산업용 IoT환경에서 기계학습을 통한 비정상 디바이스 판별

노태균*, 이수연**, 정태명***
*성균관대학교 전자전기컴퓨터공학과
**R&D센터, 지어소프트
***성균관대학교 소프트웨어학과
*tkroh0198@skku.edu
**sylee@gaesoft.co.kr
***tmchung@skku.edu

A Study of Improving System Security Using Abnormal Devices Detection in Industrial IoT Environment

Tae-Kyun Roh*, Soo-Yeon Lee**, Tai-Myung Chung***
*Dept of Electrical and Computer Engineering, SungKyunKwan University
**Research & Development Center, Gaesoft
***Dept of Software Engineering, SungKyunKwan University

요 약

다양한 센서들과 디바이스들이 실시간으로 정보를 주고받는 산업 IoT환경에서 싱크노드에게 속하는 하위 센서 및 디바이스들을 통한 데이터 손실 및 시스템 마비를 발생시킬 수 있는 상황이 발생될 수 있다. 따라서 본 논문은 위의 상황을 고려하여 센서 및 디바이스의 정상범주를 파악하고 비정상적인 디바이스를 판별을 통해 시스템 보안성을 향상시키는 방안을 제시한다. 싱크노드에 속하는 센서 및 디바이스들의 로그데이터를 통해 주성분 분석법을 통해 데이터의 차원을 감소시키고 차원 감소시킨 데이터를 K-means 클러스터링 알고리즘에 적용하여 정상범주 내에 속하지 않는 디바이스를 판별하여 비정상 센서 및 디바이스를 판별한다. 비정상 데이터로 판별된 센서 및 디바이스의 모니터링을 통해 시스템의 보안성을 향상시킬 수 있도록 한다.

1. 서론

Industry 4.0의 시작으로 IoT 기기들은 여러 산업 환경에서 물리적 세계와 가상 세계를 연결하는 매개체 역할을 하고 있다. 산업용 IoT(Industrial Internet of Things)에서는 여러 ICT(Information & Communication Technology) 기술들과 무선센서 네트워크 및 통신 기술들이 상호 연결되어 작동하고 있다. 이러한 환경에서는 비정상 디바이스의 침입으로 인해 시스템의 장애나 데이터 유출이 발생할 수 있다. 본 논문에서는 주성분 분석(PCA)과 K-means 클러스터링 알고리즘을 통해 디바이스의 로그 데이터를 분석하여 비정상 디바이스를 판별하고 판별된 비정상 디바이스의 접근을 제어하는 메커니즘에 대해 제안한다.

2. 관련연구

1) 주성분 분석(PCA)

위의 환경에서 비정상 디바이스를 판별하기 위한 알고리즘으로는 주성분 분석(PCA)과 K-means 클러스터링 알고리즘을 사용하여 판별한다. 주성분 분석은 고차원의 데이터에서 패턴을 찾는 알고리즘중 하나이다. 기계학습 알고리즘에서 데이터를 인공 신경망에 넣기 전에 전처리과정에서 많이 사용되는 분석법이다. 주성분 분석을 통해 데이터의 범위를 재조정하고 데이터의 평균을 0으로 맞춰줌으로써 주성분 분석은 고차원의 데이터를 중요한 차원의 성분을 유지시키며 저차원의 데이터로 만들어주는 분석법이다[2]. 차원을 감소시키는 이유는 K-means 클러스터링 알고리즘에 다차원의 데이터를 입력할 경우 결과 값의 정확도가 낮아지게 된다. 따라서 주성분 분석으로 차원을 감소시켜 K-means 클러스터링 알고리즘에 적용한다.

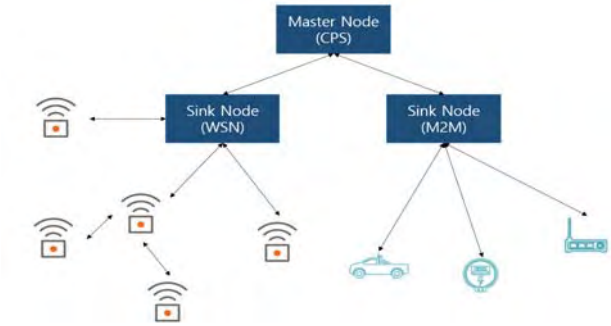
2) K-means 클러스터링

K-means 클러스터링 알고리즘은 각 구간을 나눈 다음 중심을 찾고 찾은 중심을 기준으로 다시 구간을 나누어 변경사항이 있을 경우 다시 중심을 찾아나가는 방식으로 평균 군집화를 시키는 과정을 의미한다. 사람이 개입하여 k개의 클러스터 개수를 선택한 후 데이터가 분포된 공간상에 클러스터 중심으로 가정할 임의의 지점 k개를 선택한다. 각 데이터는 근처에 있는 클러스터의 중심에 할당되고 각 클러스터 중심을 해당 클러스터에 속한 데이터들의 평균으로 조정한다. 위의 과정을 클러스터의 중심이 변하지 않을 때까지 반복하게 되고 클러스터의 중심이 변하지

않는다면 클러스터의 중심으로 결정되는 군집화 알고리즘을 의미한다.

3. 본론

산업용 IoT환경은 CPS(Cyber Physical System)가 마스터노드로서 하위 싱크노드들을 관리하고 제어한다[1]. 하위 싱크노드로서는 WSN(Wireless Sensor Network)과 M2M(Machine to Machine)을 싱크노드로 관리하고 있다. 본 논문에서는 CPS를 마스터 노드, WSN과 M2M을 싱크노드로 가정한다. WSN은 같은 네트워크 안의 모든 센서들의 집합체를 의미하며 모든 센서들을 실시간으로 관리하고 장소에 상관없이 컴퓨팅 환경에 접속 가능하게 한 임베디드 무선 네트워크 기술이다. M2M은 CPS에서 분석된 데이터를 받아 M2M에 속하는 디바이스들의 실시간 행동 제어 및 상태를 관리하고 제어한다. 그리고 여러 센서와 디바이스들이 지속적으로 추가되고 제거된다. (그림 1)은 싱크노드들의 하위 센서 및 디바이스의 연결 관계를 나타내는 그림이다.



(그림 1) 싱크노드와 센서 및 디바이스의 연결관계

위와 같은 상황에서 비정상적인 센서 혹은 디바이스의 추가로 인해 시스템 장애 및 데이터 손실 또는 유출이 발생할 수 있다. 따라서 새로운 디바이스의 추가되었을 때 추가된 디바이스의 로그 데이터를 사용하여 주성분 분석으로 데이터의 차원을 줄인다. 이 후 차원을 줄인 데이터를 K-means 클러스터링 알고리즘에 적용하여 디바이스의 정상범주를 확인하고 새로운 디바이스가 추가되었을 때 정상범주 내에서 벗어나는 오차 확률을 계산하여 비정상 디바이스를 판별할 수 있다. 싱크노드 하위 센서 및 디바이스의 로그 데이터는 싱크노드에서 판단할 수 있는 동일한 형태의 로그 데이터로 생성된다고 가정한다. (표 1)은 산업용 IoT에서 사용되는 데이터 세트의 예시이다.

<표 1> 산업용 IoT의 데이터 세트

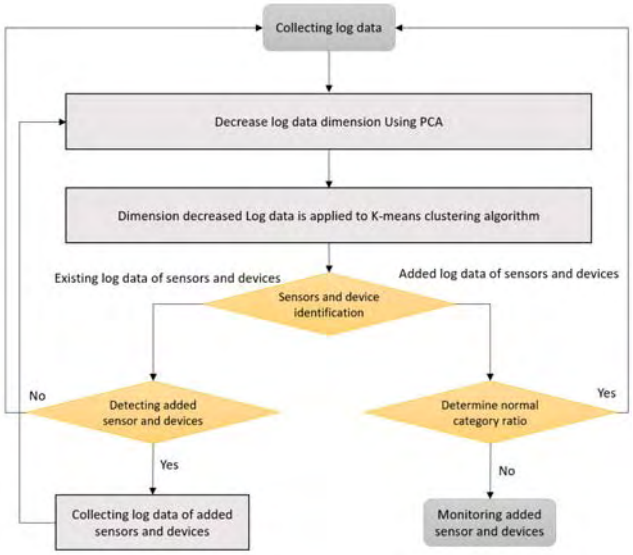
Demand_ Response	Area	Season	Energy	Cost	Distance

1) 주성분 분석을 통한 센서 및 디바이스의 차원감소

싱크노드는 하위 센서 및 디바이스를 관리 및 제어한다. 싱크노드들은 제어하는 센서 및 디바이스의 로그 데이터를 마스터노드에게 전송하게 되고 마스터노드에서는 로그 데이터들을 종합하여 K-means 클러스터링 알고리즘에 사용할 수 있는 데이터로 차원 감소시킨다. 본 논문에서는 6개의 데이터 세트을 3개로 줄여 차원감소 시켰다. 따라서 주성분 분석을 통해 센서 및 디바이스들의 로그 데이터의 차원을 낮추는 작업을 먼저 실행한다.

2) K-means 클러스터링 알고리즘을 통한 비정상 디바이스 및 센서 판별

주성분 분석을 통해 정제된 싱크노드 하위의 센서 및 디바이스의 데이터를 K-means 클러스터링 알고리즘에 적용시킨다. K-means 클러스터링 알고리즘에 적용시키면 싱크노드의 개수에 따라 클러스터가 생성된다. 이때 생성된 클러스터는 정상범주로 판단한다. 이 후 새롭게 등록된 센서 및 디바이스의 로그 데이터를 위의 설명에서 생성한 K-means 클러스터링 알고리즘에 적용하면 로그 데이터 개수에 따라 하나의 점으로 로그 데이터 개수만큼 점이 생성되며 이 점을 포인트라 부른다. 생성된 로그 데이터의 포인트를 정상 클러스터 범주 내에 속하는 비율을 계산한다. 정상 범주에 속하는 지를 판단하는 기준은 새로운 센서 및 디바이스의 포인트와 클러스터 중점과의 거리를 계산하여 판단하고 정상범주와 비정상범주에 속하는 포인트의 비율을 계산한다. 이후 관리자가 정한 비율이하로 낮아지게 되면 비정상 센서 및 디바이스로 간주하여 로그 데이터 분석 및 모니터링을 한다. 따라서 비정상 센서 및 디바이스를 발견하여 CPS의 마비 및 데이터 손실을 예방하고 관리할 수 있다.



(그림 2) 비정상 디바이스 판별 흐름도

위와 같은 방법으로 싱크노드 하위의 센서 및 디바이스의 로그 데이터를 바탕으로 주성분 분석법과 K-means 클러스터링 알고리즘을 통해 비정상 디바이스를 판별하고 판별된 센서 및 디바이스는 관리자 또는 시스템 내에서 모

니터링을 통하여 시스템의 마비 및 데이터 손실을 예방한다. (그림 2) 는 비정상 센서 및 디바이스의 판별 과정을 단계별로 보여준다.

4.결론

본 논문에서는 산업용 IoT환경에서 싱크노드들이 관리하는 센서 및 디바이스들의 로그 데이터를 통해 비정상 센서 및 디바이스를 판별하고 시스템 보안성을 향상시키는 메커니즘에 대하여 제안한다. 로그 데이터를 주성분 분석법으로 K-means 클러스터링 알고리즘의 정확성을 높이고 주성분 분석법으로 정확성을 높인 데이터를 K-means 클러스터링 알고리즘에 적용하여 클러스터를 형성한다. 이후 새롭게 등록된 센서 및 디바이스의 로그 데이터를 K-means 클러스터링 알고리즘에 적용하여 형성된 클러스터에 포함되는 데이터의 확률을 구하여 비정상 센서 및 디바이스를 판별하여 시스템의 보안성을 향상시킬 수 있다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음(과제번호 : 2015-0-00914)

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.B0184-15-1003, oneM2M Conformance 테스트 툴 및 QoS 기술 개발)

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (RF-2010-0020210)

참고문헌

[1] 노태균, 이수연, 정태명 “적절한 노드 선택에 관한 연구 : 산업용 IoT시스템에서 빠른 복구를 위한 우선순위 알고리즘을 사용한 대리노드 선택”, 한국 정보 추계 처리 학회, 2017

[2] 이수연, 정준권, 정태명 (2017). IoT 환경에서 우선순위 산출을 통한 네트워크 장애복구 연구”. 한국통신학회, 한국통신학회 학술대회논문집 , 194-195.

[3] Ahmed, Syed Hassan, Gwanghyeon Kim, and Dongkyun Kim. 2017, "Cyber Physical System: Architecture, Applications And Research

[4] Soo-Yeon Lee, Sa-Rang Wi, Eunil Seo, Jun-Kwon Jung, Tai-Myung Chung, “ProFiOt: Abnormal Behavior Profiling (ABP) of IoT devices based on a Machine Learning Approach”, 27th International Telecommunication Networks and Applications Conference (ITNAC), 2017

[5] Wu, Fang-Jing, Yu-Fen Kao, and Yu-Chee Tseng. “From wireless sensor networks towards cyber physical systems.” Pervasive and Mobile Computing 7.4 (2011):397-413.

[6] “An Efficient k-Means Clustering Algorithm: Analysis and Implementation”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 24, no 7, July, 2002

[7] Abolfazl Akbari, Arash Dana, Ahmed Khademzadeh,and Neda Beikmahdavi, “Fault Detection and Recovery in Wireless Sensor Network Using Clustering”, International Journal of Wireless & Mobile Networks, Vol 3, No 1, Feburuary, 2011