P R O J E C T D O C U M E N T

문제점 개요서

Project Name	바이너리 프로그램에서 제어 구조를 식별하는 도구 개발

14 조

202002514 안상준
202202602 손예진
202202487 박혜연

지도교수: 조은선 교수님 (서명)

Document Revision History

Rev#	Date	Affected Section	Author
1	2025/03/19	문제점 개요서 작성	박혜연
2	2025/03/20	문제점 개요서 작성	안상준

Table of Contents

1.Survey Paper - Limitations Focus

번호	연구 제목(저자)	저널/컨퍼런스(연도)	주요 내용 요약	한계점
1	Can LLMsObfuscate Code? A Systematic Analysis of Large Language Models into Assembly Code Obfuscation (Seyedreza Mohseni, Seyedali Mohammadi, Deepa Tilwani, Yash Saxena, Gerald Ketu Ndawula, Sriram Vema, Edward Raff, Manas Gaur)	AAAI(2025)	LLM의 prompting과 in-context-learning을 위한 어셈블리 코드 난독화 데이터셋 MAD를 제공하였다. 이는 LLM의 코드 난독화 성능을 테스트하기 위해 만들어진 첫 어셈블리 코드 난독화 데이터셋이다. 이 데이터셋을 GPT, CodeLLAMA, CodeGemma 등의 모델로 pre-training 하거나 prompting 함으로써 데이터셋의 학습가능성과 신뢰성을 입증하였다. Control Flow Change, Dead Code Insertion, Register Substitution 으로 세가지 난독화 기법에 대해 테스트를 진행하고, 결과를 Delta Entropy와 Cosine Similarity를 사용해 평가하였고, 사람이 평가하는 부분에서는 20년 이상 경력의 전문가들을 통해 진행되었다. 결과적으로, GPT 계열이 다른 Coder 모델보다도 난독화를 잘 수행했으며, 이 논문에 사용된 모델 중 가장 좋은 성능을 보였다.	다양한 종류의 LLM을 사용했지만, fine-tuning으로 학습한 모델이 한가지 있었고, 나머지는 in-context-learning만을 사용했다. 이는 정확한 비교가 될 수 없을 것이라고 생각한다. 뿐만 아니라 모델의 예측 결과를 평가할 Delta Entropy와 Cosine Similarity 값의 범위를 실험을 통해 선정하였지만, 이는 난독화 성능에 대한 객관적인 지표가 될 수 없다. 사람이 직접 평가한 부분에 대해서도 주관이 개입되지 않았다고는 보기 어렵다. 이 논문에서 제시한 방향과 데이터셋 구축 방안은 좋은 참고가 되었지만, 난독화 성능에 대한 객관적인 지표는 확인할 수 없었다.
2	PalmTree: Learning an Assembly Language Model for Instruction	CCS (2021)	PalmTree는 bert 기반의 모델이다. 일부를 마스킹 하는 Masked Language Model, 명령어 간의 문맥을 고려하는 Context Window Prediction, 데이터 흐름을 통해 명령어 간 데이터 흐름 관계	이 논문에서 제시한 임베딩 방안은 딥러닝 모델이 바이너리 코드를 더 잘 학습할 수 있게 해준다. 그러나 Bert 기반의 transformer 네트

	Embedding (Xuezixiang Li, Yu Qu, Heng Yin)		를 학습하는 Def-Use Prediction을 사용해 사전 학습을 진행했다. 내부 평가로는 이상치 탐지, 기본 블록 유사성 검색을, 외부 평가로는 바이너리 코드 유사성 탐지, 함수 원형 추론, 메모리 영역 분석을 활용했다. 이는 기존 임베딩 방법에 비해 높은 정확도를 보였다. 따라서 PalmTree는 기존 임베딩 방법의 한계를 극복하고, 바이너리 분석에 유용한 명령어 임베딩 방법을 제안했다. 뿐만 아니라 소스 코드와 사전 학습 모델을 공개하여 다른 연구에 사용될 수 있도록 했다.	워크를 사용하였기 때문에, 기존의 다른 모델보다 계산량이 많아 속도가 느리다는 문제가 있다.
3	Deobfuscating virtualized malware using Hex-Rays Decompiler	Virus Bulletin Conference(2023)	가상화 난독화 된 코드를 역난독화 하는 방법으로 독립적인 도구 대신 IDA Pro, Hex_rays Dccompiler를 사용하여 가상화된 코드를 역난독화하는 새로운 접근 방식을 제시. FinSpy VM을 통한 난독화된 코드에 대하여 IDA SDK의 기능을 사용하여 역난독화를 자동화하고, 가상화된 코드를 x86아키텍처로 변환하는 방법을 설명한다. Hex-Rays 마이크로코드를 사용하여 변환된 어셈블리 코드를 C로 디컴파일 하여 역난독화된 코드를 얻는 방법을 제시한다.	가상 머신의 내부 동작을 이해하고, 이를 x86과 같은 잘 알려진 아키텍처로 변환하는 것은 상당한 역공학과 분석이 필요하다. Hex-Rays Decompiler를 사용하여 코드 최적화를 수행하지만, 모든 경우세ㅓ 최적화가 효과적이지 않을 수 있다. 다른 가상머신에서 역난독화 기법이 적용될지는 명확히 알 수 없다.
4	Loki: Hardening Code Obfuscation Against Automated Attacks	USENIX Security Symposium(2022)	LOKI는 기존의 가상 머신 기반 난독화 기법을 강화하여 자동화된 역난독화 공격에 대응하는 방법을 제시. Mixed Boolean-Arithmetic(MBA) 표현식을 사용하여 코드를 난독화 하며, 기존의 공격 벡터인 Symbolic Execution, 테인트 분석, 프로그램 합성 등에 강력한 보호를 제공한다. 특히 프로그램 합성 공격의 성공률을 19%로 줄이는데 성공했다.	LOKI의 난독화 기법은 복잡하며, 이를 구현하는 데 상당한 기술적 지식이 필요하다. LOKI는 특정한 난독화 기법에 최적화되어 있으며, 다른 유형의 난독화나 플랫폼에 적용하는 데는 추가적인 연구가 필요하다.
3	VMProtect의 역공학 방해 기능 분석 및 Pin을 이용한 우회 방안	KTCCS(2021)	상용 난독화 도구인 VMProtect3.5.0을 통해 Debugger Detection, Virtualization Tools Detection을 적용시킨 실행파일을 Pin Tool을 이용하여 우회하는 방안을 제시. VMProtect	이 논문에서 제시한 우회 방안은 VMProtect 3.5.0 버전에 특화 돼 있으며, 다른 버전에서도 적용이 가능한지 확인이 필요합니다.

			의 안티리버싱 기법의 위치와 위치를 예상하고 이를 바탕으로 알고리즘을 작성하여 가상화 탐지, 디버거 탐지를 우회	해당 논문은 VMProtect에 특화돼 있으며, 다른 난독화 도구나 플랫폼에 적용하는 데는 추가적인 연구가 필요할 수 있다.

2.Limitations and Research Gaps

번호	기존 연구	한계점	연구 필요성	본 연구의 기여
1	Can LLMsObfuscate Code? A Systematic Analysis of Large Language Models into Assembly Code Obfuscation (Seyedreza Mohseni, Seyedali Mohammadi, Deepa Tilwani, Yash Saxena, Gerald Ketu Ndawula, Sriram Vema, Edward Raff, Manas Gaur), AAAI 2025	다양한 종류의 LLM을 사용했지만, fine-tuning으로 학습한 모델이 한가지 있었고, 나머지는 in-context-learning 만을 사용했다. 이는 정확한 비교가 될 수 없을 것이라고 생각한다. 뿐만 아니라 모델의 예측 결과를 평가할 Delta Entropy와 Cosine Similarity 값의 범위를 실험을 통해 선정하였지만, 이는 난독화 성능에 대한 객관적인 지표가 될 수 없다. 사람이 직접 평가한 부분에 대해서도 주관이 개입되지 않았다고는 보기 어렵다. 이 논문에서 제시한 방향과 데이터셋 구축 방안은 좋은 참고가 되었지만, 난독화 성능에 대한 객관적인 지표는 확인할 수 없었다.	우선, 다양한 LLM을 사용하여 실험해보되, fine-tuning 방법에서는 하나의 모델만을 다르게 실험하기보다는 최소 2개 이상의 모델을 같은 조건으로 학습시켜 좀 더 정확한 비교가 될 수 있도록 할 것이다. 추가적으로, 모델의 역난독화 성능을 평가하기 위한 지표에 대해서도 고민이 필요할 것 같다.	아직 LLM을 가상화 난독화에 사용한 연구가 없기 때문에 이는 보안 분야에서 새로운 접근이라고 할 수 있다. 또한, LLM에 바이너리 코드를 fine-tuning 하는 것과 in-context-learning 만을 하는 것의 차이를 비교해 바이너리 코드 분석에 어떤 방법이 적합한지 분석할 수 있을 것이라고 생각한다.

2	PalmTree: Learning an Assembly Language Model for Instruction Embedding (Xuezixiang Li, Yu Qu, Heng Yin), CCS 2021	이 논문에서 제시한 임베딩 방안은 딥러닝 모델이 바이너리 코드를 더 잘 학습할 수 있게 해준다. 그러나 Bert 기반의 transformer 네트워크를 사용하였기 때문에, 기존의 다른 모델보다 계산량이 많아 속도가 느리다는 문제가 있다.	모델의 레이어 수를 줄이거나, 더 효율적인 transformer 아키텍쳐를 사용해 성능은 유지하며 계산 비용만을 줄일 수 있을 것이다.	현재는 바이너리 코드를 벡터화 하는 방법과 바이너리 코드 분석에 LLM을 사용하는 것에 대한 연구가 따로 진행되고 있다. 그러나, 이 연구에서는 두가지를 함께 진행할 예정이다. 바이너리 코드의 특성에 맞게 벡터화 한 후 LLM을 학습시킨다면 기존보다 훨씬 좋은 성능이 나올 것이라고 생각한다.
3	Deobfuscating virtualized malware using Hex-Rays Decompiler(Georgy Kucherin, 2021)	가상 머신의 내부 동작을 이해하고, 이를 x86과 같은 잘 알려진 아키텍처로 변환하는 것은 상당한 역공학과 분석이 필요하다. Hex-Rays Decompiler를 사용하여 코드 최적화를 수행하지만, 모든 경우에서 최적화가 효과적이지 않을 수 있다.	다양한 가상머신에 적용할 수 있는 역난독화 방법이 필요하다. IDA Pro와 같은 기존의 독립적인 도구를 사용하여 코드르 해체하고 이를 x86과 같은 잘 알려진 아키텍처로 변환하여 분석하는 방식이 있다.	IDA Pro, Hex-Rays Dcompiler를 사용하여 가상화된 코드를 역난돡화 하는 효율적인 방법을 제시한다. 이러한 도구를 거친 데이터셋을 학습 데이터로 학습하여 가상화된 코드의 복접성을 줄이고 모델이 효율적으로 역난독화를 수행할 수 있도록 한다.
2	Loki: Hardening Code Obfuscation Against Automated Attacks KTCCS(2021)	LOKI의 복잡한 MBA 표현식은 분석을 어렵게 하지만, 이는 코드의 실행 속도와 공간 복잡성에 영향을 미칠 수 있다. MBA 표현식의 생성과 관리가 복잡할 수 있다. LOKI는 특정한 공격 벡터에 대해 최적화되어 있으며, 다른 유형의 공격이나 플랫폼에 적용하는 데는 추가적인 연구가 필요하다.	LOKI는 기존의 공격 벡터에 대해 강력한 보호를 제공하지만, 새로운 공격 벡터나 기술이 등장할 경우 이를 대응하기 위한 지속적인 연구가 필요하다. LOKI의 복잡한 표현식은 성능에 영향을 미칠 수 있으므로, 이를 최적화하여 오버헤드를 줄이는 연구가 필요하다.	LOKI는 다양한 난독화 기법을 조합하여 자동화된 공격에 강한 보호를 제공한다. LOKI 모델에 적용된 난독화 기법을 분석하고 난독화 된 데이터셋을 통해 LLM을 학습시켜 난독화 된 코드를 역난독화 하는 능력을 개선시킬 수 있다..
5	VMProtect의 역공학 방해 기능 분석 및 Pin을	이 논문에서 제시한 우회 방안은 VMProtect 3.5.0 버전에 특화되어 있으	다양한 환경에서 활용할 수 있도록 확장하는 연구가 필요하다. Pin을 사용하	VMProtect의 옵션과 역난독화 및 가상화 감지 방법에 대해 설명하며, 해당 탐지 기술을 Pin

	이용한 우회 방안 KTCCS(2021)	며, 다른 버전이나 스프트웨어 환경에서는 호환성 문제가 발생할 수 있다. VMProtect와 같은 상용 난독화 도구는 지속적으로 업데이트되고 있으며, 새로운 버전에서는 기존의 우회 방안이 효과적이지 않을 수 있다. Pin을 사용하여 우회하는 과정에서 성능 손실이 발생할 수 있다.	여 우회 코드를 자동화하는 과정에서 성능 손실을 줄이는 연구가 필요하다. 이 논문에서 제시한 우회 방안이 악의적으로 사용되지 않도록 하는 대책을 마련하는 연구가 필요하다.	Tool을 통해 우회하는 방법에 대해서 설명하고 있다. VMProtect로 난독화 된 코드에 대하여 Pin Tool을 이용하여 trace를 뽑아 LLM 학습 데이터로 활용할 수 있다. 이를 통해 데이터셋의 크기를 줄이며 feature를 쉽게 뽑을 수 있고, 코드를 동적으로 분석이 가능해져 역난독화를 좀 더 정확하게 수행하는데 기여할 수 있다.