

# 데이터 중심 저장 기법을 위한 효율적인 센서 데이터 압축 알고리즘

## An Efficient Sensor Data Compression Algorithm for Data-Centric Storages

노규종\*, 여명호\*, 성동욱\*, 복경수\*\*, 유재수\*  
충북대학교\*, (주)가인정보기술\*\*

Kyu-jong Rho\*, Myung-ho Yeo\*, Dong-ook Seong\*,  
Kyoung-soo Bok\*\*, Jae-soo Yoo\*  
Chungbuk National Univ.\*, GainIT Ltd.\*\*

### 요약

데이터 중심 저장 기법은 센서 네트워크에서 측정값을 효율적으로 저장하고 관리하기 위한 대표적인 연구 중 하나이다. 이 기법은 센서가 센싱한 데이터를 특정 센서 노드로 보내 저장한다. 하지만, 측정값을 네트워크 내부의 특정 노드로 전송하는 과정에서 많은 에너지를 소모하는 문제점을 가지고 있다. 본 논문에서는 데이터 중심 저장 기법에서 데이터의 전송 비용을 줄이기 위한 새로운 접근의 센서 데이터 압축 알고리즘을 제안한다. 제안하는 기법은 데이터를 전부 전송하는 것이 아니고 안전영역(Safe Region)을 설정한다. 그 후 이전 측정값을 기준으로 현재 측정값이 안전영역을 벗어나는 경우만 데이터를 전송함으로써 데이터의 전송을 줄이고 네트워크의 수명을 연장 시킨다. 성능평가 결과, 기존 기법에 비해 제안하는 기법의 에너지 소모가 60% 감소하였다.

### Abstract

Data-centric storage schemes(DCS) are one of representative researches that efficiently store and manage sensor readings in sensor nodes in the sensor networks. In DCS, a sensor sends the sensed data to a specific node in order to store them. However, it has a problem that sensor nodes consume a lot of energy for transmitting their readings to remote sensor node. In this paper, we propose a novel sensor data compression algorithm to reduce communication costs. The proposed algorithm does not transmit all the sensed data and establishes a safe region. It transmits the sensed data only when current measurement is out of the safe region. As a result, the proposed algorithm extends network life time and reduces data transmission. It is shown through performance evaluation that our proposed algorithm reduces energy consumption by about 60% over the conventional algorithm.

## I. 서론

최근 저전력, 저비용 통신기술의 발전에 힘입어 센서 네트워크에 사용되는 센서 노드의 성능은 향상되고 크기는 점차 작아지고 있다. 이러한 센서 노드를 이용하여 자연 환경 조사, 군사 응용 시스템, 환경 변화 감시 등의 여러 응용분야에서 무선 센서 네트워크의 활용에 대한 연구가 진행 되고 있다[1]. 일반적으로 센서 네트

워크를 활용한 서비스는 센서에서 수집한 데이터를 가공 처리 하여 제공하는 서비스 이다. 따라서 센서에서 수집된 데이터의 저장과 관리는 중요한 요소이다. 대표적인 데이터 저장 기법은 데이터 저장 위치와 기법에 따라 외부 저장 기법(External Storage : ES), 내부 저장 기법(Local Storage : LS), 데이터 중심 저장기법(Data Centric Storage : DCS)으로 구분 된다[2]. 외부 저장 기법은 수집된 데이터를 기지국에 저장하는 기법이다. 이 기법은 수집된 모든 센싱 데이터를 기지국으로 전송하기 때문에 높은 라우팅 비용이 소모된다. 라우팅 비용을 줄이기 위해 센싱 데이터를 수집한 노드

본 연구는 교육과학기술부와 한국산업기술재단의 지역혁신 인력양성사업으로 수행된 연구결과임

에 저장하는 내부 저장 기법이 제안되었다. 하지만 질의를 처리하기 위해 질의를 전체 네트워크에 배포해야 하는 단점이 있다. 이러한 단점을 해결하기 위해 DCS 기법이 제안되었다. 이 기법은 데이터 값의 범위마다 저장을 담당할 센서를 지정하고 해당 범위의 데이터가 발생하면 할당된 센서에 데이터를 저장한다. 이 기법은 질의를 네트워크 전체에 배포할 필요 없이 정확한 위치로 라우팅을 하기 때문에 질의의 불필요한 확산을 막을 수 있다. 하지만 발생하는 데이터를 모두 해당 저장 영역으로 전송해서 근사 데이터를 요구하는 응용분야에는 적합하지 않다.

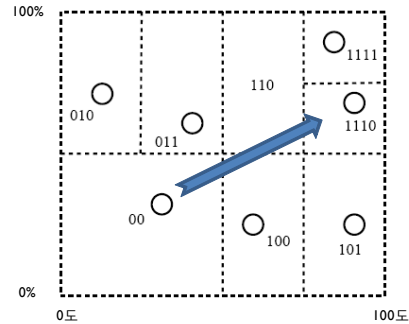
본 논문에서는 데이터 중심 저장 기법에서 데이터의 전송 비용을 줄이기 위한 새로운 접근의 데이터 압축 알고리즘을 제안한다. 제안하는 알고리즘은 데이터 전송 시 데이터를 전부 전송하는 것이 아니고 안전영역(Safe Region)을 설정하고 이전 측정값을 기준으로 현재 측정값이 안전영역을 벗어나는 경우만 데이터를 전송함으로써 데이터의 전송을 줄이고 네트워크의 수명을 연장 시킨다. 성능평가 결과 제안하는 기법이 DIM 기법에 비해 에너지 소모가 60% 감소 했다.

본 논문의 구성은 다음과 같다. II장에서 관련 연구에 관해 기술하고 III장에서 제안하는 데이터 저장 및 전송 기법과 질의 처리과정에 대해 설명하고, IV장에서 기존 연구와의 시뮬레이션을 통해 제안하는 기법의 우수성을 증명한다. V장에서는 결론 및 향후연구를 기술한다.

## II. 관련 연구

대표적인 DCS 기법으로는 GHT[3], DIM[4], KDDCS[5]가 있다. GHT기법은 공간 해싱 기법을 이용하여 데이터가 저장될 지리적인 위치를 랜덤하게 선정하고, 데이터의 저장, 탐색, 질의를 위해 해당 위치로 패킷을 GPSR[6]기법으로 최단 경로 라우팅을 하는 기법이다. DIM기법은 다차원 속성 값을 기반으로 센서 네트워크의 공간 영역을 매핑하면서 센서 노드에 데이터를 저장하는 기법이다. 공간 분할 과정은 x좌표와 y좌표를 번갈아 나누고 하나의 센서가 남을 때 까지 영역을 분할한다. 분할 후 해당 영역에 포함된 센서는 해당 영역의 데이터 값을 저장한다. 그림 1은 DIM에서의 데이터 전송기법을 보여준다. 노드 ID 00인 센서가 75

도~100도, 50%~75%의 값을 센싱한다면 센싱한 데이터를 노드 ID 1110으로 전송한다.



▶▶ 그림 1. DIM에서의 데이터 전송 기법

KDDCS 기법은 센서가 배포된 지리적 영역에 각 차원별로 데이터의 범위를 할당하고, 센서의 배포위치를 기반으로 K-D 트리 공간 분할을 통해 하나의 센서 노드가 하나의 분할 영역에 남게 될 때 까지 분할 한다. 이 기법은 데이터 라우팅시 상위 영역 정보를 가지고 있는 센서 노드를 방문하여 데이터의 라우팅 경로를 지정한다. 하지만 DCS 기법은 데이터가 발생할 때 마다 데이터를 저장 영역으로 모두 전송하기 때문에 데이터 전송에 많은 에너지를 소모한다. 따라서 데이터 전송비용을 줄이고 네트워크의 수명을 연장하기 위한 에너지 효율적인 기법에 대한 연구가 필요하다.

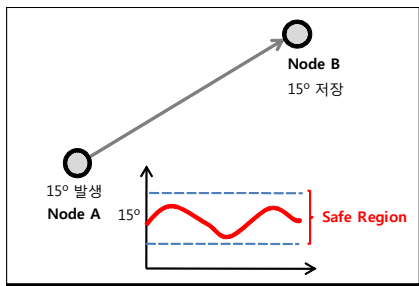
## III. 데이터 중심 저장 기법을 위한 제안하는 센서데이터 압축 기법

본 장에서는 제안하는 데이터 중심 저장 구조를 위한 압축 기법을 기술한다. 먼저, 센서데이터의 근사성을 이용하기 위한 안전영역(Safe Region) 개념을 설명하고, 데이터의 전송 및 저장 과정을 상세히 기술한다.

### 1. 특징

제안하는 기법은 그림 2와 같이 안전영역을 설정하고 이전 측정값을 기준으로 현재 측정값이 안전영역을 벗어나는 경우만 데이터를 전송한다. 일반적인 DCS 기법은 데이터가 발생하면 지정된 센서로 데이터를 전송하

여 저장한다. 하지만 데이터가 발생할 때마다 해당 데이터가 저장 될 영역으로 계속 전송을 하기 때문에 불필요한 전송 비용이 소모된다. 센서 네트워크는 근사 데이터를 이용하는 대표적인 응용 분야이다. 센서 자체의 오차 범위로 인해 측정값은 어느 정도의 오차를 가지며, 무선으로 데이터가 전송이 되기 때문에 환경적인 제약에 따라 오류가 발생하기 쉽기 때문이다. 또한, 센서 네트워크의 응용 분야에 따라 정확한 물리량을 측정하기보다 근사 데이터의 수집을 목적으로 하기도 한다. 예를 들면, 0.5도 단위로 센싱하는 센서는 20.2도 같은 데이터는 20도 또는 20.5도로 센싱 하기 때문에 정확한 값이 아니다. 그렇기 때문에 센서 데이터에 대한 근사적인 접근을 활용하여 오차 허용범위( $\epsilon$ )를 설정할 수 있고, 측정값이 해당 범위를 벗어나는 경우만 측정값을 전송함으로써 데이터 전송 비용을 줄일 수 있다.



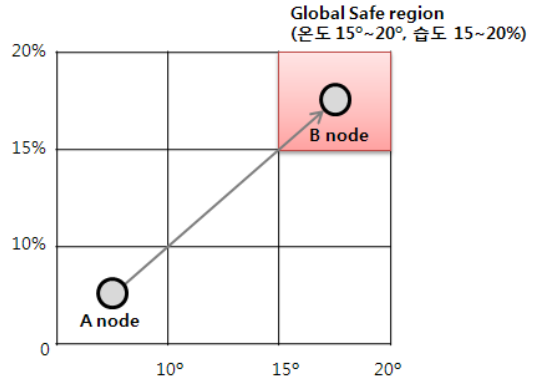
▶▶ 그림 2. 노드 A에서의 안전영역(Safe Region) 설정

2. 데이터 전송 및 저장

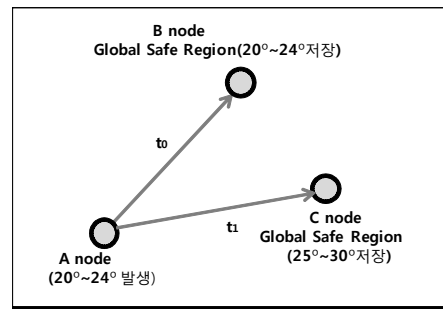
제안하는 기법은 측정값이 오차 허용범위 이내에 존재하는 경우, 데이터를 전송하지 않는다. 데이터를 전송하지 않는 오차 허용 범위를 안전 영역(Safe Region)으로 정의한다. 안전 영역은 전역 안전 영역(Global Safe Region)과 지역 안전 영역(Local Safe Region) 으로 구분된다.

첫 번째, 전역 안전 영역은 한 개의 센서가 담당하는 영역을 의미한다. 그림 3과 같이 온도 15도~20도, 습도 15%~20%의 값이 센싱되면 해당 데이터는 모두 B 노드로 전송된다. 처음으로 데이터가 발생하면 해당 데이터를 저장하는 B 노드에 보고하고 전역 안전 영역을 설정한 후 데이터를 전송한다. 데이터의 전송은 해당 전역 안전 영역을 벗어나기 전까지 수행되며, 센싱 값이 전

역 안전 영역을 벗어나면 B 노드에 해당 영역을 벗어났다고 보고한다. B 노드는 A 노드에서 전송한 데이터를 시간 값과 함께 데이터 저장 테이블에 저장한다.



▶▶ 그림 3. 전역 안전 영역(Global Safe Region)의 설정



(a) 전역 안전 영역의 변경

데이터 저장 (Data Storage)		
센서 아이디	데이터	시간
A	20	$t_0$

B node

데이터 저장 (Data Storage)		
센서 아이디	데이터	시간
A	25	$t_1$

C node

(b) 데이터 저장 테이블의 구조

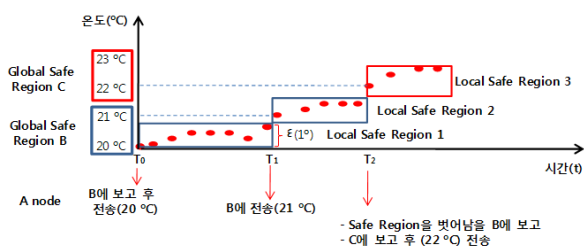
▶▶ 그림 4. 전역 안전 영역의 변경과 데이터 저장 테이블의 구조

그림 4(a)는 전역 안전 영역의 변경을 나타낸다. 처음에 A 노드는  $T_0$ 의 시점에 20도~24도의 값을 센싱한다. 이 값을 전송하기 위해 B 노드를 전역 안전 영역으로 설정한 후 데이터를 전송한다. 그 후  $T_1$ 의 시점에 25도~30도의 값이 발생하면 해당 전역 안전 영역을 벗어나는 값이기 때문에 A 노드는 B 노드에 범위를 벗어났다는 메시지를 보내고 C 노드를 전역 안전 영역으로 설정한다. 그 후 센싱된 데이터는 모두 C 노드로 전송

한다. 그림 4(b)는 데이터 저장 테이블의 구조를 나타낸다. 데이터 저장 테이블에는 고유한 센서의 아이디와 데이터 그리고 시간이 저장된다. B 노드의 저장테이블에는 T0 시점에 A 노드가 센싱한 20도의 값이 저장된다. T1 시점에서 A 노드는 25도를 센싱하고 이 값은 C 노드의 저장 테이블에 저장된다.

두 번째, 지역 안전 영역(Local Safe Region)은 오차 허용 범위내에서 데이터가 전송되지 않는 영역이다. 일반적으로 센서 네트워크상의 데이터는 시간의 흐름에 따라 크게 변화하지 않고 유사한 값이 지속적으로 발생한다는 특성이 있다. 제안하는 기법은 일단 데이터가 센싱되면 처음 센싱된 값을 가지고 전역 안전 영역을 설정한다. 다음으로 오차 허용 범위( $\epsilon$ ) 설정하고 오차 허용 범위( $\epsilon$ ) 내의 영역을 지역 안전 영역으로 선정한다. 데이터의 전송 여부는 지역 안전 영역 내에서 설정한 오차 허용 범위( $\epsilon$ )에 따라 결정된다.

그림 5는 지역 안전 영역의 설정과 값을 전송방법을 나타낸다. T0 시점에 A 노드에서 20도가 센싱이 되면 B 노드를 전역 안전 영역으로 설정하고 값을 전송한다. 전송한 후 다음 전송은 지역 안전 영역의 값인 오차 허용 범위( $\epsilon$ )를 1도로 정하고 값의 차이가 1도 이상 차이가 나지 않으면 전송하지 않는다. T1 시점에 21도가 센싱이 되면 값의 차이가 1도 이상 나기 때문에 값을 전송한다. T2 시점에 22도가 센싱이 되면 해당 전역 안전 영역을 벗어나기 때문에 B 노드에 해당 전역 안전 영역을 벗어났다고 보고 후 C 노드를 전역 안전 영역으로 선정하고 값을 전송한다. 오차 허용 범위( $\epsilon$ )의 설정은 응용에 따라 다르게 할 수 있다. 만약 정확한 값을 요구하는 응용에서는 오차 허용 범위( $\epsilon$ )를 작게 설정해서 더욱 정확한 값을 전송해야 하고 그렇지 않다면 오차 허용 범위( $\epsilon$ )를 크게 설정해서 근사 데이터 값을 전송한다. 제안하는 기법은 이처럼 전역 안전 영역과 지역 안전 영역을 이용하여 데이터 압축의 효과를 내며 데이터의 전송 빈도를 줄여 에너지 소모를 줄인다.



▶▶ 그림 5. 지역 안전 영역의 설정과 값의 전송

## IV. 성능평가

### 1. 실험 환경

제안하는 기법의 우수성을 보이기 위해서 기존에 제안된 히스토그램 기반 기법에서 수행한 실험과 동일한 방법으로 성능 평가를 수행하였다. 실험에 사용한 데이터는 인텔 연구실[7]에서 1개월 동안 54개의 센서로부터 수집한 실제 환경 데이터이다. 중앙에 기지국이 위치한 100m X 100m의 영역 공간을 가정하였으며, 임의로 54개의 센서 노드를 위치시키고 센서 노드는 다중-홉으로 통신한다고 가정하였다. 임의의 네트워크 토폴로지(network topology)에서 연결성을 보장하기 위해서 통신 반경은 18로 설정하였다. 표 1은 임의로 구성된 1,000개의 센서 네트워크로부터 측정된 평균 중간 노드의 수, 평균 전송 거리이다. 센서 노드의 통신 반경은 18m로 설정한다. 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/m<sup>2</sup>으로 설정한다. 메시지 수신에 소모되는 에너지 모델은 {메시지 크기} x {수신 비용}이며, 수신 비용은 50nJ/b으로 설정한다.

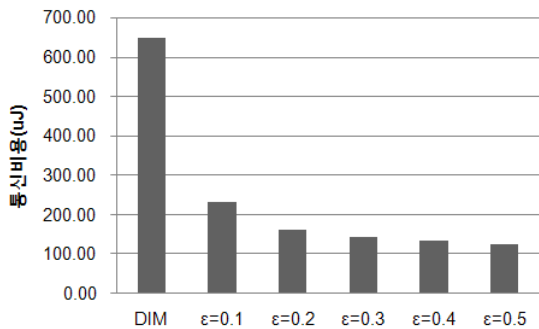
표 1. 성능 파라미터

센서노드의수(N)	54
중간노드의평균수(Nni)	25.7
통신반경	18m
평균통신홉수	4.26
측정값의 범위	10 ~ 40도
측정값의 수	94156

### 2. 실험 결과

시뮬레이션은 54개의 센서 노드를 균등하게 배포한 환경에서 수행하였다. 제안하는 기법의 환경변수인 전역 안전 영역은 10~40도 범위를 54개로 균등하게 분할 설정하였고, Local Safe Region( $\epsilon$ )을 0.1~0.5의 범위 내에서 0.1단위로 변화시켜가며 테스트하였다. 그림 6은 통신비용 측면에서 제안하는 기법과 DIM을 비교한 결과이다. 비교 결과, DIM에 비해 제안하는 기법의 에너지 소모가 약 60% 감소한 것을 보여준다. DIM의 경우 네트워크에서 발생하는 모든 데이터를 아무런 가공

처리 없이 전송하여 높은 통신비용을 소모한다. 하지만 제안하는 기법의 경우 발생하는 데이터의 유사한 정도를 판단하여 설정된 오차 허용범위( $\epsilon$ ) 이내의 데이터가 발생하였을 때, 이러한 데이터의 전송을 불필요한 데이터 전송로 간주하고, 이를 제거함으로써 많은 통신비용적 측면의 이득을 가진다. 실험결과 제안하는 기법에서 오차 허용범위가 증가함에 따라 발생하는 통신비용의 이득이 증가하는 것을 확인할 수 있으며, 0.3을 기점으로 수렴하는 것을 확인할 수 있다.



▶▶ 그림 6.  $\epsilon$ 에 따른 통신 비용 비교

## V. 결론 및 향후 연구

본 논문에서는 데이터 중심 저장 기법에서 데이터의 전송 비용을 줄이기 위한 새로운 접근의 데이터 압축 알고리즘을 제안하였다. 제안하는 기법은 발생한 데이터를 전부 전송하는 것이 아니라 안전영역을 설정하고 오차 허용 범위를 사용하여 오차 허용 범위를 넘었을 때만 데이터를 전송함으로써 데이터의 압축의 효과와 함께 데이터 전송량을 줄이는 기법이다. 성능평가 결과 기존 기법에 비해 데이터 전송 비용이 60% 감소하였다. 향후 연구는 시간의 근사 속성을 이용하여 질의 처리 비용을 줄이는 것이다.

## ■ 참고 문헌 ■

[1] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient Algorithms for Mining Outliers from

Large Data Sets", In Proceedings of ACM SIGMOD'00, New York, NY, USA, 2000.

- [2] M. Aly, P. K. Chrysanthos, and K. Pruthis, "Decomposing data-centric storage query hot-spot in sensor networks", In Proceedings of The 3<sup>rd</sup> Annual International Conference on Mobile and Ubiquitous Systems, 2006.
- [3] S. Ratnasamy, B. Karp, S. Shenker, D. Estrin, R. Govindan, L. Yin, F. Yu, "Data-centric storage in sensornets with GHT, a geographic hash table", MONET vol. 8, no. 4, 2003.
- [4] X. Li, Y. J. Kim, R. Govindan and W. Hong. "Multi-dimensional Range Queries in sensor Networks.", In Proceedings of ACM Conference on Embedded Networked Sensor Systems, 2003.
- [5] M. Aly, K. Pruthis, and P. K. Chrysanthos "KDDCS: A LoadBalanced In Network Data Centric Storage scheme for Sensor Networks", In Proceedings of ACM CIKM'06, 2006
- [6] B. Karp, H. T. Kung. "GPSR: Greedy Perimeter Stateless routing for wireless sensor networks", In Proceedings of ACM Mobicom'00, 2000
- [7] Intel Lab Data, <http://berkeley.intel-research.net/labdata/>, Apr. 2004.