Journal of KIIT. Vol. 21, No. 5, pp. 131-137, May 31, 2023. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2023.21.5.131

트랜잭션 분석 기반 블록체인 네트워크 이상 탐지

고 광 만*

Transaction Analysis based Blockchain Network Anomaly Detection

Kwang-Man Ko*

이 논문은 2019년도 상지대학교 교내 연구비 지원에 의한 것임. 이 논문은 부분적으로 2022년도 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. RS-2022-00207391, Development of Hashgraph-based Blockchain Enhancement Scheme and Implementation of Testbed for Autonomous Driving)

요 약

최근 블록체인의 가치가 상승함에 따라 다양한 보안 위협이 등장하고 있다. 또한 블록체인 네트워크에 비정상적인 행위를 빠른 시간내에 탐지하는 연구가 중요한 상황이다. 본 논문에서는 블록체인 네트워크 보안 위협과 이상 탐지 요소를 탐지하기 위해 트랜잭션 정보를 기반으로 데이터를 수집하고 비지도 학습을 활용하여 이상 탐지에 대한 실험 결과를 제시한다. 본 논문의 학습 모델은 하나의 블록당 탐지 시간이 0.058ms로서 초당 17,000여개의 블록의 정상/비정상 행위 여부를 판단이 가능하다. 또한 학습 평가 점수는 Accuracy 98.2%, False Negative 0.6%, False Positive 1.2%이다. 비정상 행위로 발생한 블록 100개 중 99.4개의 블록을 비정상 행위로 올바르게 탐지하고 정상 행위로 발생한 블록 약 100개 중 99.8개의 블록을 정상 행위로 올바르게 탐지한다. 종합적으로 100개의 블록이 생성된 경우, 98개 블록의 정상/비정상 행위 여부를 올바르게 탐지하였다.

Abstract

Recently, as the value of blockchain increases, various security threats are emerging. In addition, it is important to research on detecting anomaly behavior in the blockchain network in a short time. In this paper, in order to detect blockchain network security threats and anomaly detection elements, we collect data based on transaction information and present experimental results for anomaly detection using unsupervised learning. The learning model of this paper has a detection time of 0.058 ms per block, and it is possible to determine normal/abnormal behavior of about 17,000 blocks per second. Also, the learning evaluation scores were Accuracy 98.2%, False Negative 0.6%, False Positive 1.2%. It correctly detects 99.4 blocks out of 100 abnormal behavior blocks as abnormal behavior and correctly detects 99.8 blocks out of about 100 normal behavior blocks as normal behavior. Overall, when 100 blocks were generated, normal/abnormal behavior of 98 blocks was correctly detected.

Keywords

blockchain network, machine learning, anomaly detection, cyber security

* 상지대학교 컴퓨터공학과 교수
- ORCID: https://orcid.org-0000-0002-7465-5400

· Received: May 09, 2023, Revised: May 18, 2023, Accepted: May 21, 2023
· Corresponding Author: Kwang-Man Ko
Dept. of Computer Engineering, Sangji University
Tel.: +82-33-730-0486, Email: kkman@sangji.ac.kr

Ⅰ. 서 론

블록체인은 네트워크 내의 모든 참여자가 공동으로 트랜잭션 정보를 검증, 기록, 보관함으로써 중앙 집중형 서버 없이도 트랜잭션에 대한 무결성, 투명성 및 신뢰성을 보장하는 기술로서 다양한 분야에서 활용되고 있다[1].

블록체인의 가치가 상승함에 따라 블록체인 또한 해커의 주요 공격 대상이 되고 있다. 블록체인 기술의 처리 속도 한계와 트랜잭션 정보 관리에 대한 부담을 악용한 가용성 저하 공격, 트랜잭션 유효성 검사 시간을 지연시키는 DDoS 공격, 블록체인 익명성을 악용한 비정상 트랜잭션 등 다양한 보안 위협이 증가함에 따라, 블록체인 보안 사고에 의한 경제적 손실이 매년 증가하고 있어 기술 발전 및 도입을 방해하고 있다[2].

최근에는 국내외에서 블록체인 2.0 기술이 공공·민간 부분에 적극 도입되고 있으며 특히 금융, 물류, 의료, 분산인증 등 전 산업 분야에서 활용될 가능성이 높아지고 있어서 블록체인 네트워크 보안 위협을 신속하고 정확하게 탐지하는 기술에 대한 연구가 필요한 상황이다.

본 논문에서는 블록체인 네트워크 보안 위협과 이상 탐지 요소를 탐지하기 위해 트랜잭션 정보를 기반으로 데이터를 수집하고 비지도 학습을 활용하여 이상 탐지에 대한 실험 결과를 제시하였다.

본 논문에서 구성한 시뮬레이터를 통해 생성한 학습 데이터를 바탕으로 학습된 모델 한 블록당 탐지 시간이 0.058ms로, 초당 17,000여개의 블록의 정상/비정상 행위 여부를 판단이 가능했다. 또한 학습 평가 점수는 accuracy 98.2%, false negative 0.6%, false positive 1.2%으로, 비정상 행위로 발생한 블록 100개 중 99.4개의 블록을 비정상 행위로 올바르게 탐지하고, 정상 행위로 발생한 블록 약 100개 중 99.8개의 블록을 정상 행위로 인해 생성된 블록으로 올바르게 탐지해낼 수 있었다. 종합적으로 100개의 블록이 생성된 경우, 98개 블록의 정상/비정상 행위 여부를 올바르게 탐지하였다.

본 논문의 구성은 다음과 같다. 2장은 블록체인 네트워크 보안 위협 및 이상 탐지에 대한 관련 연구를 소개한다. 3장에서는 본 논문에서 제안하고 시도하는 블록체인 네트워크 이상 탐지 모델 설계와 개발 환경을 소개한다. 4장에서는 시뮬레이션 환경을 설정하고 실험을 통해 도출한 결과와 분석 내용을 소개한다. 마지막으로 5장에서는 결론에 대해 기술하고 본 연구가 갖는 한계점 및 향후 연구에 대하여 기술한다.

Ⅱ. 연구배경 및 관련연구

2.1 블록체인 네트워크 보안 위협

블록체인 네트워크의 가용성 저하는 노드가 급증하고 트랜잭션이 증가함에 따라 처리 속도와 거래 정보 관리 부담이 증가하여 발생한다. 노드의 증가 및 트랜잭션의 증가에 따른 가용성 저하를 개선하기 위해 블록체인 네트워크 노드 인증, 트랜잭션에 대한 수수료 발생 등을 활용하고 있다.

이중 지불은 합의된 트랜잭션에 대한 블록이 생성되기 전에 보상을 받은 후 정상적인 트랜잭션을 취소 또는 재사용하는 방식으로 블록체인 네트워크 보의 보안성을 위한다.

블록체인 네트워크에서 51% 이상의 해시파워를 확보하여 거래내역 조작 또는 교체하는 공격으로 해시 파워 분배 조정, 새로운 합의 알고리즘 개발을 통해 보안 위협에 대처하고 있다[3].

DDoS 공격은 불필요한 거래 또는 불량 거래를 의도적으로 발생시켜 블록체인 네트워크 성능 저하시키는 보안 위협으로 비트코인, 이더리움 등에서 공격의 증후가 확인되었다[4].

2.2 블록체인 네트워크 보안 위협 탐지

블록체인 네트워크의 이상 탐지는 이상 탐지 데이터 수집·배포, 탐지·대응, 관리 단계를 진행된다[2][6].

수집·배포 단계에서는 블록체인 네트워크를 위협하는 정보를 수집하고 네트워크 참여 노드들에게 정보를 공유 또는 배포한다. 보안 위협 탐지·대응 단계에서는 수집한 노드 또는 트랜잭션 정보를 바탕으로 블록체인 네트워크 위협 상황을 판단하는 작업을 수행한다.

탐지 단계에서 네트워크 과부하, 위협 상황을 탐지하기 위해 네트워크의 성능을 정량적으로 평가할 수 있는 초당 트랜잭션 수, 평균 응답 지연, CPU당 트랜잭션 수, 메모리당 트랜잭션 수, 디스크 입력/출력당 트랜잭션 수, 네트워크 데이터 당 트랜잭션 수를 이용한다. 거래 프로세스에 따른 평가 지표로 피어 검색 속도, RPC 응답 속도, 트랜잭션 전파 속도, 계약 실행 시간, 상태 업데이트 시간, 합의 비용 시간과 같은 지표를 설정하고 이용한다[7].

최근에는 인공지능 기술을 기반으로 블록체인 네트워크에서 이루어지는 악성 행위, 보안 위협 패턴을 학습하고 보안 위협 행위를 탐지하는 연구가 진행되었다. [8] 연구에서는 머신러닝 기반으로 이더리움 악성 계정을 탐지하기 위해 공개 데이터를 활용하여 EtherScamDB에 이더리움 악성 계정, 정상 계정을 수집한 후 학습 데이터의 특성으로 사용하기 위하여 트랜잭션에서 42개 이상의 특성을 추출하는 연구를 진행하였다. [9] 연구에서는 OCGNN (One Class Graph Neural Network) 기반 악성 노드 및 악성 트랜잭션 탐지에 관한 연구를 수행하였다. 악성 트랜잭션을 판별하기 위한 독창적인 알고리즘을 개발하였지만 성능평가에 대한 우수성을 확인할 수 없다.

본 논문에서는 기본 연구 방법과 차별적으로 블록체인에 저장되는 트랜잭션에 대한 분석을 통해 이상 탐지 유형을 판별하는 특징을 가지고 있다.

III. 트랜잭션 기반 블록체인 네트워크 이상 탐지

3.1 전체 시스템 구성

블록체인 네트워크 이상 탐지를 위한 전체 시스템 구성은 그림 1과 같이 이더리움 노드에서 X.probe[10]를 통해 발생되는 블록의 정보는 MariaDB에 저장하고 시스템 상태에 대한 데이터는 InfluxDB에 저장한다. 수집된 블록 정보는 이상 탐지를 목적으로 특징값을 추출하는 데 활용하였다.

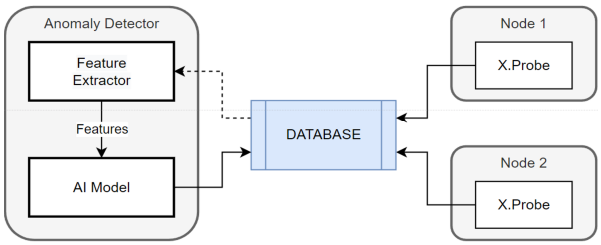

그림 1. 블록체인 네트워크 이상 탐지 시스템
Fig. 1. Blockchain network anomaly detection system

이상 탐지는 그림 2와 같이 이상 탐지 시점 기준으로 최근에 생성된 최대 100개의 블록 정보를 MariaDB에서 가져온 후, 이상 탐지에 활용할 특징값(블록 크기, 블록 내 포함된 트랜잭션 개수, 이전 블록과의 생성 간격)을 추출하는 것부터 시작된다.

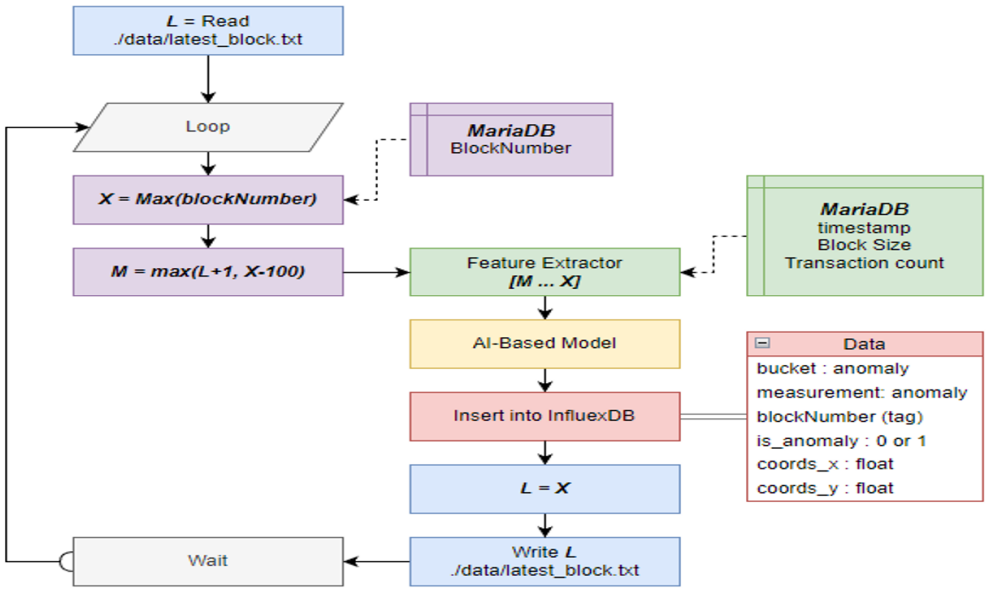

그림 2. 블록체인 네트워크 이상 탐지 과정
Fig. 2. Blockchain network anomaly detection procedure

본 논문에서는 스트리밍 데이터 환경에서 지속적인 부분 학습이 가능한 특징을 가지고 있는 Half-Space Tree[11] 이상 탐지 학습 모델을 적용하였다. Half-Space Tree 모델을 통해 정상 또는 비정상으로 판별된 이상 탐지 결과(is_anomaly), PCA를 활용하여 특징값을 2차원 그래프로 변환하여 얻은 x, y값(coords_x, coords_y)을 블록 번호화 함께 InfluxDB에 저장하는 과정으로 동작한다.

블록단위로 읽은 후에 최대 100개의 블록에 대해 특징값을 추출한 후 인공지능 기반 모델을 통해 이상 탐지된 내용은 InfluxDB에 저장하는 그 결과를 읽어 이상 탐지 과정이 진행된다.

3.2 이상 탐지 시스템 설계

본 논문에서 블록체인 네트워크의 이상 탐지 모듈 개발을 위해 표 1과 같은 라이브러리를 사용하였다.

블록체인 네트워크 이상 탐지를 위한 학습 모델은 스트리밍 데이터 환경에서 지속적으로 부분 학습이 가능하고 빠른 작동 시간이 장점인 Half-Space Tree 모델을 적용하였으며, 최근 10분 동안 생성된 블록의 평균적인 패턴의 범주를 벗어나는 블록을 비정상 트랜잭션으로 판별하였다.

또한 블록 체인 네트워크 이상 탐지 결과를 시각화를 위해 고차원 데이터를 저차원 데이터로 변환하는 차원 축소를 활용, 이 과정에서 가장 대중적이고 빠른 속도를 가진 PCA(Principal Component Analysis)[12] 모델을 활용하였다.

표 1. 개발 라이브러리
Table 1. Development libraries

Libraries	Description
pyenv 2.3.6	Python version
scikit-learn 1.1.1	Machine learning library
scikit-multiflow 0.5.3	Scikit-learn extention library
aiomysql 0.1.1	MariaDB client library
sqlalchemy 1.4.42	ORM(Object Relational Mapping) library
influxdb-client[async] 1.33.	InfluxDB client library
numpy 1.23.4	Python huge multi array library
pandas 1.4.2	Data analysis and manipulation library

Ⅳ. 이상 탐지 실험 및 분석

4.1 시뮬레이션 구성

블록체인 네트워크의 이상 탐지 시뮬레이션은 그림 4와 같은 환경에서 진행하였다.

Time Previous Block Current Block Feature Extraction Time Detection Time

```
...
2022-11-16T13:22:17.251715 :  15,015 ->  15,022 (+  7) / extract features:     4.897 ms / detection:    14.319 ms
2022-11-16T13:22:17.792188 :  15,022 ->  15,028 (+  6) / extract features:     4.917 ms / detection:    15.215 ms
2022-11-16T13:22:18.325829 :  15,028 ->  15,032 (+  4) / extract features:     4.802 ms / detection:    14.115 ms
2022-11-16T13:22:18.871347 :  15,032 ->  15,040 (+  8) / extract features:     5.223 ms / detection:    15.418 ms
2022-11-16T13:22:19.417523 :  15,040 ->  15,047 (+  7) / extract features:     5.268 ms / detection:    17.420 ms
2022-11-16T13:22:19.961139 :  15,047 ->  15,054 (+  7) / extract features:     5.595 ms / detection:    14.181 ms
2022-11-16T13:22:20.515111 :  15,054 ->  15,063 (+  9) / extract features:     4.480 ms / detection:    15.623 ms
2022-11-16T13:22:21.058983 :  15,063 ->  15,070 (+  7) / extract features:     4.725 ms / detection:    14.840 ms
2022-11-16T13:22:21.598254 :  15,070 ->  15,077 (+  7) / extract features:     4.386 ms / detection:    13.804 ms
2022-11-16T13:22:22.136261 :  15,077 ->  15,083 (+  6) / extract features:     4.364 ms / detection:    13.939 ms
2022-11-16T13:22:22.679875 :  15,083 ->  15,090 (+  7) / extract features:     5.415 ms / detection:    14.765 ms
2022-11-16T13:22:23.221686 :  15,090 ->  15,096 (+  6) / extract features:     4.824 ms / detection:    15.600 ms
2022-11-16T13:22:23.756652 :  15,096 ->  15,101 (+  5) / extract features:     4.069 ms / detection:    13.923 ms
...
```

그림 3. 이상 탐지 결과 예
Fig. 3. Example of anomaly detection results

	CPU	CPU Core & Clock	RAM	VGA	OS
Development Setup	AMD Ryzen™ 7 3700X	8 Cores 16 Threads 3.6 GHz (4.4 GHz)	DDR4 32GB	NVIDIA GeForce RTX 3060	Microsoft Windows 10 (21H1)
Test Setup	AMD Ryzen™ 3 2200G	4 Cores 4 Threads 3.5 GHz (3.7 GHz)	DDR3 64GB	AMD Radeon™ Vega 8 Graphics (CPU Integrated)	ArchLinux (kernel 5.18.16)

그림 4. 이상 탐지 시뮬레이션 환경
Fig. 4. Environment of anomaly detection simulation

시뮬레이션이 시작되면 블록체인에 생성된 마지막 블록 번호를 기록한 후 시뮬레이터가 10분간 데이터 입력 요청을 정상 행위, 비정상 행위 두 가지로 나누어 요청하였다. 정상적 요청 행위로 초당 10~20회, 비정상적인 요청은 초당 100회의 데이터 입력 요청을 발생시켰다. 또한 차후 과정에서 활용하기 위해 비정상 행위 발생 시간을 기억하도록 하였다.

데이터 입력 요청이 발생되면 블록체인 내에서 트랜잭션이 생성, 블록을 생성하게 된다. 시뮬레이션 작업이 끝나면, 시뮬레이션 중 발생한 블록을 대상으로 실제 블록의 이상 여부 Answer의 결정을 수행하였다. 이때 정상 행위 중 발생한 블록은 Answer=0으로, 앞서 시뮬레이션에서 비정상 행위 발생 시간 중 생성된 블록은 Answer=1으로 결정하였다.

이 과정을 통해 결정된 Answer, 해당 블록의 특징값을 계산하여 블록 번호와 함께 별도의 파일로 저장하여 이상 탐지 AI 학습 및 평가에 활용할 수 있도록 하였다.

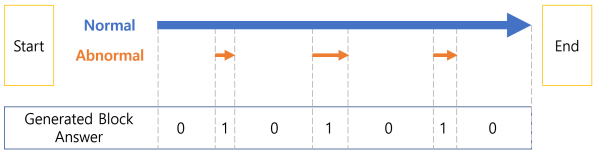

그림 6. 블록체인 이상 결정 (Answer)
Fig. 6. Decision of blockchain anomaly (Answer)

4.2 블록체인 네트워크 이상 탐지 평가

시뮬레이터를 동작시켜 얻은 특징값, Answer 데이터를 바탕으로 이상 탐지 AI을 학습시킨 후, 모델의 성능을 평가 작업을 수행하였다. 이때 시뮬레이터는 총 5회(50분) 수행하여 40,274개의 학습 데이터를 수집하였고, 정상 행위로 분류된 블록 Answer=0은 38,392개로 전체의 95.32%, 비정상 행위로 분류된 블록 Answer=1 은 1,882개로 전체의 4.68%를 차지한다.

수집된 모든 블록의 특징값을 탐지 AI에 입력하여, 얻은 이상 탐지 결과(Prediction)을 결정하였다. 이때 정상 행위로 판별된 블록을 Prediction=0, 비정상 행위로 판별된 블록을 Prediction=1으로 결정하였다. 이후 Answer값과 Prediction값을 토대로 모델의 성능을 평가하였다. 학습된 모델의 성능은 표 3과 같다. 전체의 블록 100개 중 정상/비정상을 올바르게 탐지하는 Accuracy는 98.1%의 우수한 성능을 확인할 수 있지만 비정상 블록을 비정상으로 정확히 판단한는 sensitivity는 향후 연구를 통해 진행할 예정이다. 따라서 블록체인의 이상 탐지는 100개의 블록중 정확히 몇 개가 이상으로 탐지했는지에 대한 실험 결과이다.

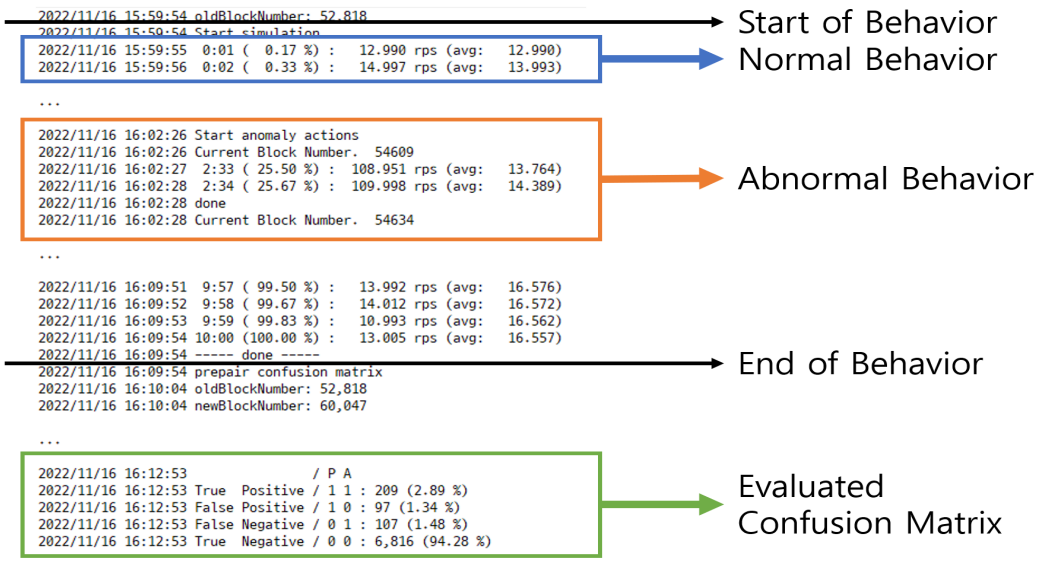

그림 5. 이상 탐지 시뮬레이션 결과
Fig. 5. Results of anomaly detection simulation

표 2. 학습 모델의 성능 평가 결과
Table 2. Results of learning model performance evaluation

Evaluation time	0.058 ms for each block
False positive	1.2 %
False negative	0.6 %
Accuracy	98.1 %
F1-score	78.0 %

또한 학습 데이터의 이상 탐지 시각화를 위해 학습 데이터에서 특징값만을 활용하여 PCA 모델을 비지도 학습하여 차원 축소를 진행하였고, 학습 데이터의 특징값을 2차원 그래프로 축소한 결과는 그림 7과 같다.

정상 블록(Answer=0)은 청록색으로, 비정상 블록(Answer=1)인 블록은 주황색으로 표시하였으며, 화살표가 가리키는 주황색 원 안에 비정상 블록이 위치하는 것을 확인할 수 있다. 또한 그림 8과 같이 정상 데이터와 비정상 데이터를 텍스트 모드에서도 확인하였다.

V. 결론 및 향후 과제

블록체인 기술이 다양한 산업 분야에서 활용되고 기술 발전이 거듭되면서 보안 위협 요소가 지속적으로 등장하고 있다. 이러한 보안 요소에 대한 선제적 대응 기술과 블록체인 네트워크의 이상 행위를 탐지 기술 개발은 매우 중요하고 시급하다. 특히, 최근 머신 러닝 기반으로 다양한 보안 위협 요소, 이상 행위 탐지 연구는 중요한 의미를 가지고 있다.

본 논문에서는 블록에 저장되는 트랜잭션 정보를 기반으로 블록체인 네트워크의 이상 탐지를 기존 학습 모델을 적용한 후 시뮬레이션을 통해 탐지율의 우수성과 시간 효율성을 확인하였다.

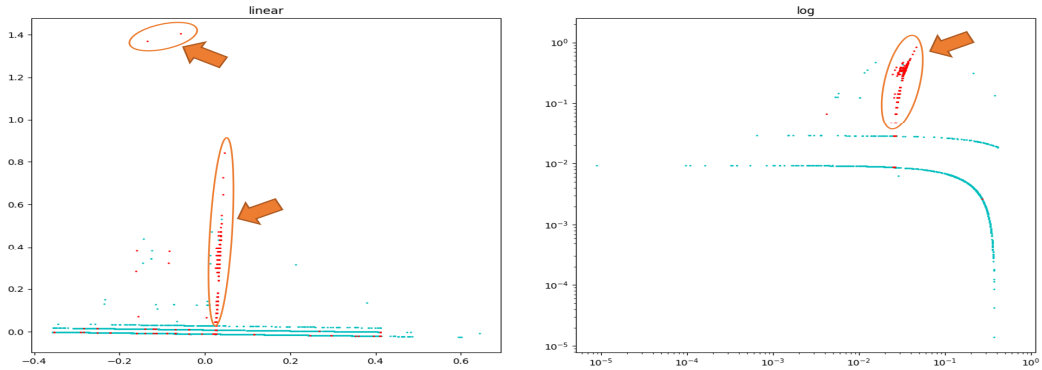

그림 7. 이상 탐지 결과의 시각화
Fig. 7. Visualization of anomaly detection results

그림 8. 정상 및 비정상 학습 데이터 비교
Fig. 8. Comparison of learning data

이러한 연구는 블록의 정보를 이용하여 이상 행위를 탐지할 수 있는 선도적 연구로서 가치가 있으며 향후에 실제 블록체인 메인넷에 탑재하여 서비스가 가능하다는 것을 확인하였다.

향후에는 본 연구의 결과를 프라이빗 블록체인에 적용하여 이상 행위 탐지 시간을 단축하는 연구와 블록의 비정상 정도를 측정하는 실험연구를 진행할 예정이다.

References

[1] E. Jun and C. Lee, "Analysis of Technology and Security Threats on Blockchain", Journal of the Korea Society of Digital Industry and Information Management, Vol. 14, No. 4, pp. 47-56, 2018.

[2] E. Lee, C. Han, J. Moon, and I. G. Le, "Blockchain Network Security Threats Detection Technique Trend Analysis", Review of KIISC, Vol. 31, No. 3, Jun. 2021.

[3] M. K. Shrivas, T. Y. Dean, and S. S. Brunda, "The Disruptive Blockchain Security Threats and Threat Categorization", 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T), Raipur, India, Jan. 2020. https://doi.org/10.1109/ICPC2T48082.2020.9071475.

[4] M. U. Hassan, M. H. Rehmani, and J. Chen, "Anomaly Detection in Blockchain Networks: A Comprehensive Survey", IEEE Communications Surveys & Tutorials, Dec. 2021. https://doi.org/10.48550/arXiv.2112.06089.

[5] Y. Chen, H. Chen, Y. Zhang, M. Han, M. Siddula, and Z. Cai, "A survey on blockchain systems: Attacks, defenses, and privacy preservation", High Confidence Computing, Vol. 2, No. 2, Jun. 2022. https://doi.org/10.1016/j.hcc.2021.100048.

[6] M. Signorini, M. Pontecorvi, W. Kanoun, and R. Di Pietro, "BAD: A Blockchain Anomaly Detection Solution", IEEE Access, Vol. 8, pp. 173481-173490, Sep. 2020. https://doi.org/10.1109/ACCESS.2020.3025622.

[7] P. Zheng, et al., "A detailed and real-time performance monitoring framework for blockchain systems", 2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Practice Track, pp. 134-143, May 2018. https://doi.org/10.1145/3183519.3183546.

[8] S. Farrugia, J. Ellul, and G. Azzopardi, "Detection of illicit accounts over the Ethereum blockchai", Expert Systems with Applications, Vol. 150, Jul. 2020. https://doi.org/10.1016/j.eswa.2020.113318.

[9] V. Patel, L. Pan, and S. Rajasegarar, "Graph Deep Learning Based Anomaly Detection in Ethereum Blockchain Network", NSS 2020: Network and System Security, pp. 132-148, Dec. 2020. https://doi.org/10.1007/978-3-030-65745-1_8.

[10] J. S. Lee, "Final Technical Report: Blockchain Ledger and Smart Contract Data Structure Visualization Tool", National IT Industry Promotion Agency (NIPA), 2022.

[11] S. C. Tan, K. M. Ting, and T. F. Liu, "Fast Anomaly Detection for Streaming Data", IJCAI 2011 Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, pp. 1511-1516, Jul. 2011. http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-254.

[12] M. E. Tipping and C. M. Bishop. "Mixtures of probabilistic principal component analyzers", Neural computation, Vol. 11, No. 2, pp. 443-482, 1999.

저자소개

고 광 만 (Kwang-Man Ko)

1991년 2월 : 원광대학교 컴퓨터공학과(공학사)
1993년 2월 : 동국대학교 컴퓨터공학과(공학석사)
1998년 2월 : 동국대학교 컴퓨터공학과(공학박사)
2023년 5월 ~ 현재 : 상지대학교 컴퓨터공학과 교수
관심분야 : 프로그래밍 언어 및 컴파일러, 소프트웨어 보안