그래프 임베딩 기반의 이더리움 피싱 스캠 탐지 연구

정유영 [1], 김경태 [1], 임동혁 [2]
[1] 광운대학교 인공지능응용학과
[2] 광운대학교 정보융합학부
yycheong@kw.ac.kr, kkt9601@kw.ac.kr, dhim@kw.ac.kr

Ethereum Phishing Scam Detection Based on Graph Embedding

Yoo-Young Cheong[1], Gyoung-Tae Kim[1], Dong-Hyuk Im[2]
[1]Dept. of Artificial Intelligence Applications, KwangWoon University
[2]School of Information Convergence, KwangWoon University

요 약

최근 블록체인 기술이 부상하면서 이를 이용한 암호화폐가 범죄의 대상이 되고 있다. 특히 피싱 스캠은 이더리움 사이버 범죄의 과반수 이상을 차지하며 주요 보안 위협원으로 여겨지고 있다. 따라서 효과적인 피싱 스캠 탐지 방법이 시급하다. 그러나 전체 노드에서 라벨링된 피싱 주소의 부족으로 인한 데이터 불균형으로 인하여 지도학습에 충분한 데이터 제공이 어려운 상황이다. 이를 해결하기 위해 본 논문에서는 이더리움 트랜잭션 네트워크를 고려한 효율적인 네트워크 임베딩 기법인 trans2vec 과 준지도 학습 모델 tri-training 을 함께 사용하여 라벨링된 데이터뿐만 아니라 라벨링되지 않은 데이터도 최대한 활용하는 피싱 스캠 탐지 방법을 제안한다.

1. 서론

블록체인은 양 당사자 간의 거래와 관련된 정보를 검증 가능하고 영구적으로 기록하는 분산형 공개 원장이다 [1]. 나카모토 사토시는 비트코인 프로젝트를 시작하여 블록체인을 이용한 최초의 성공적인 암호화폐를 소개하였다 [2]. 이더리움은 스마트 컨트랙트를 지원하는 블록체인 플랫폼이며 사용하는 암호화폐 이더(ether)를 기반으로 고속 성장하여 현재 비트코인 이후 최대 규모의 플랫폼이다. 최근 이러한 암호화폐에 대한 관심이 높아지며 다양한 사이버 범죄의 대상으로 문제가 되고 있다. 특히 이더리움에서는 2017 년 이후 피싱(Phishing) 사기 사건이 전체 사이버 범죄의 50% 이상을 차지하고 있으며, 이더리움 거래 보안의 주요 위협원이 되고 있다 [3].

피싱 사기는 신뢰할 수 있는 개체로 위장하여 사용자의 민감 정보를 얻으려고 시도한다. 이더리움에서의 피싱 사기는 피해자를 통해 거래 취소나 변경이 불가능한 이더를 피의자 측으로 이전시키는 것으로 발생할 수 있다 [4]. 피싱 스캠 탐지 문제는 다양하게 논의되었으며, 여러 가지 방법이 제안되었다 [5]. 그러나 이메일과 웹 사이트에 의존하는 기존의 일반적인 피싱 범죄 수법과 다르게 블록체인 플랫폼에서의 피싱 수법은 더 다양한 소스에 의존한다. 따라서 블록체인 플랫폼에서의 피싱 탐지 문제를 해결하기 위한 연구들이 진행되고 있다 [1] [6].

개방성과 투명성이 특징인 블록체인 암호화폐는 거래 기록에서 정보 추출이 가능하여 [7] 추출한 정보를 이용하여 피싱을 탐지한다. 트랜잭션 이력을 네트워크로 모델링하면, 노드는 고유한 주소이고 노드 엣지는 두 주소 사이에 하나 이상의 이더 전송이 존재하는 것을 의미한다. 그러나 피싱 탐지를 위한 트랜잭션 네트워크 사용은 극심한 데이터 불균형 문제가 존재한다. 이더리움의 블록 탐색 및 분석 플랫폼인 etherscan.io 에 따르면, 총 주소 및 트랜잭션 수는 각각 5 억개, 38 억 개를 초과하는 반면, 라벨링된 피싱 주소는 2041 개에 불과한다 [1]. 따라서 데이터 불균형은 피싱 탐지 문제에서 지도 학습 접근법을 사용할 경우 성능에 영향을 미칠 수 있다. 또한 이러한 대규모 네트워크 데이터를 학습에 사용할 경우 feature 선택이 성능에 중요하다.

이에 본 연구는 트랜잭션 네트워크에서 그래프 임베딩을 사용하여 feature 를 추출하고, 라벨링된 데이터를 생성하여 분류하는 준지도 학습 알고리즘으로 피싱 노드를 탐지하는 방법을 제안한다.

2. 연구 배경 및 관련 연구

네트워크 임베딩을 기반으로 한 이상 탐지 알고리즘은 현재 여러 업계에서 많은 관심을 끌고 있다. 네트워크 임베딩은 노드 간 임베딩 매핑 기능을 학습하여 d 차원 feature 공간에서 이웃 노드의 공존 가능성을 극대화하는 것을 목표로 한다. 노드 간의 구조적 관계를 포착하는 방법 중 랜덤 워크 기법으로는 DeepWalk [8], Node2vec [9] 등이 존재하며, 이더리움 피싱 스캠 탐지 연구로 Wu, Jiajing et al [1] 은 대규모 이더리움 트랜잭션 네트워크에서 거래 금액과 타임스탬프 정보에 편향된 feature 를 추출하는 Trans2vec 기법을 제안하였다.

라벨링된 데이터와 라벨링되지 않은 데이터를 모두 사용하는 준지도 학습 알고리즘인 Tri-training [10] 은 세 개의 분류기를 사용하는 알고리즘으로, 라벨링되지 않은 데이터에 레이블을 지정한 후 최종 분류를 결정하는 기법이다. He, Ying et al [11] 은 Node2vec 그래프 임베딩으로 데이터 feature 를 추출 후 Tri-training 을 사용하여 Internet water army 를 탐지하는 모델을 제안하였다. 그러나 블록체인 피싱 스캠 탐지를 위한 모델이 아니며, 사용하는 그래프 임베딩 기법은 트랜잭션 네트워크에서 구체적인 거래 정보를 효과적으로 추출하는 방법이 아니다.

따라서 본 연구에서는 이더리움 트랜잭션 네트워크에서 효과적으로 feature 를 추출하기 위하여 Trans2vec 을 사용하며, 데이터 불균형 문제를 개선하기 위하여 Tri-Training 으로 라벨링되지 않은 데이터에 레이블을 지정한 후 피싱 및 비피싱 노드를 분류한다.

3. Method

본 논문에서 피싱 스캠 탐지 모델은 그림 1 과 같이 몇 가지 단계로 구성되어 있다. (1) 이더리움 클라이언트를 통해 수집된 트랜잭션 이력과 Etherscan.io, EtherScam DB 에서 라벨링된 피싱 주소를 결합하여 노드를 피싱 및 기타 주소로 분류하고 엣지가 두 주소 간의 트랜잭션을 나타내는 이더리움 트랜잭션 네트워크를 구축한다. (2) 네트워크에서 거래 금액과 타임스탬프 정보 feature 를 추출하기 위하여 Trans2vec 으로 임베딩 한다. (3) 준지도 학습 알고리즘인 Tri-training 으로 피싱 및 비피싱 노드를 분류한다.

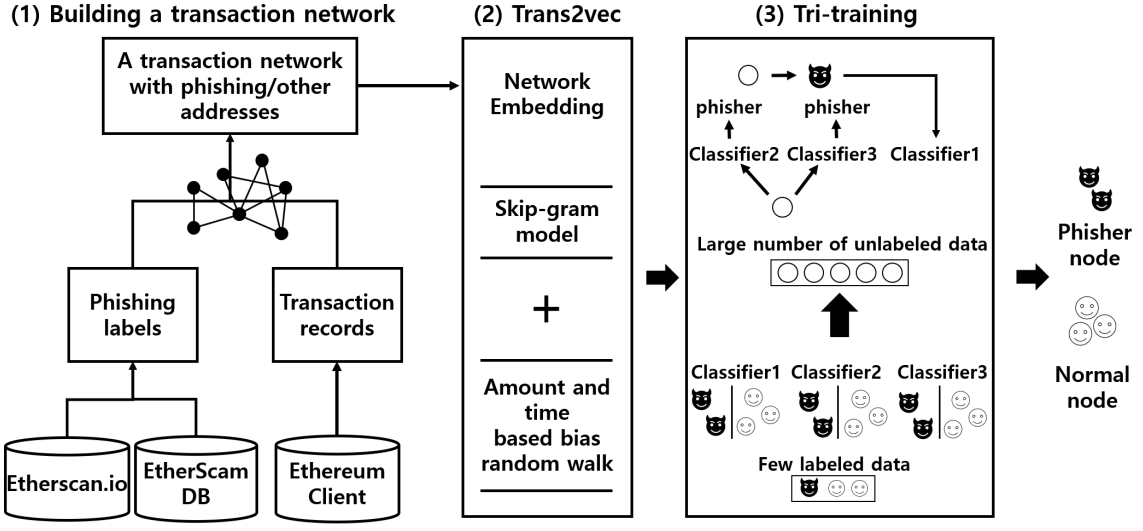
네트워크 임베딩에서 랜덤 워크를 수행함으로써 대규모 네트워크는 샘플링된 노드 시퀀스 집합으로 변환된다. Trans2vec 은 랜덤 워크 기반 그래프 임베딩 기법으로, 거래 금액과 타임스탬프 정보에 편향된 샘플링을 한다. 이더리움 트랜잭션 네트워크에서는 거래 금액이 클수록 두 노드 간의 관계가 강해진다는 것을 의미하며, 소스 노드로부터 랜덤 워크를 수행하여 매 단계마다 매개 변수 α 값을 통해 트랜잭션 금액과 시간 사이의 값을 조절하여 샘플링한다. 또한 특정 노드에 대한 이웃 노드 시퀀스로부터 노드들의 발생 확률을 최대화하는 함수 f 를 최적화하기 위하여 스킵그램 모델을 사용한다.

Tri-training 은 세 가지 분류기를 사용하는 알고리즘이다. L 은 라벨링된 데이터 집합이고 U 는 라벨링되지 않은 데이터 집합일때, 초기 분류기 (M1, M2, M3) 는 L 에서 샘플링된 데이터 L1, L2, L3 로 학습된다. 이후 M2 및 M3 분류기는 U 에서 동일한 샘플을 예측한다. 두 분류기의 예측 결과가 동일하면 해당 샘플은 높은 신뢰도를 갖는 것으로 간주되며, 세 번째 분류기 M1 의 라벨링된 훈련 세트에 추가된다.

4. 실험 및 평가

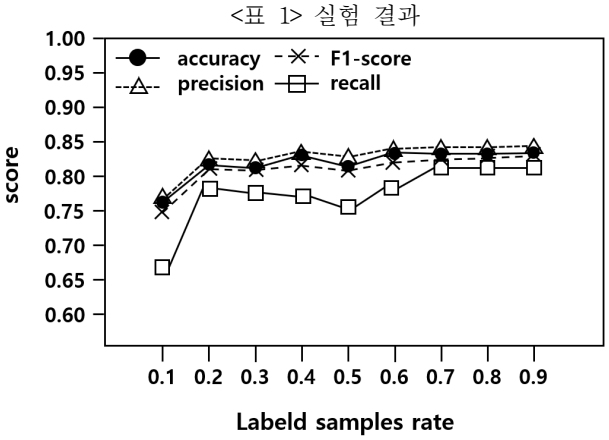
Etherscan.io 에서 제공하는 피싱 주소를 이용하여 라벨링된 주소와 라벨링된 주소의 1 차 거래 기록을 포함하는 데이터로 피싱 주소 445 개를 포함한 트랜잭션 네트워크를 구축하였다.

Trans2vec 으로 얻은 벡터 표현은 64 차원이며, Tri-training 알고리즘에서는 세 가지 분류기를 사용하므로 랜덤 포레스트, BP Neural networks 및 Naïve Bayes 알고리즘을 사용하였다. 적은 양의 라벨링된 데이터에서



(그림 1) 이더리움 피싱 스캠 탐지 구조

<표 1> 실험 결과



좋은 결과를 얻을 수 있는 지 확인하기 위하여 라벨링된 데이터 비율을 10%부터 90%까지 증가시켜 실험하였다. 실험에서 성능을 평가하기 위해 accuracy, Recall, Precision, F1-score 평가 지표를 사용하였으며, 실험 결과는 표 1 과 같다.

10%의 비율에서만 약간 다른 수치를 보이지만 기본적으로 모든 평가지표가 비율별로 변동하지 않았다. 이는 라벨링된 데이터가 적은 조건에서도 대략적인 탐지 결과를 얻을 수 있음을 나타낸다.

5. 결론

본 논문에서는 트랜잭션의 특징에 맞는 Trans2vec 네트워크 임베딩 기법과 라벨링 되지 않은 데이터도 학습에 사용하기 위하여 tri-training 준지도 학습 알고리즘을 함께 사용한 탐지 방법을 제안하였다. 실험을 통하여 레이블이 적은 암호화폐 트랜잭션 네트워크 데이터에 대하여 기존 라벨링된 데이터를 기반으로 3 가지 분류기의 시너지를 활용하여 레이블을 확장시켜, 라벨링된 데이터 비율이 적어도 탐지 결과를 얻을 수 있음을 확인하였다.

본 연구는 이더리움 플랫폼으로 특정하여 진행하였으나 추후 연구를 통하여 다른 블록체인 플랫폼으로 확장할 수 있을 것이며, 레이블 확장에 있어 피싱 스캠 탐지에 좀 더 효과적인 모델이 될 수 있도록 성능 개선 연구를 진행할 것이다.

Acknowledgement

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2021-0-00231, 빅데이터 대상의 빠른 질의 처리가 가능한 탐사 데이터 분석 지원 근사질의 DBMS 기술 개발, 50%)과 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음(IITP-2022-2018-0-01417, 50%)

참고문헌

[1] Wu, Jiajing, et al. "Who Are The Phishers? Phishing Scam Detection on Ethereum via Network Embedding" IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020.

[2] Nakamoto, Satoshi. “Bitcoin: A peer-to-peer electronic cash system.” Decentralized Business Review, 21260, 2008.

[3] Conti, Mauro, et al. “A Survey on Security and Privacy Issues of Bitcoin” IEEE Communications Surveys & Tutorials, 20(4), 3415-3452, 2018.

[4] Low, Kelvin FK, et al. “Legal risks of owning cryptocurrencies” Handbook of Blockchain, Digital Finance, and Inclusion, Volume 1, London, Academic Press, 2018.

[5] Khonji, Mahmoud, et al. “Phishing detection: A literature survey” IEEE Communications Surveys & Tutorials, 15(4), 2091-2121, 2013.

[6] Chen, Liang, et al. “Phishing scams detection in Ethereum transaction network” ACM Transactions on Internet Technology (TOIT), 21(1), 1-16, 2020.

[7] Lin, Dan, et al. “Modeling and understanding Ethereum transaction records via a complex network approach” IEEE Transactions on Circuits and Systems II: Express Briefs, 67(11), 2737-2741, 2020.

[8] Perrozzi, Bryan, et al. “Deepwalk: Online learning of social representations” 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York (USA), 2014, 701-710.

[9] Grover, Aditya, et al. “node2vec: Scalable feature learning for networks” 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, San Francisco (USA), 2016, 855-864.

[10] Zhou, Zhi-Hua, et al. “Tri-training: Exploiting unlabeled data using three classifiers” IEEE Transactions on knowledge and Data Engineering, 17(11), 1529-1541, 2005.

[11] He, Ying, et al. “Semi-supervised internet water army detection based on graph embedding” Multimedia Tools and Applications, 1-22, 2022.