LPWA 네트워크 환경에서 다변량 가우스 분포를 활용하여 이상탐지를 위한 머신러닝 기법 연구

이상진 · 김기천*

건국대학교

Study of Machine Learning Method for Anormaly Detection Using Multivariate Gaussian Distribution in LPWA Network Environment

Sangjin Lee · Keecheon Kim*

Konkuk University

E-mail : vitamin8096@konkuk.ac.kr, kckim@konkuk.ac.kr*

요 약

최근 사물인터넷(IoT) 기술의 혁신적 발전과 함께 사물 간 초연결적 사회를 맞이하게 되었다. 본 논문은 사물인터넷의 LPWA 네트워크 환경 내에 발생할 수 있는 보안적인 부분에 초점을 맞추었으며, 기기의 비·의도적인 이상행위를 탐지 및 차단할 수 있는 차세대 IPS/IDS를 고려한 머신러닝 기법을 제안하고자 한다.

ABSTRACT

With the recent development of the Internet (IoT) technology, we have come to a very connected society. This paper focuses on the security aspects that can occur within the LPWA Network environment of the Internet of things, and proposes a new machine learning method considering next generation IPS / IDS that can detect and block unexpected and unusual behavior of devices.

키워드

Security, IoT, LPWA, Machine Learning

Ⅰ. 서 론

최근 사물인터넷(IoT)의 지능화 및 네트워크화 기술이 급속도록 발전함에 따라, 인간의 생활의 삶의 질이 향상되었으며 기업의 생산성 증대 및 혁신적인 공공서비스 가능하도록 돕고 있다.[1] 그 중, 대용량 또는 고속 통신에는 적합하지 않으나, 저전력으로 소량의 데이터를 가지고 넓은 범위의 통신을 할 때 유리한 LPWA(Low Power Wide Area) 네트워크 기술의 필요성이 확대 되고 사용 수도 증가하고 있다.[2]

따라서 LPWA 네트워크 환경에서 보안적인 측면 또한 고려 사항 중의 하나이며, 본 고에서는 LPWA 환경에서 동작하는 기기의 비·의도적인 이상행위를 탐지하기 위한 방식을 제안하고자 한다.

Ⅱ. 관련 연구

2-1. 이상탐지

이상탐지(Anomaly Detection)이란 이름에서 알 수 있듯이 특이한 것들을 찾아내는 것이다. '상하다' 라는 의미는 기존의 것들과 다르다는 이야기이며, 이상한 것들의 본질은 '모르고 있음이 알려지지 않은 것' 이라는 데 있다. 즉, 관측하고 이해 할 수 있게 된 '이상행위' 는 더 이상 '이상행위' 가 아니다. [3]

2-2. 비지도 학습

머신러닝은 크게 지도학습과 비지도 학습으로 나뉜다. 지도학습(Supervised Learning)이란 x값에

대한 y값이 존재하는 데이터들로부터 결과 값을 얻어내는 학습 방법이며 분류(Classification), 회기(Regression) 등의 분석을 통하여 문제를 해결 할 수 있다. 비지도 학습(Unsupervised learning)의 경우, x값에 대한 y값이 존재하지 않는다. 따라서 답이 정해져 있지 않은 문제를 해결 할 때 주로 쓰이며, 비지도 학습의 대표적인 문제 해결 방식으로는 군집화(Clustering) 분석을 예로 들 수 있다.

2-3. 군집화 분석

군집화 분석이란 주어진 데이터들의 특성을 고려해 데이터 집단(Cluster)을 정의하고 데이터 집단의 대표할 수 있는 대표점을 찾는 방법이며, 군집 형성은 비슷한 특성을 가진 데이터들의 집단이라 말할 수 있으며 반대로 데이터들의 특성이 다르다면 다른 클러스터에 속해야 한다.[4]

2-4. 다변량 가우스 분포

비지도 학습에 속하며 군집화를 통하여 이상탐지를 할 수 있는 방법 중의 하나이며, 기존의 가우스 분포에 비해 다차원적으로 분석이 가능하다는 것이 특징이다.

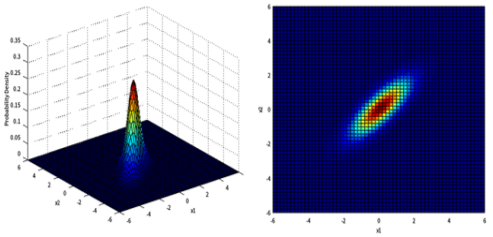

그림 1. 다변량 가우스 분포의 표현

다변량 가우스 분포에서는 평균은 μ(Mu)로, 표준편차의 제곱 형태인 분산은 Σ(Sigma)로 표현하며 다음과 같이 표현 할 수 있다.

$$p(x;\mu=\begin{bmatrix}\mu_a\\ \mu_b\end{bmatrix},\Sigma=\begin{bmatrix}\Sigma_{aa}\ \Sigma_{ab}\\ \Sigma_{ba}\ \Sigma_{bb}\end{bmatrix})$$

또한, μ(Mu) 값과 Σ(Sigma)은 아래의 공식을 통해 도출해 낼 수 있으며

$$\mu=\frac{1}{m}\sum_{i=1}^{m}x^{(i)},\Sigma=\frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu)(x^{(i)}-\mu)^T$$

이를 아래의 식에 대입하게 되면 p(x)의 대한 함수를 구할 수 있다.

$$p(x)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

이때, 이상탐지를 위하여 정해놓은 ε(Epsilon)에 대해 $p(x)<\varepsilon$ 을 만족하면, 비정상적 또는 이상행위로 간주할 수 있다.

2-5. LPWA(Low Power Wide Area)

사물 인터넷(IoT)은 다양한 목적으로 사용이 되며 기기를 운용하기 위한 전력을 공급함에 있어서 큰 제약이 있을 수 있다. 쓰레기의 양을 측정해주는 센서를 그 예로 들 수 있으며, 위의 경우에 저전력 장거리 통신의 특징을 가진 LPWA 기술을 사용하여 비교적 적은 대역폭으로 배터리를 보다 효율적으로 사용할 수 있다.

표 1. LPWA 기술 종류[5]

구분	LoRa	Sigfox	LTE MTC	LTE NB-IoT
주파수 대역	비 면허대역 (920MHz)	비 면허 대역 (920MHz)	면허 대역 (LTE 대역)	면허 대역 (In-band, Guard-band)
표준화 단체	LoRa Alliance	ETSI	3GPP	3GPP
표준화 단계	표준 완료	표준 완료	Cat. 0/1: 표준 완료 Cat. M: 표준화 진행 중	표준 진행 중 (Rel.13) : 9 월 예상
Max. Data Rate	5.47kbps	1kbps	Cat.1: DL/UL 10/5Mbps Cat.0: DL/UL 1Mbps Cat.M: DL/UL 0.2Mbps	200 kbps
Cell Coverage	~10 km	~10 km	~10 km	~5 km
상용화	기 상용화	기 상용화	기 상용화(Cat.1)	'17 년 상반기 예상
Device Stack	Non-IP	Non-IP	IP	Non-IP, IP
Module Chip 가격	약 5$~10$ 수준	약 5$~10$ 수준	약 20$ 수준	약 10$ 수준

LPWA 내의 LoRa(Long Range) 망 같은 경우, Node들이 IP 없이, Gateway와 Physical Layer의 통신을 하며, Gateway는 Network Server와 LoRa MAC Layer 프로토콜로 연동이 된다.[6]

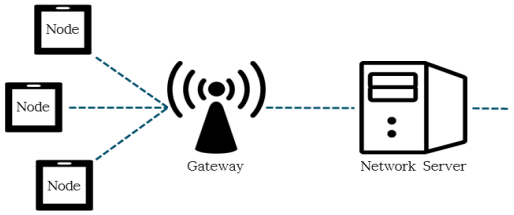

그림 2. LoRa 망 구성도

Ⅲ. 제안 기법

LPWA 네트워크 환경에서 사용되는 기기들은 기능에 따라 제한적으로 통신을 주고받는다. 따라서 데이터의 기능별로 레이블(Label)을 달아 구분을 지어 줌으로써, 해당 x값들이 불완전한 프로세스의 분포, 예를 들어, 쌍봉우리 형 등과 같은 형태를 띄지 않게 만들어 주어야 한다. 즉, 첫 번째로 가우스 분포를 이용하기 위한 조건의 만족과 이를 효율적으로 분석하기 위한 환경을 만들어 주고자 한다.

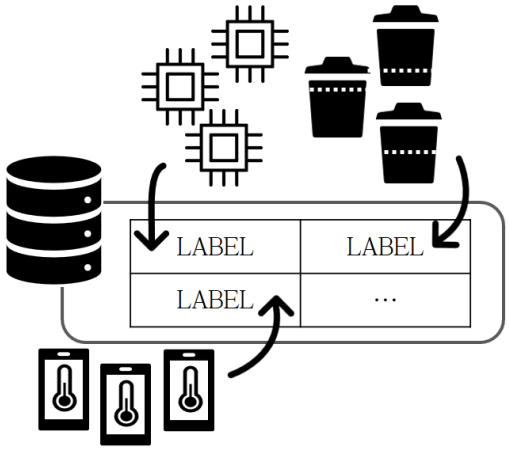

그림 3. 기능별 분류를 통한 DB에 저장

두 번째로, 기능별로 분류한 데이터들의 해당 x값들을 다변량 가우스 분포를 이용한 분석을 통하여 이상행위를 탐지 하는 방법을 제시하고자 한다.

Ⅳ. 결 론

LPWA 네트워크 환경에서 이상행위를 탐지하기 위해 기능별 레이블을 부여하여, 다변량 가우스 분포 알고리즘을 사용하기 위한 조건을 만족시키는 것과 동시에 분석 시, 데이터의 량을 감소시킴으로써, 속도의 증대를 가져올 것이라 예상된다. 향후 연구에서는 위에서 제시한 내용을 바탕으로 효율성 및 성능 검증을 하겠다.

참고문헌

[1] Howon Kim, Dongkyue Kim, Iot 기술과 보안, 情報保護學會誌, 第22卷 第1號, 2012. 2
[2] T.-J. Park, K.S. Lee, W.-C. Jeong, B.-C. Choi, H.-C. Bang, LPWA IoR Network Technology Trends, 2017 Electronics and Telecommunications Trends, 32권 1호 (통권 163), 2017. 02. 01.
[3]Josh Wills, Sean Owenm Uri Laserson, Sandy Ryza, 9가지 사례로 익히는 고급 스파크 분석, 한빛미디어, 2016.07.01
[4]위키백과, https://ko.wikipedia.org/wiki/%ED%81%B4%EB%9F%AC%EC%8A%A4%ED%84%B0_%EB%B6%84%EC%84%9D
[5][6] 고득녕, OSIA S&TR Journal, Vol.29, No.3, September 2016

ACKNOWLEDGEMENT

이 논문은 2017년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2017-0-00279,안전한 IoT 전용망 구축을 위한 LPWAN 침해 방지 기술 개발)