

센서 네트워크 환경에서 비컨 교환을 이용한 아웃라이어 탐지 기법

여명호*, 성동욱*, 조용준*, 유재수*

*충북대학교 정보통신공학과

e-mail : {mhyeo, sergei, yjcho}@netdb.cbnu.ac.kr, yjs@cbnu.ac.kr

An Outlier Detection Algorithm using Beacon Messaging in Sensor Networks

Myung-Ho Yeo*, Dong-Ook Seong*, Yong-Jun Cho*, Jae-Soo Yoo*

*Dept. of Information and Communication Engineering, Chungbuk National University

요 약

아웃라이어 탐지는 일반적이지 않은 특징을 가진 데이터를 찾는 기법으로 데이터 마이닝 분야의 중요한 연구 중 하나이다. 본 논문에서는 센서 네트워크를 이용한 응용분야에서 아웃라이어를 효과적으로 탐지하기 위한 인-네트워크 처리 기법을 제안한다. 비컨 교환을 통해 데이터 교환 비용을 줄이고, 슬라이딩 윈도우를 이용하여 센서 노드간 불필요한 데이터 수신(overhearing)을 줄인다. 실험 결과, TDMA 를 이용하여 모든 데이터를 기지국으로 전송하는 Naïve 기법에 비해 평균 86.92% ~ 133.67% 네트워크 수명이 연장되었다.

1. 서론

아웃라이어 데이터 집합에서 일반적이지 않은 특징 데이터를 추출하는 방법으로 기존 데이터 마이닝 분야의 중요한 연구 중 하나이다[1, 2, 3, 4]. 환경, 건설, 군사 분야의 모니터링 용도로 각광받고 있는 센서 네트워크 분야에서도 아웃라이어 탐지는 중요한 의미를 지닌다. 예를 들면, 화재, 지진, 해일과 같은 자연 재해의 진조현상 및 다리, 빌딩과 같은 건축물의 안전도를 모니터링 용도로 사용될 수 있다. 또한, 군사 응용 분야에서는 화학 물질이나 방사능의 누출을 효과적으로 탐지할 수 있는 방법이다. 하지만, 기존에 제안된 아웃라이어 탐지 기법은 중앙처리를 요구하며, 센서 네트워크에 적용할 경우, 모든 데이터를 기지국으로 수집해야 한다. 하지만, 데이터의 수집 과정에서 많은 에너지를 소모하기 때문에 에너지를 효율적으로 사용하기 위한 방법의 연구가 필요하다.

효과적인 데이터 수집을 위한 최근 연구 동향은 응용 분야에 해당하는 질의를 각 센서 노드로 전달하고, 질의에 해당하는 결과만을 기지국으로 수집하는 인-네트워크 처리(In-network processing) 기법들이 제안되고 있다. 아웃라이어 탐지를 위한 대표적인 인-네트워크 처리 기법으로는 [5]와 [6]이 있다. [5]는 거리 기반의 로컬 아웃라이어 탐지 기법으로, 각각의 센서 노드는 자신이 수집한 데이터 내에서 데이터 값의 거리

를 비교함으로써 아웃라이어를 탐지한다. 특정 시간대의 슬라이딩 윈도우를 설정하고, 윈도우 내에 임의로 추출한 샘플 데이터의 비교를 통해 아웃라이어를 탐지한다. 하지만, 센서 노드가 과거 데이터를 비교하기 위해서는 별도의 저장 공간이 필요하고, 비정상적인 센서 노드가 포함된 경우, 아웃라이어를 탐지 못하는 경우가 발생한다. [6]은 히스토그램을 이용한 글로벌 아웃라이어 탐지 기법을 제안한다. 글로벌 아웃라이어 탐지 기법은 로컬 아웃라이어 탐지 기법과 달리 특정 시점에 전체 센서 네트워크의 수집된 데이터를 기반으로 아웃라이어를 탐지하는 방식이다. [6]은 데이터 수집에 소요되는 통신 비용을 줄이기 위해 원본 데이터를 수집하는 대신 데이터 분포 특성을 나타내는 히스토그램을 수집하는 방법을 제안한다. 히스토그램을 수집한 기지국은 히스토그램을 분석하여 아웃라이어, 잠재적인 아웃라이어, 비-아웃라이어 데이터를 판별하게 된다. 하지만, 정확한 아웃라이어 탐지를 위해서 잠재적인 아웃라이어 판별된 센서 데이터는 다시 센서 네트워크로 전송해야하며 이 과정에서 많은 에너지를 소모하는 문제점을 가지고 있다.

본 논문에서는 데이터 교환 비용을 줄이기 위해서 비컨(beacon) 메시지를 교환하는 기법과 불필요한 데이터 수신을 방지하기 위한 슬라이딩 윈도우 기법을 제안한다. 가장 직관적인 인 네트워크 처리 방법으로 자신의 데이터를 주변 센서 노드와 교환하는 방법이다. 모든 센서 노드는 자신의 원본 데이터 대신 비컨 메시지를 주변 센서 노드와 교환하고, 이를 수신한 센서 노드는 해당 데이터가 자신의 수집 데이터와 특정 범위 이내 있는지 여부를 판단하고 그 수를 카운

이 논문은 2008년도 교육과학기술부로부터 지원(지역거점연구단육성사업/충북 BIT 연구중심대학육성사업단)과 한국과학재단의 지원(No. R01-2006-000-1080900)을 받아 수행된 연구임.

트한다. 카운트는 자신의 수집 데이터와 근사한 데이터의 수를 의미하기 때문에 일정 임계치 이하의 수를 카운트한 경우 센서 노드 스스로 아웃라이어 여부를 판단하게 된다. 시물레이션을 통해 모든 데이터를 기지국으로 전송하여 처리하는 TDMA 스케줄에 따라 데이터를 기지국으로 수집하는 Naïve 기법과 성능을 비교하였으며, 그 결과 네트워크의 수명이 00% 향상되었다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 거리 기반의 아웃라이어를 정의하고, 제 3 장에서는 본 논문에서 제안하는 비컨 교환을 이용한 아웃라이어 탐지 기법을 제안한다. 제 4 장에서는 시물레이션을 통해 제안된 기법의 성능을 비교 평가하고, 제 5 장에서는 결론과 향후 연구방향을 기술한다.

2. 거리 기반의 아웃라이어 정의

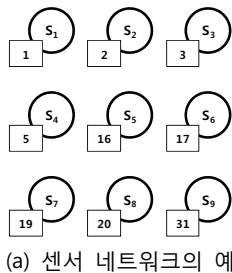
아웃라이어는 다른 데이터 포인트와 비교하여 매우 다른 데이터 포인트를 의미한다. 본 논문에서는 [6]과 같이 거리 기반의 아웃라이어를 탐지하기 위한 연구를 고려한다. 여기서 “거리”란 데이터 포인트간의 거리를 의미한다. 예를 들면, 데이터 포인트 A 와 B 가 각각 속성값 10 과 14 에 해당한다고 할 때, 거리는 4 가 된다. 거리 기반의 아웃라이어를 탐지하기 위해서 최근접 이웃(Nearest Neighbor)의 개념을 이용한다. 어떤 데이터 포인트 p 로부터 거리에 따라 데이터 포인트를 $p_1, p_2, \dots, p_k, \dots$ 과 같이 정렬한다고 가정하

자. 이때, 각 데이터 포인트는 $|p_1-p| \leq |p_2-p| \leq \dots \leq |p_k-p|$

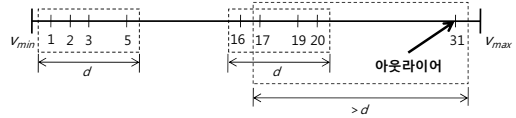
$\leq \dots$ 의 식이 성립한다. 데이터 포인트 p 와 p 의 k 번

째 최근접 이웃(k -th Nearest Neighbor)까지의 거리를 $D^k(p) = |p_k - p|$ 로 정의한다. 이때, 두 가지 종류의 거리 기반 아웃라이어는 다음과 같이 정의할 수 있다.

정의(1). 만약 $D^k(p)$ 가 특정 거리 d 보다 크거나 같으면 ($D^k(p) \geq d$), 데이터 포인트 p 를 아웃라이어 $O(d,k)$ 라고 정의한다.



(a) 센서 네트워크의 예



(b) 센서 데이터의 분포

(그림 1) 거리 기반의 아웃라이어 $O(d,k)$ 정의 ($d=5, k=3$)

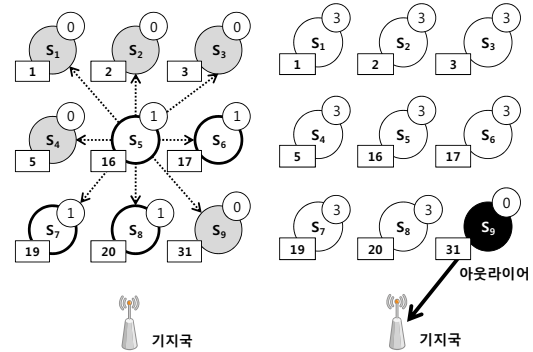
(그림 1(a))와 같은 단일 홉으로 연결된 센서 네트워크가 있다고 가정할 때, 각 센서가 측정한 센서 데이터는 (그림 1(b))와 같은 분포를 보인다. 아웃라이어 $O(5,3)$ 를 탐색한다고 할 때, 정의(1)에 의하여 $k(=3)$ 번째 최근접 이웃까지의 거리가 $d(=5)$ 보다 큰 31의 값을 측정할 센서 노드 $\{S_9\}$ 는 아웃라이어에 속한다.

3. 제안하는 아웃라이어 탐지 기법

3.1. 특징

가장 직관적인 인 네트워크 아웃라이어 탐지 기법으로 자신의 데이터를 주변 노드와 교환함으로써 센서 노드 스스로 아웃라이어를 판별하는 방법을 고려할 수 있다. 먼저, 모든 센서 노드가 자신의 데이터를 다른 센서 노드에게 알리고, 데이터를 수신한 센서 노드는 수신한 데이터의 범위가 자신의 데이터와 d 범위 이내인지 판별하고 카운트를 증가시킨다. 결국, 모든 센서 노드가 다른 센서 노드의 센서 데이터를 인지함으로써 스스로 아웃라이어 여부를 판별한다.

(그림 2)와 같은 모든 센서 노드가 단일 홉 통신이 가능한 센서 네트워크가 있다고 할 때, 아웃라이어 $O(5,3)$ 를 탐지한다고 가정하자.



(a) 데이터 교환 (b) 아웃라이어 판별 및 전송

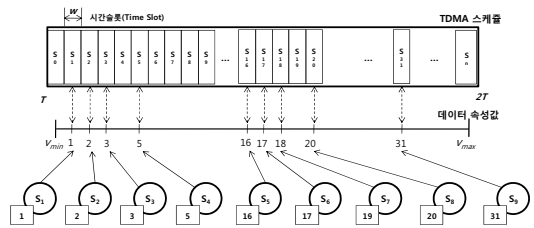
(그림 2) 데이터 교환을 통한 아웃라이어 탐지 기법

먼저, 각 센서 노드 $S_1 \sim S_9$ 는 (그림 2(a))와 같이 자신의 데이터를 전송하게 된다. 센서 노드 S_5 의 경우, 센싱한 값 16 을 주위로 전송하고, 이를 수신한 모든 센서 노드는 16 과 자신의 센싱값이 $d(=5)$ 범위 이내에 있는지 비교한다. 그리고, $d(=5)$ 범위 이내에 있는 $\{S_6, S_7, S_8\}$ 은 자신의 카운터를 1 증가 시킨다.

모든 센싱값이 교환된 이후, (그림 2(b))와 같이 자신의 카운터 값이 $k(=3)$ 이하인 센서 노드 S_0 는 아웃라이어로 결정되고, 자신의 값을 기지국으로 전송한다. 아웃라이어로 판별된 센서 데이터만 기지국으로 전송하기 때문에 기지국이 상대적으로 멀리 위치한 경우 효과적이다. 하지만, 기지국이 센서 네트워크와 인접한 경우, 불필요한 데이터 수신으로 인해 TDMA 에 따라 기지국으로 데이터를 수집하는 경우 보다 비효율적이다. S_5 의 경우, $\{S_1, S_2, S_3, S_4\}$ 와 데이터와 거리 $d(=5)$ 이상의 데이터를 수집했기 때문에 $\{S_1, S_2, S_3, S_4\}$ 의 데이터를 수신시 카운트를 하지 않는다. 이렇게 카운트에 활용되지 않는 데이터의 수신은 S_5 의 입장에서 불필요한 수신이다. 센서 노드의 수가 증가함에 따라 데이터 수신에 따른 에너지 소모가 증가하기 때문에 에너지 소모를 최소화하기 위해서는 전송되는 데이터의 크기를 줄이고, 불필요한 데이터의 수신을 방지해야 한다. 본 논문에서는 이를 위한 데이터 전송 크기를 줄이기 위한 데이터 인지 비컨 교환 기법과 불필요한 데이터 수신을 방지하기 위한 슬라이딩 윈도우를 이용한 상태 변환 기법을 함께 제안한다.

3.2. 데이터 인지 비컨 교환 기법

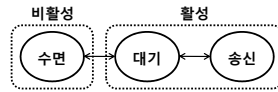
본 논문에서는 먼저 (그림 3)과 같이 센싱한 데이터값에 따라 동적으로 메시지의 전송 시점을 결정하는 데이터 인지 비컨 브로드캐스트 기법을 제안한다. 센싱하는 데이터 속성값의 범위가 ($v_{min} \sim v_{max}$) 일 때, 데이터 수집에 소요되는 시간 T 의 시간을 n 개의 시간 슬롯으로 분할하고, 해당 속성값의 범위를 일치시킨다. 따라서 작은 값을 가진 센서 노드일수록 빠른 시간 슬롯을 할당한다. 각 센서 노드가 센싱 데이터를 수집하는 주기와 데이터 속성값의 범위를 알고 있다고 가정할 때, 이러한 일련의 처리 과정은 중앙의 제어없이 각 센서 노드에 의해 결정된다. 모든 센서 노드가 동기화되어 있다고 가정할 때, 각 센서 노드는 데이터의 크기에 따라 데이터 전송 시점을 달리하며, 센서 데이터가 아니라 단순히 작은 크기의 데이터를 송신하고 수신함으로써 원본 센서 데이터를 계산할 수 있다. 예를 들어 속성값이 (1, 100)의 범위를 가지고 있고, 100 개의 시간 슬롯으로 TDMA 가 구성된다고 할 때, 센서 데이터 1 을 센싱한 센서 노드는 시간 슬롯 S_1 에 할당되어 데이터 송신을 시도하고, 이때 데이터를 수신한 센서 노드는 수신한 데이터의 내용과 관계없이 주변에 센서 데이터 1 을 가진 센서 노드가 존재함을 인지할 수 있다. 즉, 데이터를 전송하는 시점인 시간슬롯이 센서 데이터의 값을 나타냄으로써 전송 패킷에 센서 데이터의 값을 포함하지 않더라도 데이터의 수신 시점을 이용하여 센서 데이터의 값을 파악할 수 있다. 따라서, 센서 데이터 자체를 전송하지 않고, 비컨 메시지와 같은 작은 크기의 메시지를 이용하여 센서 데이터를 효과적으로 전송하게 된다.



(그림 3) 데이터 인지 동적 TDMA 스케줄링 기법

3.3. 슬라이딩 윈도우를 이용한 센서 노드 상태 변환

센서 노드의 상태는 (그림 4)와 <표 1>과 같이 크게 비활성과 활성 상태로 나누어진다. 비활성 상태 즉, 수면 상태에는 에너지 소모를 최소화 하기 위해서 데이터 송수신을 하지 않는다. 활성 상태는 데이터 송/수신 여부에 따라 대기상태와 송신상태로 나누어진다. 에너지 사용을 최소화 하기 위해서 가능한 수면상태의 시간을 늘리고, 활성 상태의 시간을 줄여야 한다. 센서 네트워크에서 데이터를 수집할 때, 에너지 사용을 최소화하기 위해서 TDMA 방법을 널리 사용한다. 수집에 소요되는 시간을 T 라고 할 때, T 시간을 시간 슬롯 단위로 분할하고, 각 시간 슬롯에 센서 노드를 배정한다. 센서 노드는 수면상태를 유지하며 자신의 시간 슬롯 시점에서 활성화하여 데이터 통신을 수행한다.

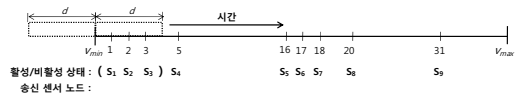


(그림 4) 센서 노드의 상태

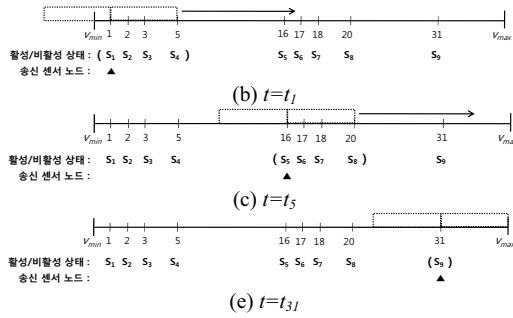
<표 1> 센서 노드의 상태

상태	비활성	활성	
	수면상태	대기상태	송신상태
데이터수신가능	X	O	X
데이터송신가능	X	X	O

거리 d 이내에 포함된 센서 노드간의 독립적인 통신을 보장하기 위해서 슬라이딩 윈도우를 이용한다. 슬라이딩 윈도우의 크기는 $2d-w$ 이며, 슬라이딩 윈도우에 포함된 센서 노드는 활성상태에서 데이터를 수신하거나 송신할 수 있다. 슬라이딩 윈도우의 중앙에 위치한 시간 슬롯에 포함된 센서 노드는 송신 센서 노드로서 주위로 자신의 데이터를 전송한다. 예를 들면, (그림 5)과 같이 센서 노드가 스케줄링되어 있다고 가정할 때, 각 시간에 따라 슬라이딩 윈도우 범위의 값을 수집한 센서 노드의 상태가 변환함으로써 데이터의 송신과 수신 여부를 결정된다.



(a) $t=t_0$



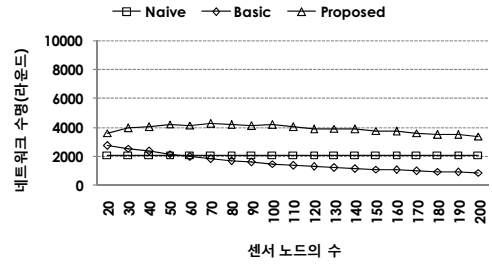
(그림 5) 슬라이딩 윈도우를 이용한 센서 노드의 상태 변환

4. 성능평가 및 분석

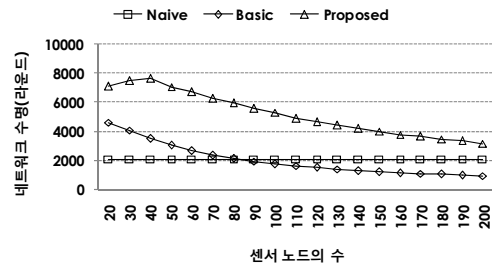
제안하는 알고리즘의 우수성을 보이기 위해 기존에 제안된 연속적인 스카이라인 질의 처리 알고리즘과 성능을 비교평가 하였다. 표 1 은 시뮬레이션에서 사용된 환경 변수이고, 미국 워싱턴주에서 측정된 온도 데이터 모델을 성능평가에 사용하였다. 센서 노드의 메시지 전송에 소모되는 에너지 모델은 {메시지 크기}*{전송 비용}+{증폭 비용}*{거리}²이며, 전송 비용은 50nJ/b, 증폭 비용은 100pJ/b/m² 으로 설정하였다. 메시지 수신에 소모되는 에너지 모델은 {메시지 크기}*{수신 비용}이며, 수신 비용은 50mJ/b 으로 설정하였다.

<표 2> 성능평가 환경	
파라미터	값
센서 개수	20~200
센서 식별자	4byte
센싱 데이터의 크기	4byte
비컨 데이터의 크기	1bit
d	1 ~ 3
k	10%
초기 센서 에너지	0.01J

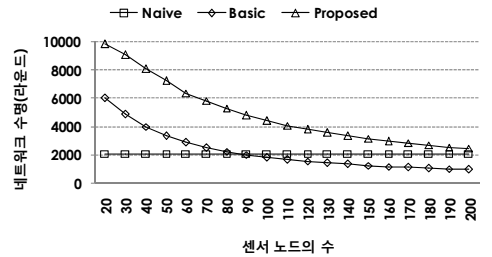
(그림 6) ~ (그림 8)은 센서 노드 수가 변경됨에 따라 TDMA 스케줄을 이용하여 모든 데이터를 기지국으로 전송하여 처리하는 기법(Naive)과 제안하는 기법의 네트워크 수명을 평가한 결과이다. Basic 기법은 슬라이딩 윈도우 기법을 사용하지 않고 단순히 비컨 교환을 이용하여 아웃라이어 데이터를 검출하는 방법이고, Proposed 는 슬라이딩 윈도우 기법을 적용한 방법이다. 센서 노드의 수가 작은 경우, Naive 기법보다 비컨을 교환하는 Basic 기법이 좋은 성능을 보이지만, 센서 노드의 수가 증가함에 따라 주변 센서 데이터의 수신(overhearing)이 증가하기 때문에 성능이 나빠진다. 슬라이딩 윈도우 기법을 함께 사용하여 불필요한 데이터의 수신을 방지한 경우, Naive 기법보다 평균 86.92% ~ 133.67% 네트워크 수명이 연장된다. d 값이 증가함에 따라 아웃라이어의 수는 줄어들어 일부 센서 노드의 수 (20~40) 일때 성능이 좋아지지만, 센서 노드의 수가 증가함에 따라 전체적으로 overhearing 이 증가하여 네트워크 수명이 단축된다.



(그림 6) 네트워크 수명 비교(d=1)



(그림 7) 네트워크 수명 비교(d=2)



(그림 8) 네트워크 수명 비교(d=3)

5. 결론 및 향후 연구

본 논문에서는 비컨 교환을 이용한 아웃라이어 탐지 기법을 제안한다. 데이터 인지 비컨 교환 기법을 통해 데이터 전송량을 줄이고, 슬라이딩 윈도우 기법을 이용하여 불필요한 데이터의 수신(overhearing)을 방지하였다. 향후 연구는 트리 구조와 클러스터링 구조와 같은 계층적 네트워크 토폴로지를 가진 네트워크로 제안하는 기법을 확장하는 것이다.

참고문헌

[1] S. Ramaswamy, R. Rastogi, and K. Shim. "Efficient algorithms for mining outliers from large data sets." In SIGMOD' 00, New York, NY, USA, 2000.
 [2] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. "LOF:identifying density-based local outliers." In

SIGMOD' 00, New York, NY, USA, 2000.

- [3] S. D. Bay and M. Schwabacher. "Mining distance-based outliers in near linear time with randomization and a simple pruning rule." In KDD' 03, New York, NY, USA, 2003.
- [4] A. Lazarevic and V. Kumar. "Feature bagging for outlier detection." In KDD' 05, New York, NY, USA, 2005.
- [5] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos. "Online outlier detection in sensor data using non-parametric models." In VLDB 2006, 2006.
- [6] B. Sheng, Q. Li, W. Mao and W. Jin. "Outlier Detection in Sensor Networks." In MobiHoc'07, 2007.