논문 2011-48CI-5-13

콘텐츠 기반 무선 센서 네트워크 이상 탐지 기법

(A Contents-Based Anomaly Detection Scheme in WSNs)

이 창 석*, 이 광 휘*

(Chang-Seuk Lee and Kwang-Hui Lee)

요 약

데이터 중심적인 네트워크인 무선 센서 네트워크는 대량의 센서 노드들이 광범위한 지역에 조밀하게 분산 배치되어 동작한다. 센서 노드들은 일반적으로 열린 환경에서 독립적으로 동작하기 때문에 보안 공격에 취약하다. 본 논문에서는 무선 센서 네트워크를 위한 콘텐츠 기반 이상 탐지 기법을 제안한다. 제안 기법은 무선 센서 네트워크의 특징인 특정한 현상을 여러 개의 센서 노드가 동시에 감지한다는 특성과 센서 노드에서 측정된 데이터인 콘텐츠는 어떤 특정 범위 안에서 변한다는 특성을 이용한다. 제안 기법은 훈련 단계, 적용 단계와 보정 단계로 구성되며 적용 단계에서 거리 기반 이상 탐지(distance-based anomaly detection) 기법을 이용하여 얻게 된 이상치 후보를 보정 단계로 보낸다. 보정 단계는 동일한 현상을 동시에 감지한 센서 노드들의 데이터로 구성된 콘텐츠 테이블과 이상치 후보를 비교, 분석함으로써 이상 탐지 기법의 성능을 향상시킨다. 시뮬레이션을 통해 제안 탐지 기법이 높은 탐지율과 낮은 오탐율을 가진다는 것을 확인할 수 있었다.

Abstract

In many applications, wireless sensor networks could be thought as data-centric networks, and the sensor nodes are densely distributed over a large sensor field. The sensor nodes are normally vulnerable in terms of security since they are very often deployed in a hostile environment and open space. In this paper, we propose a scheme for contents-based anomaly detection in wireless sensor networks. In this scheme we use the characteristics of sensor networks where several nodes surrounding an event point can simultaneously detect the phenomenon occurring and the contents detected from these sensors are limited to inside a certain range. The proposed scheme consists of several phases; training, testing and refining phases. Anomaly candidates detected by the distance-based anomaly detection scheme in the testing phase are sent to the refining phase. They are then compared in the sink node with previously collected data set to improve detection performance in the refining phase. Our simulation results suggest the effectiveness of the proposed scheme in this paper evidenced by the improvements of the detection rate and the false positive rate.

Keywords : 무선 센서 네트워크, 이상 탐지, 침입 탐지, 콘텐츠 기반

Ⅰ. 서 론

무선 센서 네트워크는 수백 혹은 수천 개의 센서 노드들로 구성되어 있으며 센서 노드들은 제한된 에너지, 낮은 계산 능력과 작은 메모리 용량 등 제한적인 자원을 가지고서 광범위한 지역에 조밀하게 분산적으로 배치되어 동작하는 경우가 일반적이다. 이렇게 배치된 각 센서 노드는 측정된 데이터를 싱크 노드로 전송한다. 무선 센서 네트워크는 군사, 비즈니스, 산업, 그리고 의료 분야 등 다양한 분야에서 활용 되고 있다. 예를 들어서 군사적 목적의 감시, 환경과 주거 모니터링과 홈 헬스 케어 등의 응용들이 있다[1].

무선 센서 네트워크에서 측정된 데이터는 여러 원인에 의하여 변형될 수 있으며, 예를 들어 센서 노드의 오작동으로 인하여 데이터에 노이즈가 추가되어 에러가 발생할 수 있으며, 또한 중복된 데이터를 수신하고 이로 인해 데이터가 불일치하는 경우가 발생할 수 있다.

* 정회원, 창원대학교 컴퓨터공학과
(Dept. of Computer Engineering, Changwon National University)
※ 이 논문은 2009-2010년도 창원대학교 연구비에 의하여 연구되었음.
접수일자: 2011년6월30일, 수정완료일: 2011년8월29일

또한 통신과정에서 데이터가 분실될 수도 있다. 이뿐 아니라 서비스 거부, 블랙 홀, 도청 등과 같은 악의적인 공격들에 의해 데이터가 오염되거나 변질될 수 있다. 즉, 악의적인 공격자들이 데이터를 조작할 수 있다[2]. 이러한 환경에서 수집된 센서 데이터는 신뢰성이 문제될 수 있으며 데이터의 품질과 처리 결과에 큰 영향을 줄 수 있다[3].

이러한 데이터의 품질과 관련하여 잘못된 데이터에 대한 정의를 몇 연구자들이 하였다. 측정된 데이터들의 집합과 불일치하는 값들을 보통 이상치(anomaly)라고 하며 Hawkins[4]는 이상치를 다른 관찰들과 서로 다른 관찰로서 다른 방법을 사용하여 생성된 결과라는 의심을 불러일으키는 관찰로 정의한다. 한편 Barnett과 Lewis[5]는 데이터 집합의 나머지와 불일치하는 관찰(혹은 관찰의 부분집합)을 이상치로 정의한다. 그리고 무선 센서 네트워크에서 이상치는 측정된 데이터의 정상적인 패턴과 두드러지게 벗어나는 측정값들로 정의된다[6].

이상치 탐지에 관련된 연구들은 크게 모델 기반(model-based) 기법, 밀도 기반(density-based) 기법, 군집 기반 기법(clustering-based), 거리 기반(distance-based) 기법 등으로 구분될 수 있다[7]. 모델 기반 기법은 확률분포 인수를 계산하여 확률분포 모델을 생성하고 이 모델과 부합되지 않는 객체를 이상치로 판단한다. 이 기법은 통계 분포를 알 수 없거나 훈련 데이터가 없는 경우에는 모델 생성이 어렵다는 단점을 가지고 있다. 밀도 기반 기법은 특정한 거리 이내에 속하는 객체의 수를 나타내는 객체의 밀도를 구하여 밀도가 낮으면 이상치로 판단한다. 이 기법은 데이터가 상이한 밀도를 가지는 영역을 포함하는 경우에는 정확하게 이상치를 식별할 수 없는 단점이 있다. 군집 기반 기법은 관련된 객체들의 군집을 탐색하는 것으로 다른 군집들과 멀리 떨어진 작은 군집들을 폐기하는 방법이다. 생성된 이상치의 집합은 사용된 군집수와 데이터에 존재하는 이상치의 존재유무에 따라서 왜곡될 수 있으며 특정한 데이터 유형에만 적합하다. 거리 기반 기법은 객체들 간의 거리를 계산하여 거리가 멀수록 이상치로 판별되는 확률을 높게 추정하는 방법이다. 전역 임계값(threshold)을 사용하기 때문에 상이한 밀도를 가지는 영역의 데이터 집합은 처리할 수 없으며 임계값의 선택에 민감하게 좌우된다.

무선 센서 네트워크에서 에너지, 메모리, 계산 용량과 통신 대역폭 등의 자원이 한정되어 있으므로 이러한 특징을 고려하여 에너지 소모를 최소화 하고 작은 용량의 메모리를 사용하면서 계산 작업을 줄이는 이상 탐지(anomaly detection) 기법이 필요하다. 그리고 무선 센서 네트워크를 위한 이상 탐지는 네트워크 트래픽을 줄이고 네트워크의 수명을 연장하기 위한 통신 오버헤드를 감소시키는 방안이 필요하다. 즉, 무선 센서 네트워크의 수명은 제한된 에너지 자원의 소비를 최소화함으로써 연장될 수 있다. 에너지 소모는 계산보다는 통신할 때 더 많은 에너지를 소모하는 것이 일반적이다.

무선 센서 네트워크에서 특정한 현상을 감지한 데이터인 콘텐츠는 어떤 특정 범위 안에서 변하는 특성을 가지며 동일한 현상을 적어도 하나 이상의 센서 노드들이 감지해서 동일한 콘텐츠를 싱크 노드로 송신할 수 있다. 이러한 특성을 이용해서 본 논문에서는 온도, 소리와 스피드 등과 같은 환경적 요소를 감지하는 데이터 중심적인 무선 센서 네트워크에서 사용될 수 있는 콘텐츠 기반 이상 탐지 기법을 제안한다. 제안 기법은 이상 탐지에 필요한 계산들이 싱크 노드에서 이루어짐으로써 센서 노드들의 자원 소비를 최소화 하고 통신 오버헤드를 줄일 수 있다. 본 논문에서 시뮬레이션을 통해 환경에 따른 최적의 임계값을 찾고 제안 기법이 오탐율(false alarm rate)을 낮추며 탐지율(detection rate)은 높이는 성능을 가진다는 것을 보여주었다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 무선 센서 네트워크에서 기존 이상 탐지 기법의 관련 연구들에 대해 살펴본다. Ⅲ장에서는 본 논문에서 제안하는 이상 탐지 기법의 구조, 동작 및 특성에 대하여 설명하고 Ⅳ장에서 성능에 대한 시뮬레이션 결과를 보여준다. 그리고 마지막으로 결론과 향후 연구 방향을 언급한다.

Ⅱ. 관련 연구

센서 노드들은 일반적으로 누구나 접근이 가능한 열린 환경에서 독립적으로 동작하기 때문에 보안 공격에 취약하다[2]. 이에 대한 많은 연구가 진행되었으며 마지막 방어책으로 이상 탐지 기법이 연구되고 있다. 표 1은 기존에 제안된 무선 센서 네트워크를 위한 이상 탐지 기법들을 요약하였다. [8], [10]과 [13]은 탐지 노드가 분산적으로 배치되고 서로 협력해서 탐지하는 방법을 제안하였으며 본 논문에서는 이러한 협력적인 방법을 활용하였다. 기존에 제안된 이상 탐지 기법을 아래에서

표 1. 기존 무선 센서 네트워크 이상 탐지 기법
Table 1. Anomaly detection schemes in WSNs.

제안자	탐지 방법	탐지 위치	동작 계층	탐지 대상	망 이동성
C. E. Loo 등	Statistical-based 분산/비협력	센서 노드	네트워크 계층	Routing 공격 (Periodic error route attack)	Static/Mobile
Da Silva 등	Statistical-based 분산/비협력	모니터 노드	응용, 네트워크, MAC, 물리	Worm hole, 데이터 변경, selective forwarding, black hole, jamming	Static
Bhuse 등	Signature-based 분산/협력	센서 노드	응용, 네트워크, MAC, 물리	위장 공격(위조 패킷)	Static/Mobile
K. Ioannis 등	분산/협력	센서 노드	응용 계층	공격	Static
W. Du 등	Statistical-based 분산/협력	센서 노드	응용 계층	잘못된 위치 정보	Static
V. Chatzigiannakis 등	Statistical-based 하이브리드	그룹 대표 노드	응용 계층	상호 연관된 이상 데이터 (잘못된 데이터 삽입)	Static/Mobile

간략하게 설명한다.

Bhuse 등은 응용, 네트워크, MAC과 물리계층과 같은 다양한 계층에서 이상치를 탐지할 수 있는 여러 가지 경량의 기술들을 제안하였다[8]. 이 방법은 낮은 오버헤드로 인해 에너지를 효율적으로 사용한다. 센서 네트워크의 다양한 계층에서 존재하는 RSSI 값, round trip 시간 등의 시스템 정보를 이용한다. 물리와 응용 계층에서 제안된 기술은 높은 오탐율을 만드는 단점이 있다.

V. Chatzigiannakis 등은 센서 네트워크에 있는 서로 다른 노드들로부터 데이터를 수집하여 이상치를 탐지하는 방법을 제시하였다[9]. PCA(principal component analysis)를 적용하여 여러 센서 노드들로부터 수신된 여러 개의 Metrics을 처리한다. 이러한 접근 방법은 센서 노드들의 여러 개의 그룹과 서로 관련된 이상치를 탐지하는데 유용하다. 무선 센서 네트워크의 에너지를 효율적으로 사용하는 방법이 부족하며 그룹 대표 노드가 상대적으로 많은 에너지를 소비하는 단점이 있다.

W. Du 등은 LAD(localization anomalies detection) 방법을 제안하였다[10]. 이웃 노드의 그룹 회원 정보와 노드의 배치 정보를 이용한다. 이 정보는 계산된 위치 정보와 관측한 위치 정보가 일치하는 지를 비교하여 일치하지 않을 경우에 LAD는 이상치로 판단한다. 무선 센서 네트워크의 특징인 이웃노드의 정보를 이용하여 효율적인 이상치를 탐지한다. 그러나 특정한 문제인 위치 정보의 이상치만을 탐지하는 한계를 가진다.

C. E. Loo 등이 제안한 방법은 라우팅 공격과 같은 네트워크 계층에서 발생되는 침입들을 탐지하기 위한 기법이다[11]. 클러스터링 알고리즘을 사용하여 정상 네트워크 트래픽 모델을 만들고 이 모델을 트래픽에 적용하여 이상치를 탐지한다. 라우팅 공격에 광범위하게 적용될 수 있는 트래픽 특성들을 이용함으로써 새로운 형태의 라우팅 공격을 탐지할 수 있는 장점이 있다. 센서 노드에서 이상 탐지가 실행됨에 따라 통신 오버헤드가 발생된다. 에너지를 효율적으로 사용하기 위한 방안이 제시되지 않았다.

Da Silva 등은 무선 센서 네트워크를 위한 분산된 IDS를 구축하는 고수준의 방법론을 제안하였다[12]. 특정한 모니터 노드에서 탐지된 이벤트들을 분석하여 만들어진 네트워크 행태에 기반을 둔 통계적 접근 방법을 적용하였다. 모니터 노드는 1홉 거리에 있는 침입자를 감시하는 이웃 노드들을 모니터링 한다. IDS가 네트워크에 분산 배치됨에 따라 분산 시스템의 확장성과 견고함을 가진다. 하지만 모니터 노드를 어떻게 결정할 것인지와 모니터 노드가 얼마나 필요한지를 언급하지 않았다.

K. Ioannis 등은 센서 노드들이 서로 협력하여 침입을 탐지하는 방법을 제안하였다[13]. 침임 탐지를 위해 센서 노드들이 협력함으로서 얻을 수 있는 이점들과 이론적 제한 사항들을 제시하였으며 일반적 공격 모델에서 성공적으로 침입자를 탐지할 수 있음을 실험을 통해서 보여주었다. 그러나 공격자가 한 명이라는 특정한 환경에서 동작하는 단점이 있으며 에너지를 효율적으로 사용하기 위한 방안이 없다.

기존의 관련 연구들은 무선 센서 네트워크의 자체 망에 대한 이상 여부를 탐지하는 방법들만이 제안 되었고

연구되었다. 그러나 수집된 센서 데이터의 오류나 악의적인 공격에 의한 데이터 조작 등과 같은 무선 센서 네트워크의 콘텐츠에 대한 이상치를 탐지하는 방법들은 많은 연구가 되지 않았으며 무선 센서 네트워크를 위한 효과적이고 효율적인 방법들이 제시되지 않았다.

Ⅲ. 콘텐츠 기반 이상 탐지 기법

3.1 콘텐츠 기반 이상 탐지 모델

무선 센서 네트워크는 데이터 중심적인 네트워크의 특징을 가진다. 또한 온도, 소리와 스피드 등과 같은 환경적 요소를 측정하는데 주로 사용되며 이런 요소들은 어떤 특정 범위 안에서 변하는 특성이 있다. 본 논문에서는 이상치를 탐지하기 위해 Knorr 등[5]에 의해 처음 제안된 거리 기반 이상 탐지 (distance-based anomaly detection) 기법을 적용하였다. 이상치는 $DB(p, D)$로 정의되며 이는 주어진 패턴 x에 대하여 참조 데이터 중 최소 비율 $p \times 100\%$ 이상의 데이터가 x로부터 거리 D보다 멀리 위치하는 패턴이다. 비율 p와 거리 D는 사용자에 의해 정해지는 모수이기 때문에 이 모수들의 변화에 따라서 성능의 편차가 큰 단점이 있다. 본 논문에서는 이러한 단점을 보정 단계(refine phase)에서 보완함으로써 기존의 이상 탐지 기법의 성능을 향상시킨다.

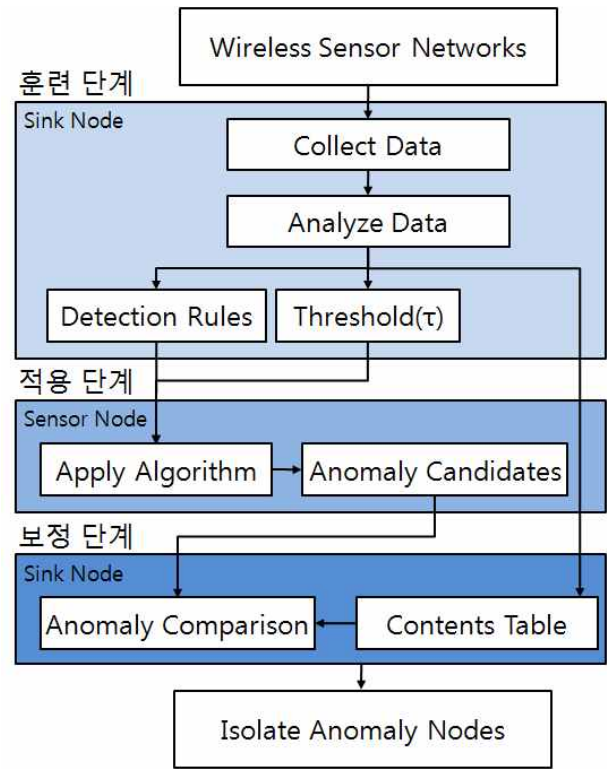

그림 1. 제안 모델의 동작 과정
Fig. 1. Workflow of anomaly detection.

본 논문에서 제안하는 이상 탐지 기법은 훈련 단계(training phase)와 적용 단계(testing phase), 그리고 보정 단계(refine phase)로 구성된다. 그림 1에서 제안 모델의 동작 과정을 단계별로 보여준다.

제안하는 이상 탐지 기법은 수집된 데이터를 싱크 노드에서 분석하고 처리한다. 센서 노드는 훈련 단계에서 만들어진 임계값을 이용하여 이상치 발생 여부를 감시하고 이상치가 발생하면 싱크 노드에게 이상치 후보군을 전달한다. 이상치 후보군은 싱크 노드에서 보정 단계를 거치면서 최종적으로 이상치 여부를 판단하게 된다.

본 논문에서 사용되는 무선 센서 네트워크는 모니터링 지역에서 센서 노드가 밀집되어 있어 특정한 현상을 적어도 하나 이상의 센서 노드가 동시에 감지한다고 가정한다. 그러므로 특정한 한 현상은 한 개 이상의 노드에 의해서 싱크 노드에게 전달되고 이렇게 전달된 데이터는 콘텐츠 테이블에 저장되어서 보정 단계에서 이상치 후보군과 비교할 때 사용된다.

본 논문에서 제안하는 기법은 높은 계산과 처리 능력을 필요로 하는 훈련 단계와 보정 단계는 싱크 노드에서 처리되고 임계값과 비교하여 이상치 여부를 판단하는 적용 단계는 센서 노드에서 처리된다. 이상 탐지를 위한 모듈을 분산 배치함으로써 무선 센서 네트워크의 에너지 소비와 통신 오버헤드를 최소화하는 구조를 가지도록 하였다.

이상 탐지 모델은 높은 탐지율(true positive rate, TPR)과 낮은 오탐율ffFalse positive rate, FPR)이 요구되고 탐지율과 오탐율이 성능 평가 지표로 사용되는 것이 일반적이다.

표 2. 탐지 파라미터들
Table 2. Detection parameters.

파라미터	정의
TP(True Positive)	이상 징후가 발생한 후 실제 경보가 발생한 경우
FP(False Positive)	트래픽에 대해 잘못된 경보를 발생한 경우
TN(True Negative)	이상 징후가 없으며 경보를 발생하지 않는 경우
FN(False Negative)	이상 징후가 발생한 후 탐지를 못한 경우

탐지율은 식 (1)과 같이 계산한다.

$$TPR = \frac{TP}{TP + FN} \quad (1)$$

오탐율은 식 (2)와 같이 계산한다.

$$FPR = \frac{FP}{FP + TN} \quad (2)$$

표 2는 탐지율과 오탐율 계산에서 사용되는 파라미터들을 나타낸다.

3.2 훈련 단계(Training Phase)

훈련 단계는 싱크 노드에서 이루어지며 센서 노드로부터 수집된 데이터를 트래픽 표본으로 만들어서 정상적인 데이터와 이상치를 식별하기 위한 모델을 만든다. 트래픽 표본은 정상적인 데이터와 이상치를 모두 가지고 있으며 이상치는 정상적인 데이터보다 통계적으로 발생하는 횟수가 적다고 가정한다. 즉, 이상 데이터의 발생은 특정 비율보다 낮다고 생각한다. 수집된 데이터는 데이터 분석을 하게 되는데, 이 과정은 데이터의 범위를 일치시키고 분포를 유사하게 만들기 위해 데이터를 정규화 처리한다. 정규화가 된 데이터는 특정 관측값(t_i)이 자료의 평균(t)으로부터 몇 개의 표준 편차(t)만큼 떨어져 있는가를 나타내며 상대적 위치를 말하는 척도로 일반적으로 사용된다.

평균값을 사용하는 정규화 방법을 이용하며 정규화는 식 (3)과 같이 계산한다.

$$Normalised\, t_i = \frac{t_i - \text{평균}(t)}{\text{표준 편차}(t)},\ t \in T \quad (3)$$

(T는 정규분포 확률 변수)

데이터 분석 과정이 끝나면 정규화가 된 트래픽 표본을 이용하여 임계값을 구하는 계산 과정을 수행한다. 이 계산은 임계값의 증가에 따른 TN(true negative)값과 TP(true positive)값을 비교하여 최적의 임계값을 구한다. 최적의 임계값은 TN값이 가장 작을 때 TP값이 가장 큰 경우의 임계값을 선택한다. 이 단계에서 만들어진 임계값은 센서 노드의 이상 탐지 모듈에서 이상 탐지 후보군을 생성할 때 사용된다.

3.3 적용 단계(Testing Phase)

적용 단계는 정상 데이터 또는 이상 데이터 여부를 판단하여 이상치 후보를 만드는 단계이다. 이 단계는 센서 노드에서 콘텐츠를 검사하여 이상 징후의 발생을 싱크 노드에게 통보하며 전달되는 패킷의 데이터 필드에 이상 징후의 발생 여부 필드를 포함시킨다. 이상 징후를 통보하는 별도의 패킷을 만들지 않음으로써 통신 오버헤드를 최소화 할 수 있다.

Algorithm 1. Anomaly Detection

```
INPUT: 데이터 D;
OUTPUT: 이상치 후보 C;
1: for i = 1 to count(D) do
2:    if( D_i >= Threshold(τ) )
3:       C = D_i as anomalous;
4:    return 이상치 후보 C;
5: end for
```

이상 탐지 모듈이 적재된 센서 노드는 임계값을 이용하여 이상 징후의 발생 여부를 감시한다. 싱크 노드로 전송된 이상치 후보는 보정 단계를 거쳐서 최종적으로 이상치 여부를 판단하게 된다.

3.4 보정 단계(Refine Phase)

본 논문에서 사용되는 무선 센서 네트워크는 모니터링 지역에서 특정한 현상을 적어도 하나 이상의 센서 노드에서 감지하여 싱크 노드로 전송한다. 동일한 현상을 감지한 센서 노드들에게서 전송된 데이터는 싱크 노드의 콘텐츠 테이블에 저장되어 관리된다. 콘텐츠 테이블은 동일한 현상을 동시에 감지한 센서 노드들과 각 센서 노드들이 측정한 데이터로 구성된다.

보정 단계는 싱크 노드에서 이루어지며 적용 단계에서 이상 데이터로 판단되어 싱크 노드로 전송된 이상치 후보와 싱크 노드에서 관리하고 있는 콘텐츠 테이블을 비교한다. 이상치 후보와 콘텐츠 테이블에 저장된 데이터를 비교하여 일치하면 정상 데이터로 분류하고 해당 센서 노드를 이상치 후보에서 정상 센서 노드로 분류한다. 불일치할 경우는 이상 데이터로 최종 판단한다. 보정 단계를 통해서 제안 기법의 탐지율을 향상 시킬 수 있고 오탐율을 줄일 수 있다.

Algorithm 2. Anomaly Comparison

INPUT: 이상치 후보 C,
콘텐츠 테이블 N;
1: for i = 1 to count(N) do
2: if(C == N_i)
3: Label C as normal;
4: else
5: Label C as anomalous;
5: end for

Ⅳ. 시뮬레이션

4.1 시뮬레이션 환경

제안한 콘텐츠 기반 이상 탐지 기법의 성능을 평가하기 위해 Naval Research Laboratory에서 개발한 무선 센서 네트워크 시뮬레이터를 사용하였다[14]. 이 시뮬레이터는 NS-2 시뮬레이터(version 2.27)를 이용하여 무선 센서 네트워크 환경을 구축하였다. 본 논문에서 사용된 시뮬레이션 환경은 25개의 센서 노드와 1개의 싱크 노드 그리고 일산화탄소를 발생시키는 1개의 노드로 구성되었다. 모든 노드들은 450m x 450m 공간에 100m 간격으로 배치되고 센서 노드들은 설치된 이후에는 이동을 하지 않는다고 가정하였다. 이 시뮬레이션 환경은 하나의 현상만이 있다고 가정하였지만 여러 개의 현상이 무선 센서 네트워크에 존재하더라도 동일한 결과를 예상할 수 있다. 또한 제안 기법은 무선 센서 네트워크에서 생성된 콘텐츠를 이용하는 방법이기 때문에 무선 센서 네트워크의 망 환경에는 큰 영향을 받지 않는다.

제안 기법 중 훈련 단계를 시뮬레이션 하는 시나리오와 적용 단계와 보정 단계를 시뮬레이션 하는 시나리오를 작성하여 시뮬레이션 하였으며 각 시뮬레이션은 200초 동안 실행되었다. 일산화탄소를 감지한 센서 노드에서 싱크 노드로 전송된 데이터를 분석하였다.

4.2 시뮬레이션 결과

거리 기반 이상 탐지 기법은 임계값의 선택에 민감하게 성능이 좌우된다. 그리하여 본 환경에서 이러한 임계값을 구하는 계산 과정을 먼저 수행하였다.

그림 2는 임계값 변화에 따른 TN(true negative)값과 TP(true positive)값을 보여 주고 있다. 수평축은 임계값의 변화를 나타내며 수직축은 빈도를 나타낸다. 이 계산의 결과는 임계값으로 $2.8 \leq \tau \leq 3$ 설정하면 높은 성능의 이상 탐지 시스템을 구축할 수 있다는 것을 보여 주었다. 여기서 구한 임계값은 센서 노드의 탐지 모듈에 적용되어서 이상 탐지 후보군을 생성할 때 사용된다.

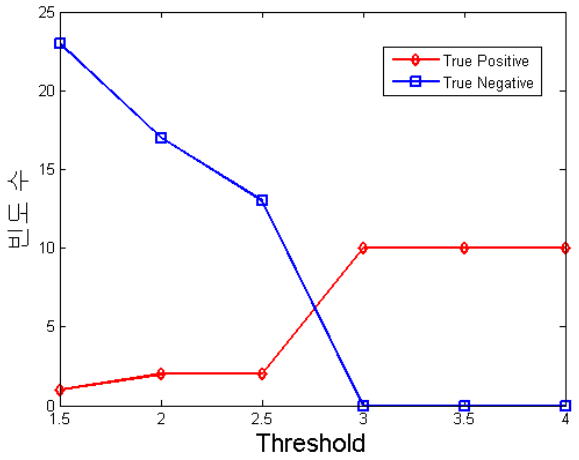

그림 2. 임계값에 따른 TN과 TP 변화
Fig. 2. The distribution of TN and TP.

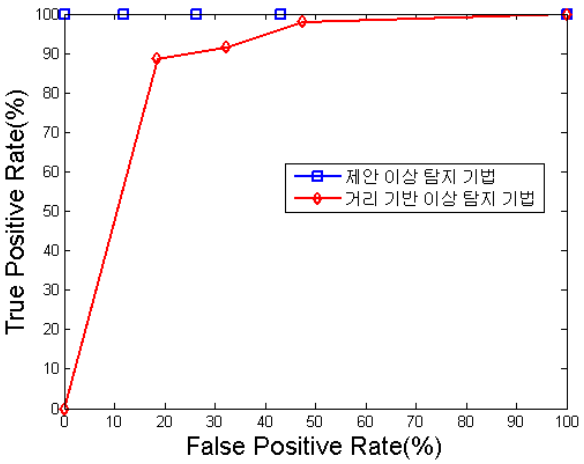

그림 3. 제안 이상 탐지 기법의 ROC 비교
Fig. 3. ROC of proposed anomaly detection.

그림 3은 본 논문에서 제안한 이상 탐지 기법의 ROC(receiver operating characteristics) 곡선을 보여준다. 수평축은 오탐율을 나타내고 수직축은 탐지율을 백분율로 나타낸다. 적용 단계에서 만들어진 이상 탐지 후보군의 ROC 곡선은 전형적인 거리 기반 이상 탐지 기법의 ROC 곡선 형태를 보여주며 오탐율이 25%일 때 90%의 탐지율 성능을 보여준다. 보정 단계를 완료한 후의 콘텐츠 기반 이상 탐지 기법은 오탐율이 0%이고 탐지율이 100%임을 보여주며 가장 좋은 이상 탐지 기법임을 나타낸다.

V. 결론 및 향후 연구 방향

본 논문에서는 무선 센서 네트워크의 특징인 특정한 현상을 여러 개의 센서 노드가 동시에 감지한다는 특성과 센서 노드에서 측정된 데이터인 콘텐츠는 어떤 특정 범위 안에서 변한다는 특성을 이용하여 콘텐츠 기반 무선 센서 네트워크 이상 탐지 기법을 제안하였다. 시뮬레이션을 통해 제안한 기법이 우수한 성능을 나타낸다는 것을 입증하였다.

향후 연구로는 본 논문에서 제안한 탐지 기법을 실제 무선 센서 네트워크 환경에 적용하여 실제 환경에서 얼마나 효율적으로 이상치를 탐지하는지를 연구하고자 한다. 그리고 이웃 노드의 개수 변화에 따른 성능 차이를 알아보고 최적의 이웃 노드를 구하는 연구와 이상치가 발생된 해당 센서 노드를 네트워크에서 고립시키는 방안에 관한 연구를 하고자 한다.

참 고 문 헌

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirici, “Wireless sensor networks: a survey,” IEEE Trans. Systems, Man and Cybernetics (B), Vol. 38, pp. 393-422, 2002.

[2] A. Perrig, et. al, “Security in wireless sensor networks,” CACM, Vol. 47, No. 6, pp. 53-57, 2004.

[3] F. Martinicic, et. al., “Distributed event detection in sensor networks,” in Proc. of the International Conference on Systems and Networks Communication, pp. 43-48, 2006.

[4] D. M. Hawkins, “Identification of outliers,” Chapman and Hall, London, 1980.

[5] V. Barnett, T. Lewis, “Outliers in statistical data,” John Wiley Sons, New York, 1994.

[6] E. M. Knorr, R. T. Ng, V. Tucakov, “Distance-based outliers: algorithms and applications,” The VLDB Journal, Vol. 8, Issue 3-4, pp. 237-253, 2000.

[7] Pang-Ning tan, M. Steinbach, V. Kumar, “Introduction to Data Mining,” Addison-Wesley, 2006.

[8] Bhuse, V., Gupta, A., “Anomaly Intrusion Detection in Wireless Sensor Networks,” J. High Speed Networks, pp. 33-51, 2006.

[9] V. Chatzigiannakis, S. Papavassiliou, “Diagnosing Anomalies and Identifying Faulty Nodes in sensor Networks,” IEEE Sensors Journal, Vol. 7, 2007.

[10] W. Du, L. Fang, P. Ning, “LAD: Localization Anomaly Detection for Wirelesss Sensor Networks,” Journal of Parallel and Distgributed Computing, Vol. 66, 2006.

[11] C. E. Loo, M. Y. Ng, C. Leckie, M. Palaniswami, “Intrusion Detection for Routing Attacks in Sensor Networks,” International Journal of Distributed Sensor Networks, Vol. 2, No. 4, 2006.

[12] A. P. R. da Silva, M. H. T. Martins, et. al, “Decentralized Intrusion Detection in Wireless Sensor Networks,” in Proc. of the 1st ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks (Q2SWinet'05), pp. 16-23, New York, USA, 2005.

[13] K. Ioannis, B. Zinaida, G. Thanassis, F. C. Felix, D. Tassos, “Cooperative Intrusion Detection in Wireless Sensor Networks,” in Proc. of the 6th European Conference on Wireless Sensor Networks, 2009.

[14] I. Downard, “Simulation Sensor Networks in NS-2,” Technical Report NRL/FR/5522-04-10073, Naval Research Laboratory, Washington, D.C., U.S.A., May 2004.

저 자 소 개

이 창 석(정회원)

1996년 창원대학교 전자계산학과 학사 졸업.
1998년 창원대학교 전자계산학과 석사 졸업.
2002년 창원대학교 컴퓨터공학과 박사과정 수료.
2000년～2008년 창신대학 인터넷정보과 교수
2009년～현재 창원대학교 컴퓨터공학과 연구원
<주관심분야 : 무선 센서 네트워크, 네트워크 관리 시스템, 분산 시스템>

이 광 휘(정회원)

1983년 2월 고려대학교 전자공학과 (공학사)
1985년 2월 고려대학교 전자공학과 (공학석사)
1989년 2월 고려대학교 전자공학과 (공학박사)
1991년～1992년 영국 Wales 대학(Swansea) 및 Newbridge Networks 연구원
1992년～1993년 영국 Reading 대학 연구원
1997년～1998년 영국 Reading 대학 연구원
2001년～2002년 영국 UCL 방문연구원
2007년～2008년 영국 Reading 대학 방문연구원
1988년～현재 창원대학교 컴퓨터공학과 교수
<주관심분야: 무선 센서 네트워크, QoS/네트워크 관리, 멀티캐스팅, 모바일 컴퓨팅>