동적 비트할당 기반 센서데이타 압축 기법

(A Sensor Data Compression Algorithm based on Dynamic Bit-assignment Techniques)

이 석 재 † 박 현 호 †† 여 명 호 †† 송 석 일 ††† 유 재 수 ††††
(Seok Jae Lee) (Hyun Ho Park) (Myung Ho Yeo) (Seok Il Song) (Jae Soo Yoo)

요 약 대부분의 센서네트워크 응용들은 일정기간 동안 센서데이타를 지속적으로 수집·분석한다. 이 때 모든 센서가 계속해서 서버로 데이타를 전송하기 때문에 많은 통신비용이 소모된다. 따라서 통신비용 감소를 위한 기술은 배터리를 이용해 동작하는 무선 센서에 기반한 센서네트워크의 수명 연장을 위한 가장 핵심적인 기술이라 할 수 있다. 본 연구에서는 센서데이타를 압축·전송함으로써 통신비용을 감소시키기 위한 동적 비트할당 기법에 기반한 센서데이타 압축 기법을 제안한다. 제안하는 기법은 일정기간 동안 수집한 센서데이타의 분포를 분석해 센서데이타를 비트로 표현할 수 있는 정보를 각 센서에 배포하고, 데이타를 압축해 수집함으로써 통신비용을 효과적으로 줄일 수 있다.

키워드 : 센서데이타, 압축, 센서네트워크, 동적 비트할당

Abstract Most of the sensor applications collect and analyze sensor data within a given period of time. When sensor send a data to sink, it spend many communication cost. Accordingly, a compression algorithm is one of the most critical issues for the communication cost decrease in sensor fields. In this paper, we propose an algorithm for compressing sensor data using the dynamic bit assignment technique. In our algorithm, sink collect sensor data within a short period of time and make bit assign information. Then sink send the information to sensor. Finally, sensors compresssensing data and send to sink.

Key words : Sensor data, Compression, Sensor network, Dynamic bit assignment

1. 서 론

일반적으로 무선 센서 네트워크(wireless sensor network)는 응용에서 관심을 갖는 대상의 발생, 변화, 또는 분포 등의 정보를 일정기간 동안 지속적으로 관찰하기 위한 목적으로 구축된다. 기반이 되는 센서는 프로세서, 메모리, 통신 장치, 센싱 장치, 그리고 배터리 등으로 구성된다. 또한 응용에 따라 GPS와 같은 장치를 추가적으로 탑재하기도 한다. 일정 지역에 배포해 일정기간 동안 동작하는 센서들의 수와 배포 범위는 응용의 요구사항에 따라 다양하게 결정될 수 있다. 이렇게 다양하게 만들어 질 수 있는 배포 환경에서 센서들이 센싱한 데이타를 수집하는데 가장 효과적인 방법으로 무선 센서 네트워크를 들 수 있다.

현재 무선 센서 네트워크 환경에서 가장 많은 연구가 이루어지고 있는 주제는 배터리로 동작하는 센서의 동작 수명을 늘이기 위한 다양한 기법들에 대한 것이다. 특히 센서에서 통신이 다른 처리 비용에 비해 매우 많은 전력을 소모하기 때문에 통신비용을 줄이기 위한 많은 연구들이 진행 중이다. 센서 데이타를 수집하는 과정에서 통신에 사용되는 전력소모를 줄이기 위한 센서 데이타 압축 기법은 압축을 위한 계산에 사용되는 전력소

· 이 논문은 2007년도 정부(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. R01-2006-000-10809-0)

† 정 회 원 : 충북대학교 정보통신공학과 Post-Doc
sjlee@netdb.cbnu.ac.kr
†† 학생회원 : 충북대학교 정보통신공학과
hhpark@netdb.cbnu.ac.kr
mhyeo@netdb.cbnu.ac.kr
††† 정 회 원 : 충주대학교 컴퓨터공학과 교수
sisong@chungju.ac.kr
†††† 종신회원 : 충북대학교 정보통신공학과 교수
yjs@cbnu.ac.kr
(Corresponding author임)

논문접수 : 2007년 10월 2일
심사완료 : 2008년 4월 21일

Copyright@2008 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.

모량보다 전송을 위한 통신에 사용되는 전력소모량이 더 크다는 점에 초점을 맞추어 연구되고 있다[1-5,10].

지금까지 연구된 대부분의 센서 데이타 압축 기법들은 센싱된 데이타에 웨이블릿 변환, 양자화, 가변길이 부호화 등과 같은 신호 압축 기법 또는BWT, LZE, LZ77과 같은 코드 압축 기법들을 적용하고 있다[6-9]. 이와 같은 연구들은 각 센서에서 측정한 데이타만을 가지고 압축을 적용하는 방법이기 때문에 전체 센서 네트워크에서의 데이타 발생 분포, 빈도 등에 대한 고려는 되어있지 않다. 그러나 기상 변화 측정을 위한 센서네트워크와 같은 응용에서는 전체 센서네트워크에서 발생하는 값이 특정 값 또는 좁은 범위의 값으로 한정되는 경우가 대부분이다. 이와 같은 경우 발생하는 데이타의 특성을 각 센서가 알고 있다면 보다 효과적인 데이타 압축 기법을 만들어 적용할 수 있다.

본 논문에서는 센서의 통신 비용소모를 감소시키기 위해 데이타를 동적으로 비트화해 압축하는 효과적인 압축 기법을 제안한다. 제안하는 압축 기법에서는 먼저 일정 주기 동안 전체 센서네트워크에서 센싱된 모든 값을 압축하지 않은 상태로 싱크(sink)노드로 수집한다. 싱크 노드는 발생한 값의 분포를 분석하고, 효과적으로 압축할 수 있는 비트 할당 정보를 생성해 모든 센서들에게 배포한다. 각 센서는 이 정보를 바탕으로 센싱한 데이타를 비트화 함으로써 전송할 데이타를 효과적으로 압축한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 기존 센서 데이타 압축 기법들을 분석한다. 3장에서는 제안하는 센서 데이타 압축 기법에 대해 자세히 설명한다. 4장에서는 센서데이타 수집에 제안하는 압축 기법을 적용한 경우와 압축 기법을 적용하지 않은 경우의 통신 에너지 소모량을 비교한 결과를 기술한다. 마지막으로 5장에서 결론을 맺고 향후 연구방향을 제시한다.

2. 관련 연구

2.1 일반적인 압축 기법

2.1.1 웨이블릿 변환 압축 기법

웨이블릿(wavelet) 관련 연구는 처음 순수 수학 분야, 물리학 분야, 그리고 공학 분야에서 서로 독립적으로 연구를 수행하여 왔고, 근래에 들어 이들 서로 다른 분야에서의 접근들이 종합되게 되었다.

신호의 웨이블릿 변환(transform)은 범위(scale 혹은 frequency)과 시간(time)이라고 하는 두 개의 변수에 의해 표현된다. 웨이블릿 변환은 푸리에(Fourier) 변환과 같이 기저함수(basis function)들의 집합으로 신호를 분해하여 표현한다는 점은 같지만, 그 기저함수가 훌륭한 시간-주기 지역성(time-frequency localization)을 나타내는 특성을 가지고 있다는 점에서 다르다. 영상처리 및 압축분야와 관련해서, 웨이블릿 변환은 인간의 시각 계통과 비슷한 다해상도 분해가 가능하며 점진적 전송이 가능하고 또한 초고속 영상 압축시 블록기반 부호화에서 주로 나타나는 블록화 현상이 없다는 점 등으로 해서 근래 각광받고 있으며, 또한 현재 추진 중인 MPEG-4 표준화 작업의 주목 받는 후보 중 하나로 자리 잡고 있다.

2.1.2 RLE 압축 기법

일반적으로 압축의 원리는 중복되는 코드를 줄이는 것으로, 이런 원리에서 출발한 압축 기법은 방법상의 차이에서 몇 가지로 나뉜다. 즉, 현재 사용되는 압축 방법 중 일반적으로 사용되는 기법은 패킹(packing), 스퀴징(squeezing), 크런칭(crunching), 스쿼싱(squashing) 등 4가지 방법이다.

패킹은 RLE(Run Length Encoding, 실행 길이 부호화)라고도 하며, 압축 기법 중에서 가장 간단한 방법이다. RLE는 파일내부를 검색하여 같은 문자가 반복되면 몇 개가 반복되는지 세어서 3개로 줄여버리는 방법을 사용한다. 즉, 연속되는 문자가 있을 경우 이를 한꺼번에 하나로 줄이는 방법이다. 그림 1은 RLE 압축의 방식을 간단히 나타낸 것이다. 데이타에 한 문자가 'aaaaa'로 연속되어 있다면 그것을 'aaaaa -> a5'로 간단히 나타낼 수가 있다. 위의 예의 20문자는 'a5b4c4d3e4'로 10문자로 줄일 수가 있을 것이다.

그림 1 RLE 압축 기법의 예

RLE압축 기법의 장점은 압축과 압축을 푸는 속도가 매우 빠르고 프로그래밍 하기가 쉬우며 프로그램의 크기 또한 아주 작게 만들 수 있다는 점이다. 특히 같은 문자가 계속 반복되는 자료를 압축할 때 높은 압축률을 자랑한다. 그러나 일반적인 경우에는 압축률이 낮다. RLE 압축 기법은 3바이트로 표현되기 때문에 적어도 4개 이상의 동일한 문자가 인접해 있을 경우에만 효과가 있다. 지금은 RLE 압축 기법은 거의 사용하지 않고, 대부분 LZW방식을 사용하고 있다.

2.1.3 LZW 압축 기법

CCITT의 V.42vis에 채용되어 있는 LZW(Lempel Ziv Welch) 압축 기법은 1978년 이스라엘 Lempel과 Ziv가 처음으로 제안한 것을 1985년 현재 유니시스사의 전신인 스페리사의 Welch가 수정구현한 압축 기법이다. 이 기법은 입력 데이타 길이를 가변으로 하고 출력부호의 길이를 고정한 기법으로서 데이타 압축률이 높으며

내부 연산량이 작기 때문이 압축수행속도 측면에서는 현재까지 가장 빠른 것으로 평가되고 있다.

처음 제안된 LZW 알고리즘은 가변길이의 입력문자열을 모두 12비트의 고정 길이로 2진 부호화했으나 그 후 초기 압축효율의 개선을 위해 9비트부터 시작해 문자열 테이블 내에 할당된 부호어의 범위에 따라 12비트까지 2진 부호화하도록 개선된 LZW 알고리즘이 주로 사용되고 있음으로 여기에서는 이 후자의 알고리즘을 대상으로 하였다.

또 LZW알고리즘을 채용하고 있는 상용 압축파일에는 자주 쓰이던 PKARC, PKZIP 등이 있다. LZW알고리즘에 있어서 문자열 테이블 내에 생성되는 문자열의 개수는 사용시스템의 허용능력에 따라 확장할 수 있다. 상용 압축파일인 PKARC는 테이블 내 최대 문자열 엔트리를 4096으로 제한하고 있으며, PKZIP은 이를 8192로 제한하고 있다. 일반적으로 허용 가능한 문자열 엔트리를 늘릴수록 압축률은 향상된다.

2.1.4 허프만(Huffman) 부호화

허프만 압축 기법은 1954년 허프만에 의해 제안된 압축 방식으로 오늘날에도 널리 이용되고 있다. 이 기법은 정보원 데이타내의 각 문자에 대한 발생빈도를 조사해 자주 나타나는 문자에는 보다 짧은 부호어를, 그리고 잘 나타나지 않는 문자에는 더 긴 부호어를 할당함으로써, 전체 압축 후 부호어의 길이를 원래의 정보원 길이보다 더 축소시킬 수 있는 통계적 특성을 이용한 압축 기법이다.

빈도가 높은 정보는 적은 비트 수를 사용하여 표현하고, 빈도가 낮은 정보는 비트 수를 많이 사용하여 표현해서 전체 데이타의 표현에 필요한 비트의 양을 줄일 수 있다. 대표적인 가변 길이 코드 압축 기법이며, 영상 압축에 많이 사용된다. 그림 2는 허프만 압축 기법의 예를 나타낸다.

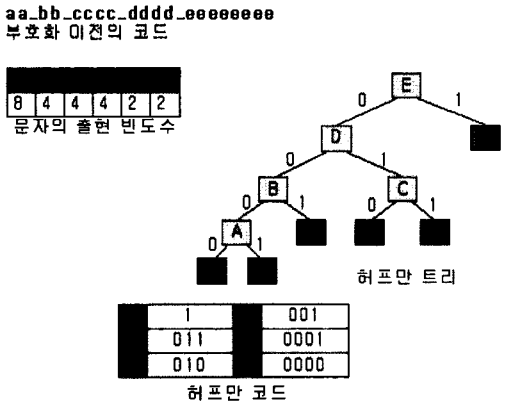


그림 2 허프만 압축 기법의 예

허프만 압축 기법을 적용한 압축 방식으로는 현재 팩시밀리 통신의 표준으로서 권고되고 있는 수정 허프만 부호가 있다. 팩시밀리 통신은 각 줄당 1728개에 달하는 데이타가 모두 흑('1'인 비트)과 백('0'인 비트)으로 이루어져 있으며, 또 이미지 정보의 경우 흑과 백의 분포가 상호 연관되어 있는 점에 착안, 각 줄당 흑과 백의 반복 개수 마다 앞의 허프만 기법을 적용해 압축 부호어를 구함으로써 데이타를 축소시킬 수 있게 된다. 현재 통계에 의하면 이 기법에 의해 원래 팩시밀리 통신의 이미지정보를 평균 약1/8이하로 축소해 전송할 수 있다고 한다.

허프만 압축 기법이 미리 조사된 정보원 데이타의 통계적 성질을 이용해 압축을 수행하기 때문에 실시간 입력 데이타에 적용하기 어려운 문제가 있다. 이런 문제점을 해결하기 위해 입력 데이타의 각 문자를 입력할 때마다 적응적으로 발생문자의 빈도수를 계산해 확률 값에 따라 허프만 부호를 할당하는 압축 기법이 개발되었으나, 압축 수행에 소요되는 시간 때문에 정보통신분야에서는 잘 이용하지 않는다.

2.1.5 압축 기법들의 특징 비교

RLE 압축 기법과 LZW 압축 기법은 고정 길이 압축 기법이고, 허프만 압축 기법은 가변 길이 압축 기법이라는 점에서 크게 구분된다. 일반적으로 압축 알고리즘의 속도 면에서 보면 RLE 압축 기법이 가장 빠르지만 압축률은 가장 낮다. RLE 압축기법은 파일 내에 동일한 문자의 연속된 나열이 있어야만 압축이 가능하기 때문이다. 이에 비해 LZW 압축 기법은 동일한 문자의 나열을 압축할 뿐 아니라, 동일한 패턴까지 압축하기 때문에 대부분의 경우에서 압축률이 가장 뛰어나다. 그러나 패턴 검색 방법이 최적화되지 않으면 성능이 떨어진다.

허프만 압축 기법은 텍스트 파일처럼 파일을 구성하는 문자의 종류가 적거나, 파일을 구성하는 문자의 빈도의 편차가 클수록 압축률이 좋아진다. 허프만 압축 기법은 많은 빈도수의 문자를 짧은 길이의 코드로, 적은 빈도수의 문자를 긴 길이의 코드로 대치하는 방법이어서 허프만 트리가 한쪽으로 쏠려 있을수록 압축률이 좋다. 그러나 빈도수가 고를 경우 허프만 트리는 대체로 균형을 이루게 되어 압축률이 현저히 떨어진다. 또한 허프만 압축 기법은 빈도수의 계산을 위해서 파일을 한번 미리 읽어야 하고, 다음에 실제 압축을 위해서 파일을 두 번 읽어야 하는 부담이 있어 실행 속도가 그리 빠르지 않다.

실제 상용 압축 프로그램들은 주로 허프만 압축 기법을 개량하거나, LZW 압축 기법을 개량해 사용하기도 하며, 혹은 이 둘과 RLE 압축 기법까지 총동원해서 최대의 압축률과 최소의 실행시간을 보이도록 최적화되어 있다.

2.2 일반적인 압축 프로그램

2.2.1 LZO

LZO(Lempel Ziv Oberhumer)는 이식이 쉬운 무손실 데이타 압축 프로그램이다. 또한 LZO는 압축률 보다 속도에 중점을 둔 실시간 작업에 적합한 데이타 압축 프로그램으로, 압축시 초당 약 5MB의 성능을 나타내며, 압축 해제 시는 을 추가적인 메모리 사용이 없기 때문에 초당 약 16MB의 성능을 나타낸다.

2.2.2 bzip2

bzip2는 Burrows-Wheeler 블럭 정렬 텍스트 압축 알고리즘과 허프만 코딩을 이용하여 파일을 압축한다. 이 압축은 전통적인 LZ77/LZ78 기반의 압축기에 의해 이루어지는 압축보다 일반적으로 상당히 좋다.

bzip2는 큰 파일을 블록 단위로 압축한다. 블록 크기는 압축된 비율과 압축과 해제에 필요한 메모리의 양 모두에 영향을 주며, 그 크기를 100,000 바이트에서 900,000 바이트(기본값)가 되도록 지정할 수 있다. 또한, 압축 해제시 압축에 사용된 블록 크기가 압축된 파일의 헤더에서 읽혀지고, bunzip2는 파일을 풀기에 충분한 메모리를 자신에게 할당한다. 압축과 해제 시에 요구량은 바이트로 다음과 같이 예측할 수 있다.

압축: 400k + (8 x 블록 크기)

해제: 100k + (4 x 블록 크기) 또는 100k + (2.5 x 블록 크기)

3. 제안하는 센서데이타 압축 기법

본 논문에서 제안하는 센서 데이타 압축 기법은 센싱 범위의 유효 값들을 비트로 표현해 전송 단위를 바이트 단위에서 비트 단위로 줄여 통신비용을 감소시킨다. 비트화를 위한 비트 할당 정보는 싱크 노드가 일정 주기마다 전체 데이타를 수집해 생성한 다음, 모든 센서들에게 배포한다. 이 장에서는 제안하는 압축 기법을 단계별로 상세하게 기술한다.

3.1 대상 센서네트워크 환경

이 절에서는 본 논문에서 대상으로 하는 센서네트워크 환경을 정의한다. 먼저 전체 센서네트워크 영역에 균일한 밀도로 센서가 배포된다고 가정한다. 각 센서는 센싱 기능, 프로세싱 기능, 무선 통신 기능을 갖추고 있으며 배터리를 사용해 동작한다. 센서들은 무선 네트워크를 구성하고 있으며, 센싱한 데이타는 정해진 라우팅 경로를 이용해 싱크 노드로 전송한다.

싱크 노드로 전송되는 센서 데이타는 센서 하드웨어에서 감지한 아날로그 값을 응용에서 사용할 수 있는 디지털 형태로 변환한 데이타이다. 일반적인 응용에서 사용할 목적으로 구축하는 대부분의 센서 네트워크에서 센서 데이타는 하드웨어 기능 또는 응용의 요구사항에 의해 제한되는 최대/최소 값의 범위 내에 있는 값이며, 최소 측정 단위 값의 배수로 표현된다. 예를 들어 온도를 센싱하는 센서 하드웨어가 측정할 수 있는 온도 값의 범위가 -50°C~100°C이고, 기기의 최소 측정 단위가 0.5°C라면 센서가 싱크 노드로 전송하는 데이타는 최대 100, 최소 -50의 범위를 가지며, 0.5의 배수로 표현되는 값들로만 표현된다.

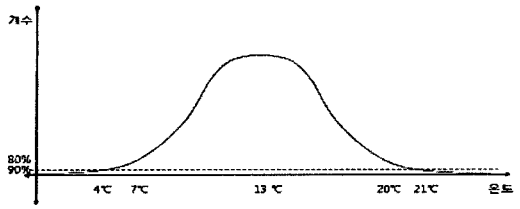
이런 센서 응용의 대표적인 예로 기상 관측 센서네트워크를 들 수 있다. 이때 측정하는 정보는 온도, 습도, 풍향, 풍속, 기압, 일사량, 방사능, 강우량 등이며, 대개 하나의 센서 장치에서 이런 여러 가지 정보를 측정할 수 있는 여러 개의 센싱 모듈을 갖추고 있다. 센서 장치는 센싱된 각각의 데이타를 모아 하나의 패킷으로 만들고 정해진 주기에 따라 싱크 노드로 전송하는 기능을 갖추고 있다.

센서데이타를 싱크로 수집할 때에는 센서네트워크 내에서 라우팅 경로를 설정해 멀티홉(multi-hop) 전송을 이용한다. 통신비용이 거리의 제곱에 비례해 늘어나기 때문에 에너지가 한정적인 센서에서는 통신 거리를 가급적 줄이는 것이 유리하다. 싱크가 정보를 센서에 배포하는 경우에는 방송(broadcast)을 통해 전체 센서네트워크의 모든 센서들에게 한 번에 데이타를 전송할 수 있다. 일반적으로 싱크는 전원을 별도 공급받기 때문에 통신비용을 고려하지 않는다. 그러나 싱크가 센서 중 하나일 경우에는 통신비용을 줄이기 위해 데이타 수집과 마찬가지로 멀티홉 전송을 이용하는 것이 적합하다.

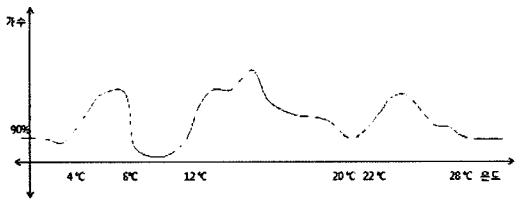
3.2 비트할당 정보 생성

일반적으로 센서네트워크에서 발생하는 데이타는 균일한 분포를 나타내는 경우 보다 특정 분포를 나타내는 경우가 많다. 이런 분포의 대표적인 형태로 가우시안 분포, 특정 값 또는 몇몇 값들을 중심으로 하는 군집형태 분포 등을 들 수 있다. 본 논문에서는 이런 특징을 활용해 센서 데이타를 비트화 해 크기를 줄이는 압축 기법을 제안한다. 비트화를 이용한 일반적인 압축 기법들은 각 데이타의 분포에 따라 가변길이 비트를 적용해 압축 효율을 높인다. 그러나 가변길이 비트화는 고정길이 비트화에 비해 처리 비용이 크기 때문에 센서와 같은 제한된 자원을 갖는 환경에 적합하지 않다. 따라서 본 논문에서는 주어진 데이타를 고정 길이 비트로 표현해 압축하는 기법을 제안한다.

제안하는 압축 기법에서 데이타를 비트화하기 위해 먼저 센서 네트워크에서 현재 수집한 센서 데이타들에 대한 분포 특성을 분석한다. 비트 할당 정보를 생성하기 위해 싱크 노드는 전체 센서 네트워크의 모든 센서들로부터 일정기간 동안 센싱한 데이타를 원본 그대로 수집한다. 싱크 노드는 수집한 데이타의 분포를 분석하고,



(a) 가우시안 분포



(b) 군집 분포

그림 3 센서데이타 발생 분포 예

분포 형태에 따라 비트 할당을 위한 정보를 생성한다. 그림 3은 센서네트워크에서 발생할 수 있는 데이타 분포 형태를 그래프로 나타낸 것이다.

3.2.1 가우시안 분포를 나타내는 경우

그림 3(a)와 같이 센서 데이타가 가우시안 분포를 나타내는 경우는 특정 목적을 위해 구축된 센서네트워크에서 가장 일반적으로 볼 수 있는 형태의 분포라 할 수 있다. 센서네트워크가 지역적으로 인접한 범위에만 배포되고, 장애물이나 지형적인 차이가 적을 경우 센서들은 측정 범위의 중심에 해당하는 값을 집중적으로 센싱하게 되고, 측정 범위의 최대, 최소값으로 갈수록 적은 수의 센서들이 값을 센싱하게 될 것이다. 이러한 환경에서 극단적인 경우, 영역 내의 모든 센서가 하나의 동일한 값을 센싱하는 경우가 발생할 수도 있다.

가우시안 분포를 나타내는 경우 비트 할당을 위한 정보를 생성하기 위해 비트화 할 값의 범위를 먼저 정해야 한다. 센싱 값을 비트화한 경우 비트할당 대상 범위를 벗어난 값들은 무시하도록 처리한다. 이 경우 비트할당 범위를 어떻게 지정하는가에 따라 결과 데이타의 정확성이 달라지기 때문에 '범위 지정을 어떻게 할 것인가'가 중요한 부분이 된다.

할당할 비트 수는 지정하는 센싱 데이타 범위 내의 모든 센서 데이타를 표현 가능한 수로 정한다. 비트할당 대상 범위 지정 및 표현 비트 수는 센서네트워크 응용에서 필요로 하는 데이타의 정확성이나 압축 요구사항에 따라 다르게 지정할 수 있다. 비트할당 정보는 G(avg, bit)와 같이 영역의 중앙값과 표현 비트 수의 쌍으로 생성한다.

그림 3(a)의 예에서 비트화 대상 센서 데이타 범위를

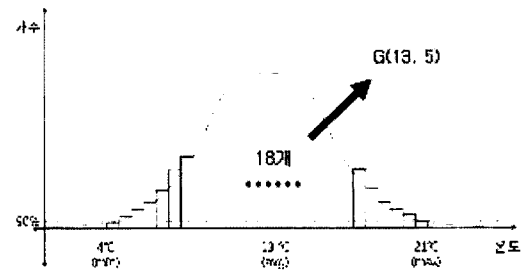


그림 4 그림 3(a)에 대한 비트할당 정보

전체 센서 데이타의 90% 이상을 포함하는 범위로 지정하면 최대값은 21°C, 최소값은 4°C가 되어 중앙값은 13°C가 된다. 이때 최소 센싱 단위가 1°C라면 비트화해 표현할 센서 데이타의 개수는 18개로 5비트면 모든 센서 데이타를 표현할 수 있다. 따라서 생성된 비트할당 정보는 G(18,5)가 된다. 또 비트할당 범위를 80%로 지정하면 중앙값은 13°C가 되며, 표현할 센서 데이타의 개수는 3개로 2비트면 모든 데이타를 표현할 수 있다. 그림 4는 그림 3(a)에 대해 비트화 범위를 90%로 지정한 경우 비트할당 결과를 나타낸다.

본 논문에서 제안하는 알고리즘에서는 비트할당 정보를 기준(중심)값과 할당 비트 수로 지정하고 있기 때문에, 그림 4에서와 같이 표현 비트를 5비트로 지정하면 대상 범위 내의 18개의 센서 데이타 개수보다 많은 32개의 값을 표현할 수 있다. 따라서 각 센서 노드는 싱크 노드가 압축 정보를 만들 때 참조한 센서 데이타 범위를 일부 벗어나는 데이타들에 대해서도 비트화를 할 수 있다.

3.2.2 군집 분포를 나타내는 경우

센서 네트워크가 지리적으로 넓게 분포하고 있으며, 지형 또는 장애물 등에 의해 센서 네트워크 환경이 몇 개의 소그룹으로 분할될 수 있다. 이 경우 각각의 소그룹은 3.2.1절의 가우시안 분포와 유사한 환경이 될 수 있다. 결과적으로 전체 센서 데이타의 분포는 몇 개의 군집을 이루는 형태로 나타날 수 있다.

전체 센서 데이타의 일정 양을 포함하는 범위를 지정하게 되면, 군집형태의 분포에서 비트화 대상 영역이 전 데이타 범위에서 연속적이지 않은 여러 개의 영역이 선정되는 형태가 된다. 또한 각 영역의 크기가 서로 다를 수 있기 때문에 이런 여러 영역에 대한 정보를 하나의 정보로 표현할 수 없다. 따라서 가우시안 분포를 나타내는 경우에 비해 좀 더 많은 정보를 센서들에게 배포해야만 각 센서들이 센싱한 데이타를 정확하게 비트화 할 수 있다.

제안하는 알고리즘에서는 센서에게 배포하는 비트할당 정보의 크기를 최소화하기 위해 비연속적인 각 군집

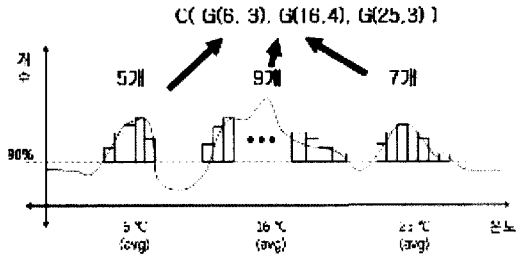


그림 5 그림 3(b)에 대한 비트할당 정보

영역을 가우시안 분포영역으로 모델링하고, 각 영역의 비트할당 정보를 모아 배포한다. 배포하는 비트할당 정보는 3.2.1절에서 제시한 가우시안 분포에 대한 비트할당 정보를 C(G1(avg, bit), G2(avg, bit), ..., Gn(avg, bit))와 같이 연속적으로 여러 개 포함된 형식으로 생성한다. 그림 5는 그림 3(b)를 비트화 범위를 90%로 지정해 비트할당 정보를 생성한 결과를 나타낸다.

3.3 비트할당 정보 배포 및 압축된 센서 데이타 수집

센서들을 처음 배포한 시점에서는 비트할당 정보가 없기 때문에 센싱한 원본 데이타를 그대로 싱크로 수집한다. 일정 기간 데이타 수집 후 싱크는 비트할당 정보를 생성하고 센서들에게 배포한다. 본 논문에서는 싱크가 센서가 아닌 별도 전력을 공급받는 컴퓨터 시스템으로 가정하므로, 비트할당 정보는 싱크가 모든 센서들에게 한 번의 방송을 통해 직접 배포할 수 있다. 이 경우 센서간 통신을 통해 비트할당 정보를 배포하는 것에 비해 센서의 에너지 소모를 줄일 수 있다.

싱크 노드가 배포하는 비트할당 정보를 수신한 이후 센서는 이 정보를 이용해 센서데이타를 비트로 변환해 전송한다. 본논문에서 대상으로 하는 센서네트워크 응용은 센싱하는 데이타의 최소 단위가 정해져 있다고 가정한다. 따라서 각 센서는 싱크 노드가 배포한 데이타 중심 값과 센싱한 데이타의 차를 이용해 주어진 비트 자릿수에 맞춰 2의 보수 형식의 비트 값으로 변환한다.

비트할당 범위를 벗어나는 센서 데이타들 중 비트할당 범위 내의 여유 비트로 표현 가능한 것들은 동일한 규칙으로 비트화해 압축된 데이타를 전송한다. 센서 데이타와 중심 값의 차가 커서 주어진 비트할당 범위에서 표현 불가능한 경우에는 센싱한 원본 데이타를 그대로 전송해 센서 데이타의 정확성이 떨어지는 것을 방지한다. 원본 데이타를 그대로 전송하는 것은 데이타 압축의 효과를 얻을 수 없지만, 비트화를 통한 데이타 압축에 의한 이득이 상대적으로 매우 크기 때문에 성능에 큰 영향을 미치지 않는다.

4. 성능 평가

표 1 성능평가 파라미터

항 목	값
센서 노드 수	3,600개
통신 반경	(1~30)k
센서 데이타 분포	가우시안, 군집 ,균등 분포
센서 데이타 최소단위	1 (℃)
센서 데이타 발생범위	-50~50 (℃)
비트화 대상 데이타 양	90~100 (%)
센싱 주기	30 (minute)
센싱 기간	1 (year)
비트화 정보 갱신주기	4~21 (day)

본 논문에서 제안한 압축 기법의 효과를 검증하기 위해 성능평가를 실시하였다. 먼저 센서 네트워크를 모델링하고, 성능에 영향을 미치는 몇 가지 요소에 대한 시뮬레이션을 실시하였다. 성능 비교는 압축을 적용하지 않고 센서 데이타 수집한 경우와 본 논문에서 제안한 압축 기법을 적용해 센서 데이타를 수집한 경우를 대상으로 하였다. 표 1은 성능 평가에 사용된 파라미터들이다.

시뮬레이션은 일정 지역에 기온을 센싱하는 센서들을 배포해 온도 데이타를 수집하는 기상 센서네트워크 환경을 대상으로 하였다. 센서 데이타 수집은 30분에 1회씩 24시간 동안 1년을 센싱한 기온을 정수형으로 싱크 노드로 전송하도록 하였다. 통신비용은 센서 데이타 수집과정에서 전송하는 데이타의 크기에 비례한다. 따라서 성능 비교를 위해 센서 데이타 수집과정에서 싱크 노드로 전송된 전체 데이타의 크기를 측정하였다. 이때 각 센서 노드는 싱크 노드까지 보내는 데이타 량과 통신거리를 고려한 전송 비용을 비교 하였다. 이 실험에서 최종 데이타를 처리하는 싱크 노드는 센서가 아닌 센서들과 통신이 가능하며, 사용자의 질의 처리 및 센서 데이타를 수집 처리하기 위한 서버 컴퓨터를 의미한다.

표 2는 압축을 하지 않은 경우와 비트할당 데이타 범위를 전체 데이타의 90-100%를 포함하는 경우로 변경해 가면서 센서데이타를 수집한 실험 결과이다. 먼저 압축 기법을 적용하지 않은 상황에서 실험한 결과 총 64,585,728,000 bit의 데이타가 싱크 노드로 전송되었다. 다음으로 센서 데이타의 비트할당 대상 범위를 비율로 지정해 압축 기법을 적용한 경우에 대해 실험을 실시하였다.

표 2 수집된 데이타 크기

비트할당 대상 데이타 범위 (%)	전송 데이타 크기 (bit)	압축비 (%)
비압축	64,585,728,000	100
90	13,625,737,728	21.1
95	10,778,019,840	16.7
100	7,946,035,200	12.3

실험 결과에서 보면 비트할당 데이타 범위를 전체 데이타의 90%를 포함하는 영역으로 할 경우, 총 13,625,737,728bit의 데이타가 수집되어 압축 기법을 적용하지 않았을 경우에 비해 전송데이타 량이 21.1%로 줄어들었음을 알 수 있다. 비트화 데이타 범위가 95%로 높아지면 압축률은 더 높아져 16.7%가 되며, 100%로 지정한 경우 12.3%까지 압축된 결과를 얻을 수 있었다.

센서데이타 수집에 소모되는 통신비용은 표 3의 수식 (1)을 이용해 계산할 수 있다. 통신비용은 데이타의 크기에 비례하기 때문에, 압축을 적용한 경우 최대 12.3%까지 전력 소모를 감소시킬 수 있어 센서네트워크의 수명을 약 8배 정도 증가시킬 수 있다. 표 4는 센서 데이타 수집시 통신비용 비교 결과를 나타낸다. E_{snd}, E_{rcv}는 무선으로 데이타를 송수신하기 위해 사용되는 비트당 소모 에너지로 이 실험에서는 MICAz에서 전송시 4317.89nJ/bit, 수신시 2028.66nJ/bit가 소모된다고 [5]에서 측정된 값을 사용하였다.

본 논문에서 제안하는 압축 기법은 압축을 위한 정보를 센서에게 배포하는 비용이 압축을 하지 않는 경우에 비해 추가적으로 더 소모된다. 센서에게 정보를 배포하는 비용이 크다면 압축으로 인한 효과가 상쇄되기 때문에 배포하는 정보의 크기와 비용을 계산해 보았다.

3장의 센서네트워크 환경에서 언급한 바와 같이 본 논문에서는 싱크의 에너지 제약을 고려하지 않기 때문에 방송을 통해 정보를 배포한다. 이 때 센서는 데이타를 송신할 때에 비해 훨씬 적은 비용을 소모한다. 센서

표 3 센서 데이타 전송/수신 통신비용 계산

파라미터	내 용
E	총 소모 에너지
S_d	데이타 크기
N_s	센서 수
N_d	센서 데이타 수집/배포 횟수
d	센서와 싱크 간 거리
E_{snd}	1bit 전송에 소모하는 에너지양
E_{rcv}	1bit 수신에 소모하는 에너지양
N_d = 주기 × 수집기간	
$E = S_d \times N_s \times N_d \times d^2 \times E_{snd}$ (1) $E = S_d \times N_s \times N_d \times E_{rcv}$ (2)	

표 4 센서데이타 수집시 통신비용

비트할당 대상 데이타 범위 (%)	통신비용 (J)	효율비 (%)
비압축	100.9	100
90	21.2	21.1
95	16.8	16.7
100	12.4	12.3

표 5 배포한 비트할당 정보 데이타 크기

비트할당 대상 데이타 범위 (%)	배포한 데이타 크기 (bit)
90	20,736,000
95	16,128,000
100	13,824,000

표 6 비트할당 정보 배포에 소모된 에너지

비트할당 대상 데이타 범위 (%)	배포 통신비용 (J)
90	1.03
95	0.80
100	0.69

의 수신 비용은 표 3의 수식 (2)로 계산할 수 있다. 표 5는 실험시 싱크가 센서에게 배포한 비트할당 정보의 총 크기이며, 표 6은 정보 배포시 센서가 소모한 비용을 계산한 것이다. 배포되는 정보의 크기와 통신비용이 데이타를 수집하는 경우에 비해 매우 적은 편이다. 압축을 통한 데이타 수집의 통신비용 절감 효과가 매우 크기 때문에 비트할당 정보 배포로 인한 추가적인 에너지 소모 부분은 무시할 수 있다.

5. 결 론

일반적으로 센서 네트워크에서 센싱되는 데이타들은 배포 환경 또는 응용 분야에 따라 특정 분포를 나타내는 형태로 나타날 수 있다. 본 논문에서는 센서 네트워크 환경에서 발생하는 다양한 데이타의 특성을 분석하고, 이결과를 바탕으로 데이타를 효과적으로 줄일 수 있는 압축 기법을 제안하였다.

제안하는 압축 기법에서는 센서데이타 분포 형태를 고려한 비트할당 정보 생성 기법을 제안하였다. 또한 비트화로 인해 수집된 센서 데이타의 정확성이 떨어지지 않도록 보완하는 알고리즘을 추가적으로 제안하였다. 그리고 성능 평가를 통해 전체 데이타를 포함하는 범위에 제안하는 압축 기법을 적용한 경우, 전체 수집 데이타의 크기를 약 12.3% 줄일 수 있음을 보였다. 압축을 통한 통신비용 절감 및 전력소모 감소로 센서 네트워크의 수명은 약 8배까지 증가시킬 수 있었다.

본 논문에서 제안하고 있는 센서 데이타 압축 기법과 함께 통신 및 신호처리 분야에서 사용하는 하드웨어적인 데이타 압축 기법들을 동시에 추가 적용한다면 센서 네트워크의 전력소모를 더욱 줄일 수 있을 것으로 기대된다. 본 논문에서는 센서의 처리를 단순화해 전력 소모를 줄이기 위해 고정길이 비트할당 기법만 적용하고 있다. 향후 압축 효과를 더 높이기 위해 가변길이 비트할당 기법을 적용한 경우에 대한 연구를 계속 수행할 것이다.

참 고 문 헌

[1] T. Arici, B. Gedik, Y. Altunbasak, and L. Liu, "PINCO: a Pipelined In-Network Compression Scheme for Data Collection in Wireless Sensor Networks," In Proceedings of 12th International Conference on Computer Communications and Networks, October 2003.

[2] E. Magli, M. Mancin, and L Merello, "Low-Complexity Video Compression for Wireless Sensor Networks," In Proceedings of 2003 International Conference on Multimedia and Expo, July 2003.

[3] S. S. Pradhan, J. Kusuma, and K Ramchandran, "Distributed Compression in a Dense Microsensor Network," IEEE Signal Processing Magazine, Volume: 19, Issue: 2, pp. 51-60, March 2002.

[4] J. Kusuma, L. Doherty, and K. Ramchandran, "Distributed Compression for Sensor Networks," In Proceedings of 2001 International Conference on Image Processing, October 2001.

[5] Naoto Kimura, and Shahram Latifi, "A Survey on Data Compression in Wireless Sensor Networks," In Proceedings of the International Conference on Information Technology: Coding and Computing, April 2005.

[6] 권영완, 김동국, 이좌형, 김윤, 정인범, "무선 센서 네트워크에서 압축을 이용한 네트워크 트래픽 감소 기법 구현", 한국정보과학회 가을 학술발표논문집, Vol.33, No.2(D), pp.699-704, 2006.

[7] D. Petrovic, R. C. Shah, K. Ramchandran, and J. Rabaey, "Data Funneling: Routing with Aggregation and Compression for Wireless Sensor Networks," In Proceedings of First IEEE International Workshop on Sensor Network Protocols and Applications, May 2003.

[8] wavelet.org, http://www.wavelet.org/

[9] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, K. Pister, "System architecture directions for network sendors," ASPLOS 2000, Cambridge, November 2000.

[10] MICA Wireless Measurement System, http://www.xbow.com/Products/Product_pdf_files/ Wireless_pdf/ MICA.pdf

[11] Junpei Kamimura, N. Wakamiya, M. Murata, "Distributed Clustering Method for Energy-Efficient Data Gathering in Sensor Networks," in Proceedings of the 1st IEEE Communications Society Conference (SECON 2004), Oct 2004.

이 석 재

정보과학회논문지 : 정보통신
제 35 권 제 2 호 참조

박 현 호

2004년 충주대학교 전자통신공학과 졸업 2005년 ㈜INE 정보보안 SE. 2007년~현재 충북대학교 정보통신공학과 석사과정 2008년 8월~현재 ㈜인프라웨어 컨버전스 사업2팀 SE. 관심분야는 데이타베이스 시스템, 무선 센서 네트워크, 네트워크 보안기술 등

여 명 호

정보과학회논문지 : 정보통신
제 35 권 제 2 호 참조

송 석 일

1998년 2월 충북대학교 정보통신공학과 공학사. 2000년 2월 충북대학교 정보통신공학과 공학석사. 2003년 2월 충북대학교 정보통신공학과 공학박사. 2003년 8월~현재 충주대학교 컴퓨터공학전공 조교수. 관심분야는 데이타베이스 시스템, 센서 네트워크, 이동객체

유 재 수

정보과학회논문지 : 정보통신
제 35 권 제 2 호 참조