https://doi.org/10.33778/kcsa.2022.22.4.017

MQTT 기반 IoT 환경에서의 PCA와 LightGBM을 이용한 공격 탐지 및 분류 방안

이 지 구*, 이 수 진**, 김 영 원*

요 약

최근 머신러닝 기반의 사이버 공격 탐지 및 분류 연구가 활발히 이루어지고 있으며, 높은 수준의 탐지 정확도를 달성하고 있다. 그러나 저 사양 IoT 기기, 대규모의 네트워크 트래픽 등은 IoT 환경에서 머신러닝 기반의 탐지모델 적용을 어렵게 하고 있다. 따라서 본 논문에서는 국방분야에서도 활용되고 있는 MQTT(Message Queuing Telementry Transport) IoT 프로토콜 환경에서 수집된 데이터세트를 대상으로, 차원축소 기법인 PCA(Principal Component Analysis)와 LightGBM(Light Gradient Boosting Model)을 이용하여 IoT 공격을 효율적으로 탐지 및 분류하는 방안을 제안하였다. 실험을 통해 제안하는 분류모델의 성능을 확인한 결과 원본 데이터세트를 약 15%로 축소하였음에도 원본 전체를 모두 사용한 모델과 거의 유사한 성능을 나타냈으며, 본 논문에서 선정한 4가지 차원축소기법과의 비교 평가에서도 가장 우수한 성능을 나타냈다.

Attack Detection and Classification Method Using PCA and LightGBM in MQTT-based IoT Environment

Lee Ji Gu*, Lee Soo Jin**, Kim Young Won***

ABSTRACT

Recently, machine learning-based cyber attack detection and classification research has been actively conducted, achieving a high level of detection accuracy. However, low-spec IoT devices and large-scale network traffic make it difficult to apply machine learning-based detection models in IoT environment. Therefore, In this paper, we propose an efficient IoT attack detection and classification method through PCA(Principal Component Analysis) and LightGBM(Light Gradient Boosting Model) using datasets collected in a MQTT(Message Queuing Telementry Transport) IoT protocol environment that is also used in the defense field. As a result of the experiment, even though the original dataset was reduced to about 15%, the performance was almost similar to that of the original. It also showed the best performance in comparative evaluation with the four dimensional reduction techniques selected in this paper.

Key words : MQTTset, PCA, LightGBM, IoT attack detection

접수일(2022년 09월 21일), 게재확정일(2022년 10월 25일)

* 국방대학교 국방과학학과(주저자)
** 국방대학교 국방과학학과(공동저자)
*** 국방대학교 국방과학학과(교신저자)

1. 서론

2012년부터 가트너(Gartner)가 발표하는 10대 전략 기술에 계속해서 선정되고 있는 IoT(Internet of Things) 기술은 4차 산업혁명 시대의 핵심 기술 중 하나로서, 사람과 사물, 사물과 사물을 연결하는 초연결사회의 구현을 가능하게 해 준다[1].

글로벌 리서치 회사인 Markets and Markets는 2019년 약 1,339억 달러로 추정된 시장 규모가 2024년까지 약 2,789억 달러가 성장할 것으로 예측했고, 2025년까지 IoT 연결 기기의 수가 250억 개에 이를 것으로 전망했다[2]. 이처럼 IoT 기술은 유용성과 확장성을 바탕으로 실생활, 산업, 의료 등 제반 분야에서 인간의 편의와 새로운 가치를 창출하고 있다.

이러한 IoT 기술은 국방 분야에서도 핵심 전략기술로 부상하고 있다. 미군은 IoT 기술을 군사/전장 IoT(Internet of Military/Battlefield Things) 기술로 확장하여 함정, 항공기, 무인 항공기, 병력 및 작전기지를 연결해 빠른 상황인식, 위협평가 및 전투를 가능하게 하는 IoT 기반의 플랫폼을 개발하고 있다[3]. 한국군 또한 국방의 효율화와 미래전을 대비하기 위해 병력, 시설, 무기체계, 지휘통신체계(C4I)를 연계하여 국방자원 및 전장 관리를 위한 국방 IoT 통합 플랫폼 개발에 힘쓰고 있다[4].

한편 이러한 IoT 기술 사용범위의 폭발적 확장과 이기종 기기의 대량 연결은 방대한 네트워크 트래픽을 생성하고 있으며, 다양한 보안 취약점의 노출로 이어지면서 전통적 방식의 보안 솔루션 적용을 더욱 어렵게 하고 있다[5]. 이에 따라 최근에는 IoT 환경에서 머신러닝 기반의 침입탐지 모델 연구가 활발하게 진행되고 있으며, 비교적 높은 수준의 탐지 정확도를 달성하고 있다. 그러나 IoT 환경에서 발생하는 대규모 데이터와 저사양 IoT 기기로 인해 머신러닝 기반 모델은 시간복잡도(time complexity)와 공간복잡도(space complexity) 측면에서 한계에 직면하며 탐지 성능이 저하되는 문제도 발생한다.

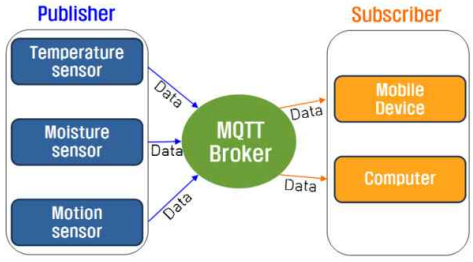
이에 본 논문에서는 이러한 문제점을 극복하기 위해 특성추출기법인 PCA (Principal Component Analysis)와 LightGBM(Light Gradient Boosting Model)을 이용하여 경량화된 탐지 모델을 생성하고 효율적으로 IoT 공격을 탐지하는 방안을 제안한다. 탐지 모델 구축 및 실험에는 국방정보사업을 통해 공급된 IoT 플랫폼에서 적용하고 있는 표준 IoT 프로토콜인 MQTT(Message Queuing Telementry Transport) 환경에서 수집한 데이터세트(MQTTset)를 이용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 선행연구를 정리하고, 3장에서는 PCA와 LightGBM 적용 방안에 대해 설명한다. 4장에서는 제안하는 방안과 다양한 데이터 차원축소 기법과의 성능평가 결과를 비교 및 분석하며, 마지막으로 5장에서 연구결과를 요약하고 결론을 맺는다.

2. 관련연구

1999년 IBM에 의해 개발되었고, 2013년 개방형 표준을 담당하는 비영리 단체인 OASIS에 의해 IoT 표준 프로토콜로 채택된 MQTT는 제한된 컴퓨팅 능력을 가지는 네트워크 환경에 최적화된 경량 메시지 전송 프로토콜로서 TCP/IP상에서 실행된다.

MQTT의 동작원리는 (그림 1)에서 보는 바와 같이 게시자(Publisher)가 브로커(broker)를 통해 메시지를 게시(publish)하면, 구독자(subscriber)가 브로커를 통해 구독(subscribe)하는 게시/구독 방식으로 이루어지며, 다중 클라이언트간 메시지 전달이 가능하다[6].



(그림 1) MQTT의 주요 구성요소 및 동작원리

MQTT는 경량 메시지 포맷을 사용하며, 80-100kb의 메모리를 사용하여 구현이 가능하다. 또한 QoS(Quality of Service)와 이벤트 방식을

통해 다수의 사용자를 지원하는 확장성을 가진다. 이러한 경량화 및 확장성 등의 장점을 바탕으로 MQTT는 전 세계 IoT 시장에서 50% 이상의 점유율을 차지하고 있으며, HTTP 프로토콜에 이어 두 번째로 많이 사용되고 있다[7].

한편 IoT 환경에서의 머신러닝 기반 공격 탐지 및 분류를 위한 연구에서 사용된 대부분의 데이터세트는 일반적인 네트워크 환경에서 수집된 데이터이기 때문에, 표준 IoT 프로토콜 환경에 적용하기 어렵다는 문제점을 안고 있다. 이러한 문제점을 인식하여, Ivan Vaccari 등[8]은 MQTT 프로토콜 환경에서 머신러닝 기반의 침입 탐지연구가 가능한 MQTTset 데이터세트를 제작하여 발표하였다.

Neenu Kuriakose 등[9]은 MQTTset을 대상으로 firefly 특성선택 알고리즘을 적용하여 데이터세트의 특성 33개 중 12개를 선택했다. 이후 SVM(Support Vector Machine), K-MEANS, RF(Random Forest) 등의 분류모델을 이용해 99%, 98%, 99.8%의 정확도를 달성하였다. 그러나 330,928개의 트래픽 데이터 중 2,000개만 클래스별로 균일하게 추출하여 실험에 사용했기 때문에 실험결과의 신뢰성은 높지 않다.

Maheshi B. Dissanayake[10]은 MLP(Multi Layer Perceptron)을 이용해 원본 데이터세트를 대상으로 90%의 정확도를 달성하였고, RF 기반의 특성 중요도(feature importance) 산출을 통해 총 33개의 특성 중 10개의 특성을 선택하였다. 이후 RF 분류모델로 실험한 결과 90%의 정확도를 지속 유지하는 것을 확인하였다.

Rachmadi[11] 등은 DoS(Denial of Service) 공격 탐지를 위해 MQTTset의 DoS 공격 클래스와 정상 트래픽 클래스 등 두 가지의 클래스만 추출하여 사용하였다. 분류모델은 AdaBoost를 이용하였으며, 실험 결과 정확도 95.84%, 재현율과 정밀도는 각각 93.28%와 98.29%를 달성하였다. 그러나 MQTTset의 5개 공격클래스 중 1개만 추출하여 이용하였기에 데이터세트의 전체적인 성능평가는 제한된다.

IoT 환경에서 특성 추출방식을 이용해 데이터차원을 감소시키기 위한 연구도 활발하게 진행되고 있다. 대표적으로 적용되는 기법으로는 오토인코더(auto encoder)와 PCA가 있다. J. Lee[12] 등은 IoT 환경에서 오토인코더 기반 특성 추출을 이용한 네트워크 침입탐지시스템을 연구했다. IoT Botnet 데이터세트인 danmini doorbell을 대상으로 오토인코더와 RF 분류모델을 통해 99% 이상의 높은 정확도를 달성하였다. 그러나 오토인코더는 신경망 깊이와 히든 노드 수의 증가에 따른 높은 연산복잡도를 야기하는 문제로 저사양 IoT 환경에서의 적용에는 한계점이 존재한다. 반면, 본 연구에서 사용하는 PCA는 선형 방식의 대표적인 차원축소 알고리즘으로 낮은 연산복잡도와 수행시간이 빠르다는 장점을 가진다.

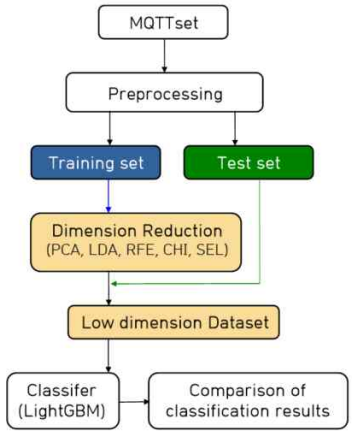
H. Kye[13] 등은 IoT 환경에서 PCA 기반의 저복잡 이상탐지 연구를 통해 기존 연구보다 낮은 연산복잡도에서 높은 탐지성능을 달성하였고, Subhash Waskle[14] 등은 KDD 데이터세트를 대상으로 PCA와 SVM, Naïve Bayes, Decision Tree 분류모델을 사용하여. 수행시간은 단축하면서도 탐지 정확도는 향상시켰다.

3. IoT 공격 탐지 및 분류 방안

3.1 제안절차

MQTTset에서의 IoT 공격 탐지 및 분류 방안은 (그림 2)에서 보는 바와 같이 구성된다. 우선 MQTTset 데이터세트에 대해 전처리 및 정규화를 실시한다. 이어서 학습 및 테스트 데이터로 분할하고, PCA를 통해 학습데이터의 특성을 추출하여, 차원을 감소시킨다. 이후 학습데이터의 차원축소 기준을 테스트 데이터에도 동일하게 적용하여 테스트 데이터의 차원을 축소한다. 마지막으로 LightGBM을 이용하여 학습데이터를 학습한 후, 테스트 데이터를 대상으로 공격 클래스를 탐지 및 분류한다.

PCA의 효과를 검증하기 위해 실험에서는 특성 추출 기법인 LDA(Linear Discriminant Analysis) [15], 특성선택 기법에서는 Chi-square test[16] 및 RFE(Recursive Feature Elimination) 알고리즘, sklearn 라이브러리의 SelectFromModel 등 총 4가지의 데이터 차원축소 기법과 비교를 실시한다.



(그림 2) 제안하는 방법의 실험과정

3.2 데이터세트(MQTTset)

MQTTset은 2020년 발표된 IoT 공인 데이터세트로 MQTT 프로토콜 환경에서 수집된 데이터로 구성되어 있다. 데이터는 MQTT 프로토콜 기반 IoT 환경에서 수집된 다양한 공격시나리오와 조도계, 온도계, 습도계, 모션센서 등 총 8개의 IoT 장치에서 생성된 정상 트래픽 데이터를 포함하고 있다. 총 특성은 33개이며, 정상데이터와 5개의 공격 클래스로 구성되어 있다. 세부구성은 <표 1>과 같다.

< 표 1> MQTTset 데이터세트 세부구성

Class	Number of data
Legitimate	165,463
Flooding Denial of service	130,233
MQTT Publish Flood	14,501
SlowITe	10,924
Malformed data	9,204
Brute force authentication	613

3.3 데이터 전처리

모델 학습 전 결측치 제거와 라벨 인코딩을 통해 범주형 데이터를 수치형 데이터로 전환하였다. 그리고 모델의 성능향상을 위해 최소값이 0, 최대값이 1이 되도록 MinMaxScaler를 이용하여 스케일링을 진행하였다.

학습 데이터와 테스트 데이터는 7:3 비율로 분할하였으며, 최종 데이터의 샘플 수는 <표 2>와 같다.

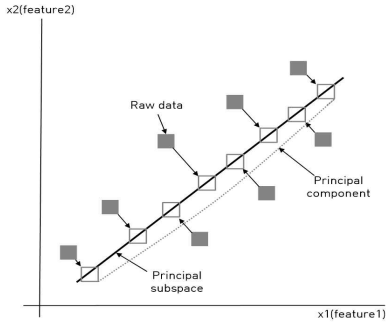
<표 2> 학습 및 테스트 데이터 구성

Class	Train	Test
Legitimate	165,463	49,639
Flooding Denial of service	130,233	39,067
MQTT Publish Flood	14,501	4,350
SlowITe	10,924	3,277
Malformed data	9,204	2,761
Brute force authentication	613	184

3.4 PCA

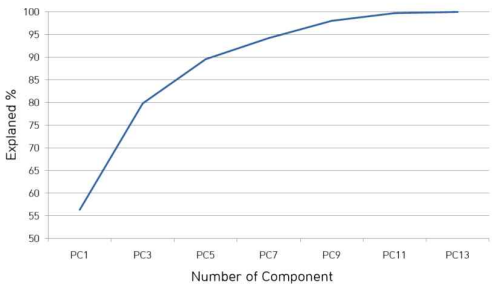
PCA는 특성추출(feature extraction)의 한 방법으로 주된 목적은 데이터의 축소, 즉 데이터의 특성 수를 감소시키는 것에 있다. PCA의 기본적인 동작원리는 주어진 레이블을 고려하지 않고, 데이터의 분산도(variance)를 잘 나타내는 주성분(principal component, 이하 PC)을 분석하여, 중요 정보만을 담은 새로운 부분 공간(principal subspace)으로의 선형변환을 통해 데이터의 차원을 축소하는 것이다.

(그림 3)은 차원감소의 원리를 보여주며, 수집된 데이터 샘플에 대한 2개의 특성값을 보여준다. 2차원으로 데이터를 표현하였으며, 원본 데이터(raw data)를 가장 많이 표현 가능한 부분 공간으로 선형변환을 하면 원본 데이터에 포함되어 있던 정보를 대부분 유지하면서도 2개의 차원을 1개의 차원으로 축소할 수 있는 PC를 생성할 수 있다.



(그림 3) PCA를 이용한 데이터 차원감소 과정

본 논문에서는 위와 같은 PCA의 원리를 이용해 PC를 추출하기 위해 sklearn 라이브러리의 PCA 모듈을 사용하였다. PC 추출결과 데이터의 정보를 의미하는 분산도가 PC 1개에서는 56%, 5개에서는 90%로 산출되었으며, PC 10개부터는 99% 이상의 분산도가 산출되었다. 이를 통해 PCA는 차원을 대폭 축소하더라도 데이터의 정보를 최대한 보존한다는 것을 알 수 있다. 실험에서는 모델의 성능이 가장 우수한 경우, 그리고 PC 개수와 성능을 지속 유지할 수 있는 PC 개수를 선정하기 위해 분류모델을 통해 성능을 평가하였다. PC 개수에 따른 분산도, 즉 데이터의 누적 설명률은 (그림 4)와 같다.



(그림 4) PC 개수에 따른 데이터의 누적 분산도

3.5 분류모델(LightGBM)

LightGBM은 트리 기반의 분류모델로 기존 트리 기반 알고리즘과는 다르게 트리 균형을 지속적으로 맞추지 않고, 최대 손실 값을 가진 리프 노드를 지속 분할 하는 리프 중심 트리분할(leaf Wise) 방식을 사용한다. 이에 따라 Gradient Boosting 모델과 비교해서 동일한 성능을 유지하면서도 학습속도는 20배까지 향상되었다[17]. 또한, 메모리 사용량이 적고, 데이터 세트 크기가 클 때 뛰어난 성능을 발휘한다.

LightGBM의 모델 성능을 평가하기 위해 PCA를 적용하지 않은 원본 데이터세트를 대상으로 XGboost, RF, KNN 알고리즘 등 3가지 분류모델과 결과를 비교하였다. 실험결과 <표 3>과 같이 LightGBM이 비교 모델 중 모든 평가지표에서 가장 우수한 성능을 나타내었으며, 수행시간 또한 가장 적게 소요되었다.

<표 3> 원본 데이터세트 대상 모델 성능 평가결과

구 분	XGB	LGBM	RF	KNN
Accuracy	0.9339	0.9398	0.9384	0.9232
Precision	0.9311	0.9414	0.9268	0.9219
Recall	0.9398	0.9398	0.9268	0.9115
F1-score	0.9321	0.9385	0.9372	0.9222
Train time	510sec	112sec	152sec	503sec
Test time	264sec	33sec	43sec	201sec

이러한 결과를 바탕으로 본 논문에서는 PCA를 통해 추출된 PC를 이용하여 LightGBM에 학습시키고, 분류성능을 평가하였다.

LightGBM 분류는 sklearn 라이브러리를 사용하였으며, 하이퍼파라미터는 n_estimator = 100, random_state = 1로 설정하였다. n_estimator는 결정 트리의 개수를 의미하며, random_state는 매 실험 시 동일 결과를 도출할 수 있도록 고정하였다.

4. 실험 및 평가

4.1 실험환경

실험은 Windows 10 Home 64bit 운용체제, Intel(R) Core(TM) i5-8520U CPU, 8G RAM 사양의 노트북에서 Google Colaboratory를 이용하였으며, GPU 하드웨어 가속기는 사용하지 않았다.

4.2 성능평가 지표

분류모델의 성능을 평가하기 위한 지표로 Accuracy, Precision, Recall, F1-score를 사용하였다. 성능 평가지표는 식 (1), (2), (3), (4)와 같이 산출된다.

$$Accuracy = \frac{(TP+TN)}{(TP+FN+FP+TN)} \quad (1)$$

$$\mathrm{Pr}ecision = TP/(TP+FP) \quad (2)$$

$$Recall = TP/(TP+FN) \quad (3)$$

$$F1-score=2\times\frac{(\mathrm{Precision}\times Recall)}{(\mathrm{Precision}+Recall)} \quad (4)$$

4.3 실험결과 및 분석

PCA를 이용해 추출한 PC를 개수 별로 분류모델인 LightGBM에 학습시켜 결과를 확인하였으며, 실험결과는 <표 4>와 같다.

PCA는 원본 데이터의 특성 33개를 PC 14개로 감소시켰을 때 약간의 성능향상을 나타냈다. 또한, PC를 5개까지 축소하였을 때에도 모델의 성능을 거의 유지하면서 수행시간은 50% 이상 단축되었다. 그러나 PC 4개부터는 성능 하락이 뚜렷하게 나타남에 따라 가장 효율적인 PC 개수는 5개로 판단하였다.

<표 4> PCA-LightGBM 성능평가 결과

구 분	특성 33개	PC 14	PC 5
Accuracy	0.9398	0.9399	0.9392
Precision	0.9414	0.9423	0.9408
Recall	0.9398	0.9399	0.9392
F1-score	0.9385	0.9386	0.9378
Train time	112sec	65sec	45sec
Test time	33sec	18sec	12sec

원본의 특성 33개와 PC 5개를 적용한 분류모델의 오차행렬(confusion matrix)은 <그림 5>, (그림 6)에서 보는 바와 같다.

(그림 5) 원본 특성 33개-LightGBM 오차행렬

	legitimate	dos	malformed	slowite	flood	bruteforce
legitimate	49003	636	0	0	0	0
dos	2918	35776	0	42	331	0
malformed	0	0	2761	0	0	0
slowite	27	269	0	1871	1110	0
flood	18	414	0	116	3802	0
bruteforce	71	22	0	2	3	86

(그림 6) PC 5개-LightGBM 오차행렬

	legitimate	dos	malformed	slowite	flood	bruteforce
legitimate	48998	636	0	4	0	1
dos	2918	35779	0	34	336	0
malformed	0	0	2761	0	0	0
slowite	35	313	0	1873	1054	0
flood	23	408	0	175	3744	0
bruteforce	81	22	0	4	2	75

PCA의 성능을 추가적으로 검증하기 위해 특성추출 기법에서는 LDA, 특성선택 기법에서는 RFE, Chi-square test 및 SelectFromModel을 사용하여 분류모델의 성능이 가장 높게 평가된 특성을 추출 또는 선택하였다. 그리고 PCA를 통해 산출된 PC 5개의 성능평가 결과와 비교하기 위해 비교 대상 모델도 5개 특성까지 단계적으로 축소해 가며 성능변화 추이를 확인하였다. 실험결과는 <표 5>과 같다.

<표 5> 모델별 성능평가 결과

구 분	특성추출		특성선택		
	PCA	LDA	RFE	CHI	SEL
특성 수	14개	5개	12개	20개	11개
Accuracy	0.9399	0.9382	0.9398	0.9398	0.9397
Precision	0.9423	0.9395	0.9419	0.9419	0.9417
Recall	0.9399	0.9382	0.9398	0.9398	0.9397
F1-score	0.9386	0.9369	0.9386	0.9385	0.9386
특성 수	5개				
Accuracy	0.9392	0.9382	0.9237	0.8245	0.8285
Precision	0.9408	0.9395	0.9228	0.8594	0.8622
Recall	0.9392	0.9382	0.9237	0.8245	0.8285
F1-score	0.9378	0.9369	0.9196	0.8068	0.8132

모델별 가장 높은 탐지 성능을 기록한 특성 수는 PCA 14개, RFE 및 SelectFromModel은 각각 12개와 11개에서 가장 높은 결과가 나타나 PCA의 특별한 장점은 확인하지 못했다. 그러나 특성 수를 더욱 감소시키면서 성능평가를 진행한 결과 PCA는 5개까지 탐지 성능을 거의 유지하였으나,

RFE와 SelectFromModel은 9개, Chi-squre test는 19개부터 뚜렷한 성능 하락이 나타났다. LDA는 알고리즘 특성상 클래스 개수보다 1개가 적은 5개부터 특성 추출이 가능하며, PCA 다음으로 좋은 성능을 나타냈다.

이러한 실험결과는 PCA가 특성선택 방식의 모델과 비교해 데이터의 차원을 2배 더 축소할 수 있으며, 이를 통해 연산복잡도와 수행시간을 단축하여 탐지모델의 효율성 향상에 기여할 수 있음을 나타낸다.

5. 결 론

본 논문에서는 국방 분야에서도 사용되고 있는 MQTT 프로토콜을 기반으로 하는 IoT 환경에서 수집된 데이터세트를 이용하여 PCA와 LightGBM 기반의 효율적인 IoT 공격 탐지 및 분류 기법을 제안하였다. 제안하는 기법은 대량의 데이터와 저 사양 IoT 환경에 적합하도록 데이터 차원감소를 통해 보다 신속하게 IoT 공격을 탐지 및 분류하면서도 성능 감소는 발생하지 않도록 하는데 중점을 두고 설계되었다.

제안한 방안을 검증하기 위해 MQTTset을 대상으로 PCA를 적용하여 데이터세트의 새로운 특성인 PC를 추출 후 LightGBM 분류모델을 통해 성능을 평가하였다. 실험결과 원본 데이터세트를 약 15%로 축소한 PC 5개에서 원본 데이터의 특성 33개를 모두 사용한 결과와 유사한 성능을 달성하였다. 또한, PCA의 효과성을 검증하기 위해 LDA, RFE, Chi-square test 및 SelectFromModel 등 다양한 차원축소 기법들과 성능을 비교한 결과 제안한 방안이 가장 우수한 성능을 나타냄을 확인하였다. 이는 동일한 데이터세트로 연구를 진행했던 기존연구[10]와 비교했을 때 더 적은 데이터 차원에서 동일 성능을 달성한 것으로서, 제안기법의 유효성을 입증했다고 볼 수 있다.

본 논문에서 제안한 방안은 국방 IoT 환경에서 사용되고 있는 표준 프로토콜 환경에서 수집된 데이터세트를 이용했다는 점, 그리고 컴퓨팅 능력이 제한되는 IoT 환경에 적합한 경량화된 공격 탐지모델을 도출했다는 점에서 향후 국방 IoT 플랫폼의 보안 강화에 큰 도움이 될 것으로 판단한다.

향후 연구에서는 일부 클래스의 오 분류 비율이 높게 나타난 원인을 분석하고, 더욱 고차원의 IoT 데이터세트를 대상으로 제안방안의 효과를 추가로 검증할 계획이다.

참고문헌

[1] Young-Teak Oh, In-June Jo, "Data Modeling for Cyber Security of IoT in Artificial Intelligence Technology", International JOURNAL OF CONTENTS, Vol. 21, No. 12, pp 58-65, 2021.

[2] Markts and Markets, "IoT Solutions and Service Market", https://www.marketsandmarkets.com/Market-Reports, TC7719, 2022.

[3] IEEE Computer Society, "Internet of Things Meets the Military and Battlefield", https:// www.computer.org/publications/tech-news/research/internet-of-military-battlefield-things-iomt-iobt, 2022.

[4] Korea Defense Industry Accociation, “Development of defense IoT platform and solution for future intelligent resource management and battlefield management system”, Defense & Technology, Vol 470, No. 28, pp 28-29, 2018.

[5] Ahmad, Rasheed, and Izzat Alsmadi. "Machine learning approaches to IoT security: A systematic literature review." Internet of Things Vol. 14, 2021.

[6] Y. Jang, J. Shim, and S. Park, “Analysis Standardized of IoT-based Low-power·Light -weight Protocol,” Journal of the Korea Institute of Information and Communication Engineering, vol. 20, no. 10, pp. 1895-1902, Oct. 2016.

[7] I. Skerritt, "IoT Developer Survey 2016," Eclipse IoT Work. Gruop, IEEE IoT Agil. IoT, 2016.

[8] Ivan Vaccari, Giovanni Chiola, Maurizio Aiello, Maurizio Monelli, Enrico Cambiaso, "MQTTset, a New Dataset for Machine Learning Techniques

on MQTT", Sensors, Vol. 20, 2020.

[9] Kuriakose, Neenu, and Uma Devi. "MQTT Attack Detection Using AI and ML Algorithm." Pervasive Computing and Social Networking. Springer, Singapore, Vol. 317, 2022. 13-22.

[10] Dissanayake, Maheshi B. "Feature Engineering for Cyber-attack detection in Internet of Things.", I.J Wireless and Microwave Technologies, Vol. 6, pp 46-54, 2021.

[11] Rachmadi, Salman, Satria Mandala, and Dita Oktaria. "Detection of DoS Attack using AdaBoost Algorithm on IoT System", 2021 International Conference on Data Science and Its Applications (ICoDSA). IEEE, 2021.

[12] Lee, Joohwa, and Keehyun Park. "Network Intrusion Detection System Using Feature Extraction Based on AutoEncoder in IOT environment." KTSDE, Vol. 8, No. 12, pp 483-490, 2019.

[13] Hyoseon Kyew, Minhae Kwon, "PCA-Based Low-Complexity Anomaly", KCIS, Vol. 46, No. 6, pp 941-955, 2021.

[14] Waskle, Subhash, Lokesh Parashar, and Upendra Singh. "Intrusion detection system using PCA with random forest approach." 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, pp. 803-808, 2020.

[15] Martinez, Aleix M., and Avinash C. Kak. "Pca versus lda." IEEE transactions on pattern analysis and machine intelligence Vol. 23, No. 2, pp 228-233, 2001.

[16] Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D. and Saeed, J. "A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction.", Journal of Applied Science and Technology Trends Vol. 1, No. 2, pp. 56-70, 2020.

[17] Ke, Guolin, et al. "Lightgbm: A highly efficient gradient boosting decision tree." Advances in neural information processing systems 30, 2017.

〔저자소개〕

이 지 구 (Ji-Gu Lee)
2011년 2월 목포해양대학교
조선해양공학 학사
2021년 3월 ~ 현재 국방대학교
국방과학학과 석사과정
email : jglee0120@gmail.com

이 수 진 (Soo-Jin Lee)
1992년 3월 육군사관학교
전산학과 학사
1996년 2월 연세대학교
컴퓨터과학과 석사
2006년 2월 한국과학기술원
전산학과 박사
2006년 ~ 현재 국방대학교
국방과학학과 교수
email : cyberkma@korea.kr

김 영 원 (Young-won Kim)
2009년 3월 해군사관학교
전산학과 학사
2021년 2월 국방대학교
국방과학학과 석사
2021년 2월 ~ 현재 국방대학교
국방과학학과 박사
email : headsun21@gmail.com