수행하였으며 그 결과 네트워크의 수명이 약 30% 증가하였다.

키워드 : 센서 네트워크, 피드백 배포, 데이터 압축, 허프만 코딩

Abstract Data compression technique is traditional and effective to reduce 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, because its communication range is restrained. Information that includes 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 based on a novel feedback diffusion algorithm in sensor networks. The base station or a super node generates the Huffman code for compression of sensor data and broadcasts it into sensor networks. Every sensor that receives 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%.

Key words : sensor network, feedback diffusion, data compression, huffman code

· 이 논문은 2010년 교육과학기술부로부터 지원(지역거점연구단육성사업/충북BIT연구 중심 대학 육성 사업단)과 한국연구재단의 지원을 받아 수행된 기초연구사업(No. 2009-0080279)의 결과임

† 학생회원 : 충북대학교 정보통신공학과
mhyeo@netdb.cbnu.ac.kr
sergei@netdb.cbnu.ac.kr
yjcho@netdb.cbnu.ac.kr
†† 종신회원 : 충북대학교 정보통신공학과 교수
yjs@chungbuk.ac.kr
(*Corresponding author*임)

논문접수 : 2009년 1월 19일
심사완료 : 2010년 1월 12일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 데이타베이스 제37권 제2호(2010.4)

1. 서 론

최근 무선 센서 네트워크는 환경 감시, 스마트 공간, 의료 기술, 농업 기술 분야와 같은 다양한 분야의 응용으로 각광받고 있다[1-3]. 이러한 센서 네트워크는 온도, 습도, 진도 등과 같은 주변 환경의 유용한 정보를 수집하고, 보다 다양한 처리를 위해서 수집한 정보를 기지국으로 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[4]. 이러한 데이터 연관성을 이용하여 센서 데이터를 압축하기 위한 많은 연구들이 진행되었다. 시간적인 제한 기법은 센서 데이터의 오차 허용치를 미리 정하고, 수집한 데이터의 변화가 오차 범위보다 작은 경우 데이터를 기지국으로 전송하지 않는다. 이때, 데이터를 수신하지 않은 기지국은 시간적인 유사도 관점에서 해당 센서 노드의 현재 데이터가 마지막으로 보고된 데이터의 오차 허용치 범위에 있다고 판단한다. 공간적인 제한 기법은 센서 노드의 통신 범위 혹은 네트워크 토폴로지 상에서 인접한 센서 노드의 평균값 혹은 센서 데이터와의 유사성을 이용하여 데이터 전송의 수를 줄인다. 기존 데이터 압축 기법의 경우, 대부분 공간적인 연관성을 활용하기 위해서 통신 범위로 제한된 지역적인 데이터만을 이용한다는 단점을 가지고 있다.

본 논문에서는 전역적인 데이터 분포 특성을 데이터 압축에 활용하기 위한 피드백 배포 기법을 제안한다. 피드백 배포 기법이란 센서 네트워크로부터 수집된 데이터를 기반으로 압축에 활용될 정보를 생성하고, 센서 네트워크로 그 정보를 배포하는 기법이다. 이 정보를 수신한 센서 노드들은 해당 센서 데이터를 압축하고, 압축된 데이터를 기지국으로 전송한다. 피드백 배포를 통해 효율적인 데이터 압축을 수행하기 위해서는 피드백 배포 시 소요되는 통신 비용을 줄이고, 피드백 정보를 통한 데이터 압축 효율을 높이는 것이 중요하다. 본 논문에서는 센서 데이터의 특성을 고려하여 효과적으로 피드백 정보의 크기를 줄이고, 일정 이상의 압축 효율을 유지하기 위해서 피드백 정보의 재배포 시점을 결정하는 기법을 제안한다. 제안하는 기법의 우수성을 보이기 위해서 시뮬레이션을 통해 성능 평가를 수행하였으며, 센서 데이터가 약 40% 압축되었고, 이를 통해 전체 네트워크의 수명이 약 30% 연장되었다.

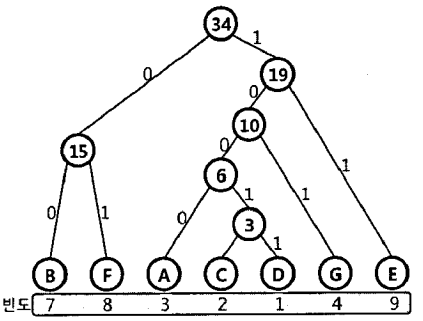
본 논문의 구성은 다음과 같다. 제2장에서는 관련연구로 센서 네트워크의 특성을 이용한 데이터압축 기법과 본 논문에서 활용 하고자 하는 허프만 코딩의 특성을 살펴본다. 제3장에서는 본 논문에서 제안하는 피드백 배포 기법을 정의하고, 피드백 정보를 이용한 데이터 압축 및 수집기법을 산술적으로 분석한다. 제4장에서는 본 논문에서 제안하는 허프만 코딩 기반 데이터 압축 기법의 특징과 상세한 과정을 기술한다. 제5장에서는 성능평가와 분석을 통해 제안하는 기법의 우수성을 보이고, 제6장에서 결론과 향후 연구 방향을 제시한다.

2. 관련연구

2.1 센서 데이터 압축 기법

센서 네트워크의 데이터는 네트워크 특성상 시간적 혹은 공간적 연관성을 지닌다[4]. 이러한 데이터 연관성을 이용하여 센서 데이터를 압축하기 위한 많은 연구들이 진행되었다. 센서에서 수집한 데이터는 환경적인 특성상 시간적 혹은 공간적 연관성을 지닌다. 이러한 센서 데이터의 특성을 활용하여 데이터 압축에 활용할 수 있다. 시간적인 특성을 이용한 압축 기법은 센서 노드에서 마지막으로 보고된 데이터를 기준으로 현재 데이터의 변화가 발생한 경우 싱크 노드로 데이터를 전송하는 기법이다. 이때, 기지국은 보고되어지지 않은 데이터의 경우 이전과 변화가 없다고 가정한다. [5]에서 제안하는 시간적인 제한 기법은 센서 데이터의 오차 허용치를 미리 정하고, 수집한 데이터의 변화가 허용치보다 큰 경우만 데이터를 기지국으로 전송한다.

공간적인 특성을 이용한 데이터 압축 기법은 이웃 노드들의 데이터와의 유사성을 이용한다. [6]에서는 평균 연산자를 이용하여 압축을 수행한다. 모든 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 비교하여 자신의 값과 동일한 경우 기지국으로 데이터를 전송하지 않는다. 데이터 압축 기법의 효율을 높이기 위해서 클러스터링과 트리 구조와 같은 인-네트워크 처리(in-network processing)를 이용하는 방법이 제안되었다[7,8]. 공간적으로 인접한 센서들의 데이터가 서로 연관성이 높은 경우, 각각의 센서들이 기지국으로 데이터를 전송하는 것보다 네트워크 내부에서 데이터를 병합한 후 전송하는 것이 더 효율적이다. [7]은 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.



(a)

심볼	허프만 코드
A	1000
B	00
C	10010
D	10011
E	11
F	01
G	101

(b)

그림 1 허프만 코딩을 이용한 가변길이 코드 할당

2.2 허프만 코딩

허프만 코딩[12]은 엔트로피 인코딩 알고리즘의 하나로써 무손실 압축기법이다. 이 기법은 심볼의 발생 빈도에 따라 최적의 가변길이 코드를 할당하는 방법을 제공한다. 그림 1은 허프만 코딩을 이용한 가변길이 코드 할당의 예를 나타낸다. 코드 생성은 심볼의 발생 빈도가 적은 쌍을 연결하고 빈도를 합하는 방식으로 그림 1(a)와 같은 확장 이진트리를 생성한다. 이때, 트리의 단말 노드는 심볼을 나타내며, 중간 노드는 빈도의 합을 나타낸다. 이렇게 생성된 트리는 심볼의 가중치에 따라 트리에서의 깊이가 결정되고, 그림 1(b)와 같이 루트 노드에서 해당 심볼의 노드까지 도달하는데 거쳐가는 비트의 셋이 해당 노드의 코드로 할당된다. 따라서 발생빈도가 높은 심볼에 대하여 길이가 짧은 비트코드를 할당하고, 발생빈도가 상대적으로 낮은 심볼에 대하여 길이가 긴 비트코드를 할당한다.

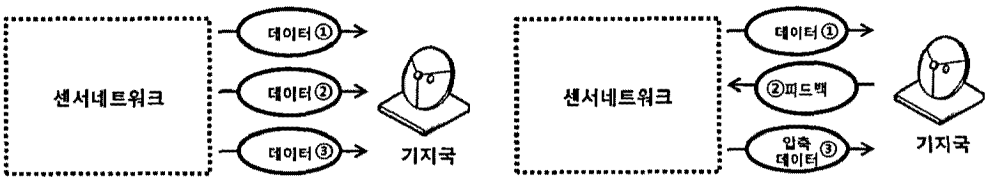
센서 네트워크에서 발생하는 에너지 소모의 상당부분이 데이터 전송에서 발생한다. 데이터 전송에 발생하는 에너지의 크기는 데이터를 전송/수신하는 횟수, 전송하는 데이터의 크기, 전송하는 거리 등의 다양한 요소가 복합적으로 작용한다. 본 논문에서는 이러한 데이터 전송에서 발생하는 에너지 소모의 특징을 고려한 데이터 압축기법을 제안한다. 기본적으로 센서 네트워크에서 발생하는 데이터는 환경적 특성으로 인해 발생하는 데이터가 급격히 변화하지 않는다. 예를 들어, 주간의 온도를 측정해보면 온도 1도가 변화하는데 약 1~2시간이 소요된다. 뿐만 아니라, 지리적으로 인접한 센서 노드들은 유사한 값들을 수집할 가능성이 높다. 즉, 센서 네트워크에서 수집되는 데이터들은 같은 시간대에 비슷한 값이 지속적으로 발생하는 특징을 보인다. 반면에 나머지 영역의 값들은 발생 하지 않거나 발생 빈도가 현저히 낮다. 이러한 특성을 반영해 효과적으로 데이터를 압축할 수 있는 방법으로 본 논문에서는 허프만 코딩 기법을 이용한다. 앞서 설명한 바와 같이 허프만 코딩은 데이터의 발생 빈도에 따라 가변 길이의 코드를 반환한다. 즉, 발생 빈도가 높은 값에는 짧은 길이의 코드가 할당되고, 발생 빈도가 낮은 값에는 상대적으로 긴 길이의 코드가 할당된다. 따라서 센서 네트워크에서 수집되는 데이터 중에서 발생빈도가 높은 데이터를 짧은 길이의 코드로 인코딩함으로써 에너지 효율적인 데이터 송/수신을 가능하게 한다.

3. 피드백 배포 기법

기존 센서 데이터의 공간적인 연관성을 이용한 데이터 압축 기법은 그림 2(a)와 같이 기지국이 센서 네트워크로부터 데이터를 수집하는 구조이다. 즉, 센서 네트워크 내부에서 각 알고리즘의 특성에 따라 압축이 이루어지며 압축된 데이터가 기지국으로 전송된다. 관련 연구에서 언급했듯이 기존 데이터 압축 기법은 각 센서 노드의 과거 데이터를 참조하거나 센서 주변 노드의 정보를 이용한다.

본 논문에서는 기지국(혹은 슈퍼노드)과 센서 네트워크의 상호작용을 통해 데이터 압축을 수행하는 방법을 제안한다. 그림 2(b)와 같이 수집된 데이터를 이용하여 데이터 압축에 참조가 되는 정보(이하 "피드백")를 센서 네트워크로 배포함으로써 센서 노드들은 자신의 데이터를 효과적으로 압축한다. 이러한 접근을 "피드백 배포 기법"이라고 정의한다. 기존 기법과 달리 센서 네트워크의 전역적인 데이터 특성을 반영하는 장점을 가지고 있다.

N 개의 센서 노드로 구성된 TAG기법[13]을 기반으로 트리 형태로 연결된 센서 네트워크가 있다고 가정하자. 각 센서 데이터 v는 압축 알고리즘에 의해 v'로 압축되고, 기지국으로 데이터를 전송하는 통신 비용은 식(1)과 같다. 이때, $avgDist$는 각 센서 노드로부터 기지국까지의 평균 거리를 나타내며, 각 센서 노드는 매주기마다 $log_2(v')$비트 크기의 데이터를 기지국으로 전송한다.



(a) 기존 데이터 수집 방법 (b) 피드백 정보를 이용한 데이터 수집 방법

그림 2 피드백 배포

본 논문에서 제안하는 피드백 배포 기법의 통신 비용은 식 (2)와 같다. N_{nl}은 단말 노드를 제외한 나머지 노드의 수를 나타내며, v''는 피드백 배포 기법을 통해 압축된 센서 데이터를 의미한다. 데이터 수집에 필요한 전체 비용(C_f)은 피드백 정보를 생성하기 위한 TAG기반의 데이터 수집 비용(C_{tag})과 피드백 정보 배포 비용(C_{fd}), 피드백 정보(F)를 통해 압축된 데이터의 수신 비용(C_{fc})의 합으로 나타낸다. 이때, T_c와 T_f는 각각 데이터의 수집 횟수와 피드백 정보의 배포 횟수를 의미한다.

$$C_{fd} = N_{nl} \cdot log_2(F),\ C_{fc} = N \cdot avgDist \cdot log_2(v'')$$

$$C_f = \frac{C_{tag} + T_f \cdot C_{fd} + T_c \cdot C_{fc}}{T_c}$$

식 (3)은 효율적인 피드백 배포 기법의 설계를 위한 통신 비용의 계산식을 나타낸다. 피드백 배포 기법을 효율적으로 설계하기 위해서 압축 기법의 수집 이득 $G_{fc}(T_c)$을 증가시키고, 순수 피드백 정보 배포 비용 $C'_{fd}(T_f, F)$을 감소시켜야 한다. 순수 피드백 정보 배포 비용은 F와 T_f에 비례하므로 본 논문에서는 피드백 정보의 크기를 줄이기 위한 기법과 피드백 정보의 배포 횟수를 줄이기 위한 기법을 연구한다.

$$C_{tag} > C_f$$

$$T_c \cdot C_{tag} > C_{tag} + T_f \cdot C_{fd} + T_c \cdot C_{fc}$$

$$T_c \cdot (C_{tag} - C_{fc}) - (C_{tag} + T_f \cdot C_{fd}) > 0$$

$$G_{fc}(T_c) = T_c \cdot (C_{tag} - C_{fc}),\ C'_{fd}(T_f) = (C_{tag} + T_f \cdot C_{fd})$$

$$G_{fc}(T_c) - C'_{fd}(T_f, F) > 0,\ C'_{fd}(T_f, F) \propto T_f, F$$

4. 제안하는 허프만 코딩 기반의 피드백 배포 기법

본 장에서는 피드백 배포 기법을 효율적으로 사용하기 위해서 피드백 정보의 크기를 줄이고, 배포 횟수를 줄이기 위한 방법을 기술한다. 피드백 배포 기법은 센서 데이터를 비트로 표현하여 데이터의 크기를 줄인다. 최소의 비트를 이용하여 데이터를 효과적으로 표현하기 위한 대표적인 방법으로 허프만 코딩 기법이 있다. 하지만, 허프만 코딩을 위하여 코드 테이블을 센서 네트워크로 배포하는 것은 많은 통신 비용을 필요로 한다. 본 논문에서는 피드백 정보의 크기를 줄이기 위해서 먼저 센서 데이터의 특징을 분석하고, 이를 고려하여 피드백 정보의 크기를 줄인다. 또한, 수집된 데이터의 압축되는 정도를 이용하여 피드백 정보의 갱신 시점을 결정한다.

4.1 피드백 정보 생성

본 논문에서는 기본적으로 센서 데이터의 빈도를 기반으로 한 피드백 정보 생성 기법을 제안한다. 일반적으로 센서 데이터는 시공간적 연관성을 가지고 있으므로 동일한 시공간에서 발생하는 데이터는 거의 동일하다. 따라서 특정 데이터 값에 대해 높은 빈도수로 발생할 가능성이 크다. 이때, 빈도가 높은 센서 데이터를 비트로 표현함으로써 전체 데이터의 압축 효율을 높인다. 이와 같은 방법은 데이터의 빈도를 이용하는 허프만 코딩(Huffman Coding) 기법과 일맥 상통한다. 하지만, 허프만 코드를 피드백 정보 형태로 배포하기 위해서 *{센서 데이터값, 비트값}* 쌍으로 구성된 테이블을 센서 네트워크로 배포하는 것은 상당한 에너지 소모를 발생시킨다. 따라서, 피드백 정보의 크기를 줄이기 위해서 센서 데이터의 특징을 이용하여 가벼운 피드백 정보를 생성한다.

특징(1). 센서 데이터는 일정 범위의 오차를 가지고 있다.

응용에 따라 센서 데이터의 오차 허용 범위를 정할 수 있으며 다음과 같은 식 (4)를 이용함으로써 오차 허용 범위 내에 있는 센서 데이터를 동일한 센서 데이터로 그룹화한다. 이때, ε는 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.

$$|v_{real} - v_{base}| \le \varepsilon$$

특징(2). 일부 소수의 센서 노드는 완전히 잘못된 값을 센싱한다.

센서 노드는 오류를 발생시키기 쉬운 환경에 배포된다. 따라서, 센서 자체의 오류가 발생하거나 배포 환경의 상황에 따라 오류가 발생한다. 따라서 다른 센서 데이터의 경향성을 따르지 않는 데이터를 보고하기도 한다. 따라서 빈도가 임계치 f_{cutoff} 보다 작은 센서 데이터는 제거함으로써 피드백 정보에 참여하는 센서 데이터

의 수를 줄일 수 있다.

그림 3은 피드백 정보를 생성하는 과정을 나타낸다. 측정값의 크기에 따른 정렬을 통해 오차 허용 범위 ε에 포함되는 센서 데이터를 그룹화하고, 센서 데이터의 빈도가 임계치 f_{cutoff} 이하인 센서 데이터 그룹은 제거한다. 그 다음, 허프만 코딩을 이용하여 비트로 변환하고, *{센서데이터값, 비트값}*으로 구성된 허프만 코드 테이블(Huffman Code Table)을 생성한다. 성능 파라미터인 ε와 f_{cutoff}은 응용 분야가 요구하는 정확도 레벨 또는 데이터 분포의 특성에 따라 정의될 수 있으며, 성능 평가의 실험 모델에서 최적의 파라미터를 찾기 위해서 값을 변화시키면서 실험을 수행하였다.

그림 4는 피드백 정보 생성의 예를 나타낸다. 예제의 응용 분야는 ±1 의 오차허용 범위를 가지고 있고, 3개 미만의 측정 빈도를 보이는 측정값을 허프만 코드 작성에 제외한다고 가정하자. 기지국은 센서 노드 S1~S7으로부터 측정값을 수집하였다. 먼저, 오차허용범위를 이용하여 측정값을 {10.5, 21.5, 1000}으로 그룹화하고, 특정 임계치 이하의 측정값을 제외한 나머지를 이용하여 허프만 코드를 작성한다.

4.2 피드백 정보를 통한 센서 데이터 압축 과정

각 센서 노드로 전달된 피드백 정보는 각 센서 노드가 센서 데이터를 양자화하는데 활용된다. 그림 5는 기지국에 수집된 데이터를 바탕으로 만든 피드백 정보를 배포하고, 피드백 정보를 이용해 수집된 데이터를 압축하여 수집하는 과정의 예를 나타낸다. 먼저 기지국에서 앞서 설명한 피드백 정보 생성 알고리즘을 이용해 센서 네트워크에서 수집된 데이터를 바탕으로 피드백 정보를 생성하고, 그림 5의 과정 1과 같이 센서 네트워크로 피드백 정보를 배포한다. 각 센서 노드들은 이 피드백 정

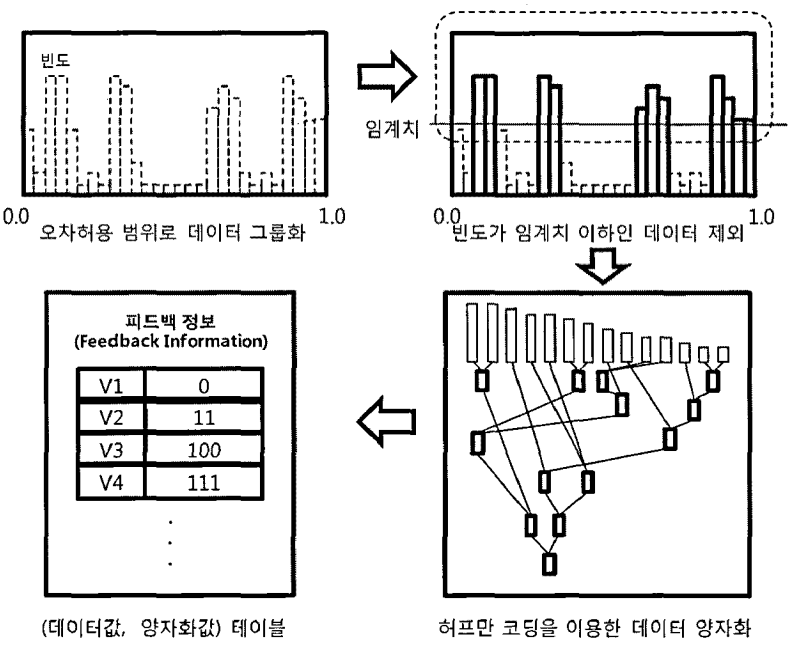


그림 3 피드백 정보 생성 과정

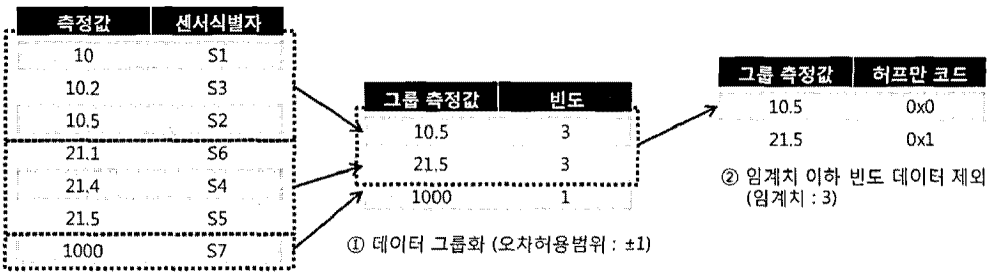


그림 4 피드백 정보 생성의 예

보를 저장하고 있으며 그림 5의 과정 2를 수행한다. 센서 자신이 측정한 센서 데이터가 피드백 정보를 이용해 압축가능 여부를 판단한다. 만약 측정 데이터가 피드백 정보에 포함되어 있는 값과 ε 이내의 오차를 가지는 경우, 피드백 정보에 있는 압축 정보를 기지국으로 전송한다. 만약 피드백 정보에 포함되어 있지 않다면, 특징(2)에 의해 데이터 전송에서 제외된다. 하지만, 응용 분야가 모든 데이터 수집을 요구하는 경우, 압축되지 않은 정보를 그대로 전송함으로써 수집 데이터의 정확도를 향상시킬 수 있다. 예를 들면, 건물 관리를 위하여 건물의 주요 20개 지역에 설치된 20개의 센서로부터 주기적으로 측정값을 수집하는 경우, 주변과 유사한 측정값을 가진 센서들은 압축을 수행하고, 빈도수가 낮은 측정값에 대해서는 원본 데이터를 그대로 수집한다. 이때, 상대적으로 빈도수가 낮은 센서 데이터만 원본 데이터로 수집하기 때문에 통신 비용이 크지 않다.

4.3 피드백 정보의 갱신

특정 시점에서 수집된 센서 데이터를 이용하여 피드백 정보를 생성하고, 배포된 피드백 정보를 통해 센서 데이터를 압축한다. 하지만 피드백 정보를 배포한 시점에서 시간이 지날수록 센서 데이터가 변하기 때문에 압축 효율이 낮아지게 된다. 만약, 압축 효율을 높이기 위해서 매주기마다 피드백 정보를 배포한다면 많은 에너지가 소모된다. 반대로 피드백 정보를 갱신하지 않으면, 압축 효율이 저하되는 문제점을 가지고 있다. 제안하는 피드백 배포 기법은 적절한 갱신 시점을 찾기 위해서 기지국에서 수집되는 데이터의 수를 계산한다. 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. 그림 6은 제안하는 허프만 코딩 기반의 피드백 배포 기법의 의사 코드(pseudo code)이다.

5. 성능평가 및 분석

5.1 실험 환경

제안하는 기법의 우수성을 보이기 위해 일반적인 데이터 수집 기법과 시뮬레이션을 통해 성능을 비교평가하였다. 표 1은 시뮬레이션에 사용된 환경 변수이다. 시뮬레이션에서 사용된 데이터는 미국 워싱턴주의 온도 데이터를 활용하였으며, 그림 8과 같은 온도 분포를 보인다[9]. 공간적인 연관성을 만들기 위해서 데이터 모델의 위상차를 두어 각 센서 노드가 서로 다른 데이터를 수집하도록 설정하였으며, 센서 노드는 그림 7과 같은 형태의 네트워크 토폴로지를 형성한다. 라우팅 트리의 레벨이 커짐에 따라 네트워크를 구성하는 센서 노드의 수는 24개~440개로 변경된다. 센서 노드의 메시지 전송에 소모되는 에너지 모델은 $\{$*메시지 크기*$\} \times (\{$*전송 비용*$\} + \{$*증폭 비용*$\} \times \{$*거리*$\})$ 이며, 전송 비용은 50nJ/b, 증폭 비용은 100pJ/b/m^{2}으로 설정하였다. 메시지 수신에 소모되는 에너지 모델은 $\{$*메시지 크기*$\} \times \{$*수신 비용*$\}$ 이며, 수신 비용은 50nJ/b으로 설정하였다[10,11]. 수집된 센서 데이터의 정확성을 보장하기 위해서 피드백 정보를 이용할 수 없는 데이터의 경우 양자화되지 않은 본래 데이터를 그대로 수집하는 정책을 사용하였다.

5.2 피드백 정보의 크기 비교

그림 9와 그림 10은 네트워크의 규모에 따른 기지국에서 생성된 제안하는 기법의 피드백 정보의 크기와 수집된 센서 데이터의 크기를 비교 평가한 결과이다. 피드백 기법은 본 논문에서 처음 제안되는 기법이기 때문에, 고정된 크기의 비트를 이용하여 센서 데이터를 압축하는 기법을 Naïve 기법이라고 정의하고, 본 논문에서 제안하는 허프만 코딩 기법과 성능을 비교하였다. 그림 8의 결과와 같이 Naïve 기법을 사용하는 피드백 기법의 경우 네트워크의 규모가 커짐에 따라 발생하는 표본이 증가하여 피드백 정보 배포비용은 급격히 증가한다. 하지만 제안하는 기법의 경우 네트워크의 규모가 증가하

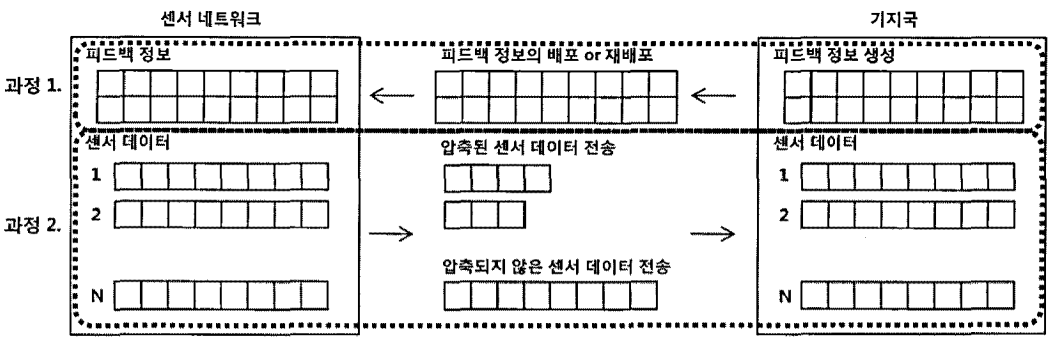


그림 5 피드백 정보 배포와 압축 데이터 수집 과정

```
입력인자 :
sensors // 네트워크를 구성하는 센서 노드
cutOff // 측정값의 발생 빈도 최소 임계치
hitRate // 사용자에 의해 정의된 양자화된 데이터 최소 수집 정도

출력인자 :
reading // 센서의 수집 데이터
feedbackInfo // 수집된 정보를 바탕으로 생성한 최종 피드백 정보

코드 :
sensors = [s0, s1, s2, ... , sn]

round = 0    // 질의 처리 횟수
freqTable = 0 // 측정값 발생 횟수 누적 테이블
WHILE network == live DO

    /* 피드백 정보 생성을 위해 모든 수집 데이터 기지국으로 전송 */
    IF round = 0 THEN
        sensors.sensing()
        sensors.sendToBasestation(sensors.readings)
    END IF

    /* 수집된 데이터를 바탕으로 피드백 정보 테이블 생성 */
    IF basestation.hitCount < hitRate // round = 0   THEN // 피드백 정보 이용 효율의 유효성 판단
        freqTable[] = null
        basestation.hitCount = 0
        FOR EACH data = basestation.recvData[i] DO
            IF freqTable.find(data, ε) THEN
                freqTable.set(data, ++freq)
            ELSE
                freqTable.add(data, 1)
            END IF
        END FOR

        /* 발생 빈도가 낮은 데이터 제거 */
        FOR EACH entry = freqTable[i] DO
            IF entry.freq < cutOff   THEN
                freqTable.del(i)
            END IF
        END FOR

        basestation.createFeedbackInfo(freqTable)
        basestation.broadcastToNetwork(basestation.feedbackInfo)
    END IF

    /* 피드백 정보를 이용한 수집 데이터 전송 */
    FOR EACH s = sensors[i] DO
        s.sensing()
        IF code = s.freqTable.findCode(s.reading) THEN
            s.sendToBasestation(code)
            basestation.hitCount++
        ELSE
            s.sendToBasestation(s.reading)
        END IF
    END FOR

    round++
END WHILE
```

그림 6 의사 코드

표 1 성능 평가 환경

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

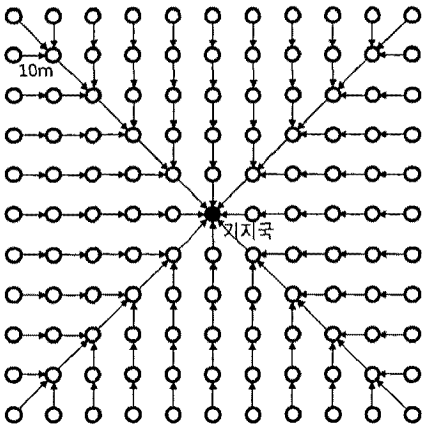


그림 7 TAG 기반의 네트워크 토폴로지

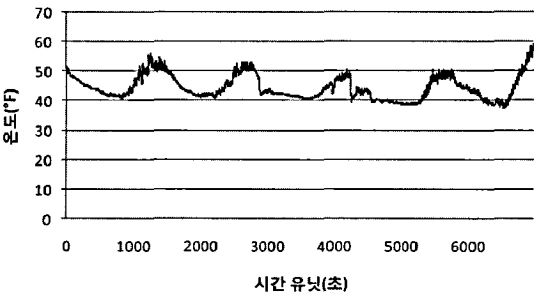


그림 8 시뮬레이션에서 사용한 데이터 모델

더라도 피드백 정보의 크기에는 큰 변화를 보이지 않는다. 그 이유는 빈도수가 낮은 데이터들을 배제시켜 피드백 정보를 생성하기 때문이다. 그림 8에서는 기존의 연구와 제안하는 기법의 피드백 정보를 통해 전송되는 데이터량을 비교한 결과 제안하는 기법을 사용했을 경우 약 40%의 데이터 전송량을 줄어드는 것을 확인할 수 있다. 이는 허프만 코딩의 이점에 의해 발생빈도가 높은 데이터에 비트수가 적은 코드를 할당하였기 때문이다.

5.3 피드백 정보의 갱신 시점

그림 11은 양자화 정보의 Hit율에 따른 네트워크의 수명을 변화를 실험한 결과이다. Hit율을 높게 설정하면 정확한 양자화 정보를 활용할 수 있지만 양자화 정보의

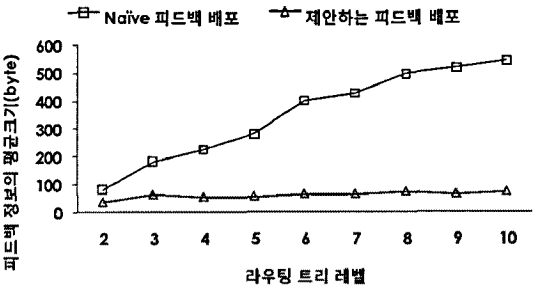


그림 9 허프만 코딩 테이블의 크기

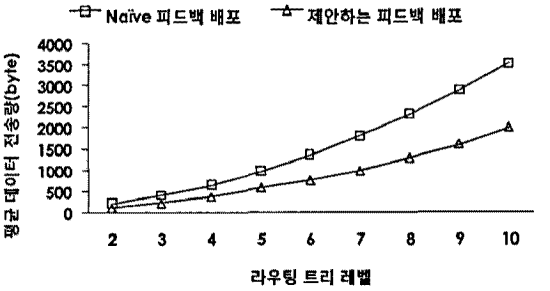


그림 10 평균 데이터 전송량

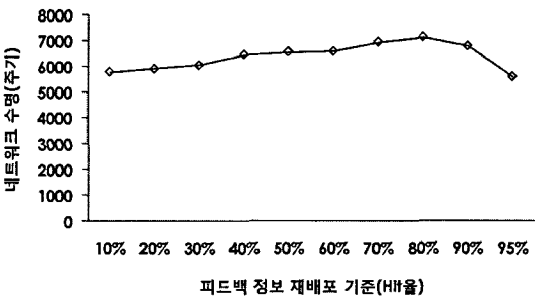


그림 11 양자화 정보 Hit율에 따른 네트워크 수명 비교

배포가 빈번하게 발생하고, Hit율을 낮게 설정하면 양자화 정보는 최신 정보를 반영하지 못하지만 배포의 빈도가 낮아지게 된다. 실험 결과 약 83% 근처에서 가장 좋은 성능을 보였고, 그보다 더 높게 설정하면 양자화 정보의 배포가 크게 증가하여 오히려 네트워크 수명에 악영향을 미치는 것을 확인하였다. 이때, 센서 네트워크에 포함된 모든 센서 노드로부터 데이터를 수집하는 것을 1주기라고 정의한다.

5.4 압축 성능 비교

그림 12와 표 2는 TAG[13]와 같이 트리 기반의 네트워크 토폴로지를 가진 센서 네트워크에서 센서 데이터를 그대로 수집하는 FM(Flooding Method) 기법과 제안하는 피드백 배포 기법의 네트워크 수명을 비교한 결과이다. 기존 압축 기법의 병행 사용 가능함을 평가하

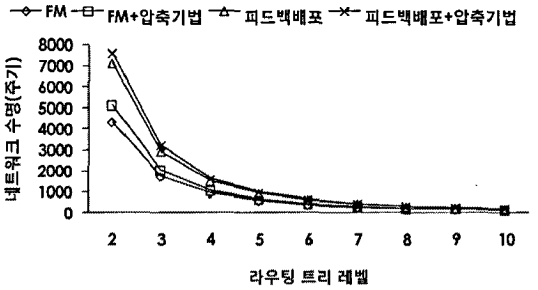


그림 12 네트워크 수명 비교

표 2 성능 향상도(FM+압축기법 vs. 제안하는 피드백 배포기법)

라우팅 트리 레벨	FM+압축기법	피드백배포	성능 향상(%)
2	5151 주기	7133 주기	27.79
3	2051 주기	2934 주기	30.10
4	1042 주기	1505 주기	30.76
5	617 주기	892 주기	30.83
6	403 주기	576 주기	30.03
7	271 주기	400 주기	32.25
8	189 주기	275 주기	31.27
9	137 주기	204 주기	32.84
10	103 주기	149 주기	30.87

기 위해서 [5]에서 사용된 압축 기법을 적용하였다. x축의 값은 라우팅 트리의 레벨을 의미한다. 실험결과 FM에 비해 제안하는 기법이 약 30%(({제안하는 기법의 네트워크 수명}-{FM 기법의 네트워크 수명})/{제안하는 기법의 네트워크 수명}*100)의 네트워크 수명 향상을 보였으며, 각 기법에 압축기법을 적용한 경우 추가적인 에너지 사용의 효율이 발생하였다. 이는 제안하는 기법의 특성상 기존 압축 알고리즘과 병행 사용이 가능하다는 것을 의미하며, 전체적으로 피드백 배포를 통한 압축을 수행함으로써 데이터 전송에 소모되는 에너지 소모가 크게 줄었다는 것을 보여준다.

6. 결 론

본 논문은 센서 데이터를 효율적으로 압축하기 위한 허프만 코딩 기반의 피드백 배포 기법을 제안한다. 공간적인 특성을 이용한 대부분의 압축 기법은 통신 범위의 제약으로 인해 지역적인 데이터 특성만을 반영하는 단점이 있다. 제안하는 기법에서 기지국(혹은 슈퍼 노드)에 의해 생성된 피드백 정보는 센서데이터의 전역적인 분포 특성을 반영하고 있기 때문에 효율적인 데이터 압축을 수행한다. 시뮬레이션을 통한 실험 결과, 기존 데이터 수집 기법에 비해 네트워크의 수명이 약 30% 향상되었고, 기존 압축 기법과 병행 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 제안하는 기법을 센서 네트워크 응용 시스템에 적용하는 것이다.

참 고 문 헌

[1] D. Estrin, L. Girod, G. Pottie and M. Srivastava, "Instrumenting the World with Wireless Sensor Networks," *Proc. of International Conference Acoustics, Speech, and Signal Processing*, pp. 2033-2036, 2001.

[2] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors," *Proc. of Comm. ACM*, pp. 51-58, May 2000.

[3] I.F. Akyildiz, W. Su, Y. Sankarasubramanism and E. Cayirci, "A Survey on Sensor Networks," *Proc. of IEEE Communications Magazine*, 2002.

[4] M. H. Yeo, M. S. Lee, J. G. Park, S. J. Lee, J. S. Yoo, "A Data-Centric Algrithm for Reducing Network Traffic in Wireless Sensor Networks," *Journal of KIISE: Information Networking*, vol.35, no.2, pp.139-148, Apr. 2008. (in Korean)

[5] M. Sharaf, J. Beaver, A. Labrinidis, P. Chryanthis, "Tina: A scheme for temporal coherency-aware in-network aggregation," *Proc. of the 2003 ACM Workshop on Data Engineering for Wireless and mobile Access*, Sep. 2003.

[6] X. Meng, L. Li, T. Nandagopal, S. Lu, "Event contour: An efficient and robust mechanism for tasks in sen-sor networks," *Proc. of Technical report*, 2004.

[7] S. Pattem, B. Krishnamachari, R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," *Proc. of International Conference on Information Processing in Sensor Networks*, pp.28-35, 2004.

[8] D. Petrovic, R. Shah, K. Ramchandran, J. Rabaey, "Data funneling: Routing with aggregation and compression for wireless sensor networks," *Proc. of the 2003 IEEE Sensor Network Protocols and Applications*, pp.156-162, May 2003.

[9] Live from Earth and Mars (LEM) Project, http://www-k12.atmos.washington.edu/k12/grayskies/, 2006.

[10] W. Heinzelman, "Application-Specific Protocol Architectures for Wireless Networks," PhD dissertation, *Massachusetts Inst. Of Technology*, 2000.

[11] X. Tang, J. Xu, "Extending Network Lifetime for Precision-Constrained Data Aggregation in Wireless Sensor Networks," *Proc. IEEE INFOCOM*, pp.1-12, Apr. 2006.

[12] D. A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," *Proc. of the I.R.E.*,

pp.1098-1102, 1952.
[13] S. Madden, M.J. Franklin, J.M. Hellerstein, and W. Hong, "TAG: A Tiny Aggregation Service for Ad Hoc Sensor Networks," *Proc. Usenix Fifth Symp. Operating Systems Design and Implementation*, pp.131-146, Dec. 2002.

여 명 호
2004년 2월 충북대학교 정보통신공학과 공학사. 2006년 2월 충북대학교 정보통신공학과 공학석사. 2010년 2월 충북대학교 정보통신공학과 박사. 2010년 2월~현재 국방과학연구소 연구원. 관심분야는 메인 메모리 기반 데이터베이스 시스템, 시공간 데이터베이스 시스템, 무선 센서 네트워크 등

성 동 욱
2005년 2월 충북대학교 정보통신공학과 공학사. 2007년 2월 충북대학교 정보통신공학과 공학석사. 2007년~현재 충북대학교 정보통신공학과 박사과정. 관심분야는 무선 센서 네트워크, 데이터베이스 시스템, LMS/LCMS, FLASH 메모리 저장 시스템

조 용 준
2007년 2월 충북대학교 정보통신공학과 공학사. 2009년 2월 충북대학교 정보통신공학과 공학석사. 2009년 3월~현재 ㈜동일기술공사 기술연구소 연구원. 관심분야는 RFID, 무선 센서 네트워크, 데이터베이스 시스템, 위치 기반 서비스

유 재 수
정보과학회논문지 : 데이타베이스
제 37 권 제 1 호 참조