코드 난독화에 강건한 GAT 기반 크립토재킹 정적 탐지 방안 제안

김민수 [1], 김미희 [2]
[1] 한경국립대학교 컴퓨터응용수학부 학부생
[2] 한경국립대학교 컴퓨터응용수학부 (컴퓨터 시스템 연구소) 교수

gsnake7@naver.com, mhkim@hknu.ac.kr(교신저자)

A Proposed Method for Robust Static Detection of Obfuscated Cryptojacking using GAT

Min-su Kim, Mi-hui Kim
School of Computer Engineering & Applied Mathematics, Hankyong National University

요 약

크립토재킹은 공격자가 사용자의 동의 없이 시스템 자원을 탈취해 암호화폐를 채굴하는 악성 행위이다. 공격자들은 탐지를 회피하기 위해 제어 흐름 그래프(CFG) 분석을 무력화하는 코드 난독화 기술을 사용할 수 있으며, 이로 인해 기존 정적 분석 기반 탐지 모델들은 난독화된 크립토재킹 탐지에 명백한 한계를 보인다. 이러한 문제를 해결하기 위해 본 논문은 다양한 난독화 기법이 적용된 바이너리의 CFG 데이터셋을 구축하고, 그래프의 구조적 노이즈 속에서 핵심 패턴을 학습할 수 있는 그래프 어텐션 네트워크(GAT) 모델을 활용하여 탐지하는 방법을 제안한다. 제안하는 모델은 여러 난독화 기법이 복합적으로 적용된 환경에서도 높은 탐지 성능을 보일 것으로 기대된다.

1. 서론

크립토재킹(Cryptojacking)은 공격자가 사용자의 동의 없이 시스템 자원을 탈취하여 암호화폐를 채굴하는 악성 행위로, 최근 클라우드 및 서버 환경에서 크게 증가하고 있다[1]. 이러한 공격은 시스템 성능 저하, 전력 소비 증가, 하드웨어 수명 단축 등 심각한 피해를 야기하며, 이는 단순한 자원 도용을 넘어 은밀하고 지속적인 공격 형태라는 점에서 그 심각성이 있다. XMRig[2]와 같은 오픈소스 채굴 프로그램은 공격자들에게 널리 악용되며, 이들의 탐지를 회피하기 위한 기술 역시 고도화되고 있다.

이러한 위협에 대응하기 위해 다양한 정적 분석 기법이 연구되어 왔다. 초기에는 악성코드의 고유 문자열이나 바이트 패턴을 미리 정의된 규칙과 비교하여 악성 여부를 판단하는 시그니처 기반 탐지가 주를 이루었으나, 이는 신종 및 변종 탐지에 명백한 한계를 보였다. 이후 CPU 가 직접 실행하는 명령어의 시퀀스로, 프로그램의 구체적인 행위를 나타내는 Opcode 시퀀스를 머신러닝으로 분석하는 방식이 등장했지만, 복잡한 제어 흐름을 모델링하는 데에는 한계가 있었다.

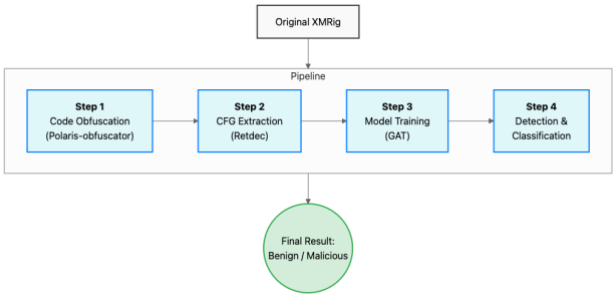
최근 Zheng 등이 제안한 MBGINet 은 제어 흐름 그래프(CFG)와 그래프 신경망(Graph Neural Network, GNN)의 일종인 그래프 동형 네트워크(Graph Isomorphism Network, GIN)를 활용해 높은 탐지 성능을 보였지만, 해당 연구에서도 결론에서 '제어 흐름 분석을 방해하는 난독화 기술' 에 취약하다는 점을 한계로 명시했다[3].

본 연구는 바로 이 지점에서 출발하여, 기존 정적 탐지 방법의 한계를 극복하기 위한 방안을 제시하고자 한다. 이를 위해 XMRig 를 기반으로 다양한 난독화 기법을 복합적으로 적용한 데이터셋을 구축하고, 이를 효과적으로 학습하여 코드 난독화에 강건한 그래프 어텐션 네트워크(GAT) 기반 정적 탐지 기법을 제안한다.

2. 제안하는 탐지 기법

본 연구는 코드 난독화 문제를 해결하기 위해, [그림 1]과 같이 총 4 단계로 구성된 GAT 기반 탐지 파이프라인을 제안한다. 먼저, (1 단계) 원본 샘플에 난독화 기법을 적용하여 데이터셋을 생성하고, (2 단계) 생성된 바이너리로부터 제어 흐름 그래프(CFG)를

추출한다. 다음으로, (3 단계) 추출된 CFG 를 그래프 어텐션 네트워크(GAT) 모델로 학습시키고, (4 단계) 학습된 모델을 통해 최종적으로 악성 여부를 판별한다. 이어지는 절에서는 각 단계의 핵심 요소인 데이터셋 구축과 GAT 모델에 대해 상세히 서술한다.



(그림 1) 제안하는 탐지 기법의 전체 파이프라인

2.1 데이터셋 구축

모델 학습을 위해 먼저, 실제 위협과 유사한 난독화 데이터셋을 구축한다. 오픈소스 암호화폐 채굴 프로그램인 XMRig 에 대하여 Polaris-obfuscator[4]를 사용해 다양한 코드 난독화 기법을 복합적으로 적용하여 빌드를 수행할 수 있다. Polaris-obfuscator 는 제어 흐름 평탄화, 간접 호출 난독화, 간접 분기 난독화, 별칭 접근 난독화, 가짜 제어 흐름 난독화, 명령어 대체 난독화와 같은 다양한 난독화 기법을 지원하여, 정적 분석을 통해 생성되는 CFG 의 구조를 의도적으로 복잡하게 만드는 데 효과적이다. 이후, 난독화된 바이너리로부터 Retdec[5] 디컴파일러 도구를 사용하여 CFG 를 추출한다. 이 과정을 통해 생성된 데이터셋은, 다양한 패턴으로 왜곡된 그래프 구조를 포함하게 되어 모델이 강건성을 갖추도록 훈련시키는 데 사용된다.

2.2 GAT 기반 난독화 대응 모델

제안하는 모델은 그래프 어텐션 네트워크(GAT)를 핵심 아키텍처로 채택한다. GAT 는 그래프 구조 데이터에 어텐션 메커니즘을 적용한 선구적인 모델로, 이웃 노드들의 중요도를 각각 다르게 학습하는 것이 특징이다 [6]. 즉, 모든 이웃 노드를 동등하게 취급하는 기존의 GNN 과 달리, GAT 는 어떤 이웃이 더 중요한 정보를 담고 있는지 스스로 판단하여 가중치를 부여한다.

이러한 GAT 의 핵심 원리는 난독화된 CFG 분석 문제에 매우 효과적으로 적용될 수 있다. 난독화 기법으로 인해 CFG 에 불필요한 노드나 엣지(가짜 제어 흐름 등)가 추가되더라도, GAT 의 셀프 어텐션(self-attention) 메커니즘은 이러한 '노이즈'들의 중요도를 낮게 학습하고, 실제 악성 행위와 관련된 핵심적인 제어 흐름 패턴에만 집중할 수 있다. 이는 여러 기법이 복합적으로 적용되어 예측 불가능하게 변형된 그래프 구조에서도 안정적인 탐지 성능을 유지하게 하는 핵심 원리이다.

이처럼 본 연구에서 제안하는 파이프라인은 오픈소스 채굴 프로그램을 통해 의도적으로 난독화된 데이터셋을 구축하는 단계와, GAT 의 어텐션 메커니즘을 통해 그래프의 구조적 노이즈 속에서 악성 행위의 본질을 학습하는 모델링 단계가 유기적으로 결합되어, 기존 정적 분석의 한계를 극복하도록 설계되었다.

3. 결론

본 연구에서는 기존 정적 탐지 기법의 코드 난독화 취약점 문제를 해결하기 위해, 다양한 난독화 기법을 적용한 데이터셋으로 GAT 모델을 학습시키는 새로운 정적 탐지 방법론을 제안하였다. 본 연구에서 제안하는 모델은 실제 공격에 널리 악용되는 오픈소스 채굴 프로그램을 대상으로 난독화 강건성을 확보하여, 기존 정적 분석의 한계를 극복하는 새로운 방향을 제시할 수 있을 것으로 기대된다.

향후 연구에서는 본 논문에서 제안한 방법론에 대한 실험을 진행하여 어떤 모델이 탐지에 좋은 결과를 보이는지 비교하는 실험을 진행할 예정이다.

사사 (Acknowledgement)

이 논문은 한경국립대학교 국립대학육성사업(2025) 지원을 받아 작성되었음

참고문헌

[1] Baggman, E. “Discovery & Mitigation of Cryptojacking in Cloud Systems using Honeypots,” Master’s thesis, KTH School of Electrical Engineering and Computer Science, Stockholm, 2024.

[2] XMRig, Available: https://github.com/xmrig/xmrig

[3] Zheng, Rui; Wang, Qiuyun; He, Jia; Fu, Jianming; Suri, Guga; Jiang, Zhengwei, “Cryptocurrency Mining Malware Detection Based on Behavior Pattern and Graph Neural Network,” *Security and Communication Networks*, vol. 2022, pp. 1-8, 2022.

[4] Polaris-obfuscator, Available: https://github.com/za233/Polaris-Obfuscator

[5] Retdec, Available: https://github.com/avast/retdec

[6] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph Attention Networks. *International Conference on Learning Representations (ICLR)*.