사이버 사건 보고서를 활용한 LLM 기반 시나리오 생성 방법론 제안

강혜진[1], 조학수[2]
[1]호서대학교 컴퓨터공학부 학부생
[2]호서대학교 컴퓨터공학부 교수
gowls203@naver.com, marius71@hoseo.edu

A Methodology for LLM-Based Scenario Generation from Cyber Incident Reports

Hye-jin Kang[1], Harksu Cho[1]
[1]Dept. of Computer Engineering, Hoseo University

요 약

사이버 공격이 고도화됨에 따라 사이버 복원력의 확보는 핵심 과제로 대두되고 있다. 사이버 복원력을 측정하기 위해 다양한 방법이 제시되고 있으나, 이를 뒷받침할 시나리오 구성은 여전히 어렵다. 본 연구는 이러한 한계를 극복하기 위해 대규모 언어 모델(LLM)을 활용하여 실제 사건 보고서 기반 공격 시나리오 생성 방법론을 제안한다. 제안된 방법론은 역할 기반 LLM 전문가 할당 방식을 통해 사건 보고서의 공백을 보완하고, 현실성 있는 공격 흐름을 자동으로 구성한다. 이를 통해 사이버 복원력 평가 및 훈련 환경에서 사용할 수 있는 시나리오를 제공하고자 한다.

1. 서론

사이버 공격이 다양하게 진화함에 따라 정부 기관과 기업에서는 시스템이 불리한 조건, 스트레스, 공격 또는 침해 상황에 직면했을 때, 이를 예측하고(Anticipate), 견디며(Withstand), 회복하고(Recover), 적응(Adapt)할 수 있는 능력인 사이버 복원력에 주목하고 있다[1]. 이를 측정하기 위해 다양한 방법이 제안되었으나, 검증을 위한 환경 구성 중 하나인 시나리오 설계는 여전히 어렵다. 기존에는 시나리오를 생성하기 위해 수동 시나리오 생성 방법에 의존하고 있으며 수동 시나리오 생성법은 높은 비용, 시간, 전문가 인력 등이 필요하다는 한계를 가진다[2].

따라서 실제 사건 보고서와 대규모 언어 모델(LLM)을 활용한 자동 시나리오 생성법이 필요하다. 실제 사건 보고서는 공격자가 어떤 방식으로 침투, 확산, 탈취를 수행했는지 보여주기 때문에 시나리오 설계의 핵심 수단이 되지만, 공개된 보고서는 추상적이고 간략하게 기술되어 있어 구체적인 공격 행위나 실행 절차는 충분히 제공하지 못한다.

본 연구는 LLM의 추론 및 생성 능력을 활용하여 추상적으로 기술된 사건 보고서를 구체적 공격 행위로 확장하고 보고서의 공백을 보완하는 시나리오 생성 방법론을 제안한다.

2. 관련 연구

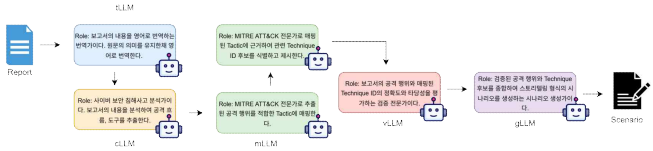
선행 연구에서는 수동 시나리오 생성의 한계를 극복하기 위해 LLM을 활용한 시나리오 생성 방법론이 다양하게 제시되었다.

Zhang et al. (2025)은 CTI(Cyber Threat Intelligence 보고서를 구조화된 공격 그래프로 변환하는 프레임워크인 AttacKG+ 제안하였다. CTI보고서는 보안 위협 정보를 자세하게 정리하여 위협을 빠르게 대응할 수 있도록 작성된 문서를 의미한다. 기존 시나리오 생성법의 한계였던 전문가 의존성의 문제를 LLM의 언어 이해 능력과 제로샷 능력을 사용하여 비전문가도 활용할 수 있는 자동화된 공격 그래프 구축을 가능하게 하였다[3]. 이는 CTI 보고서를 기반으로 LLM을 활용해 보고서의 전술 단계를 재구성하고 MITRE ATT&CK TTP 매핑을 통해 기존 모델 대비 F1 score를 향상시켰다.

Yamin et al. (2024)은 사이버 보안 훈련 시나리오 생성을 위해 LLM 기반 시나리오 생성 프레임워크인 CyExec를 제안한다. 이는 CISO와 보안 전문가 역할을 맡은 두 개의 LLM이 상호작용하는 방식으로 시나리오를 생성하며, LLM의 환각 특성을 이용하여 새롭고 복잡한 시나리오를 창의적으로 구현하였다[4].

이에 본 연구는 실제 사건 보고서를 기반으로 역할기반 LLM이 MITRE ATT&CK Tactics, Techniques, Procedures(TTP) 매핑을 수행하고 시나리오를 생성하는 방법론을 제안한다. MITRE ATT&CK은 실제 사이버 공격 사례를 기반으로 공격자의 행위를 구조화한 프레임워크로 사이버 위협에 대한 탐지 및 분석을 강화한다. 이를 통해 복원력 측정을 위한 시나리오 제공에 기여하고자 한다.

3. 제안하는 방법론



(그림 1) 시나리오 생성 파이프라인

본 연구는 다중 역할 기반 LLM 파이프라인을 통해 사건 보고서에 누락된 공격 흐름을 보완하는 시나리오 생성 방법론을 제안한다. Gosmar의 연구에 따르면 LLM에 역할을 부여하고 문제 해결을 수행하도록 프롬프트를 설계할 경우 환각(Hallucination) 현상이 완화된다고 제시한다[5]. 이를 위해 LLM의 추론 능력을 활용하고 구조화된 출력을 위해 역할 기반 LLM 전문가 할당 방식을 적용하였다. 따라서 선정된 역할은 번역가(tLLM), 사이버 보안 침해사고 분석 전문가(cLLM), MITRE ATT&CK 전문가(mLLM), 정보보안 시나리오 생성 전문가(gLLM), 검증 전문가(vLLM)로 총 5가지의 역할이다.

먼저, 실제 사건 사고 보고서를 입력하면 tLLM이 원문을 영어로 변환한다. 다양한 언어로 작성된 보고서를 통합적으로 활용하기 위함이며 이를 기반으로 cLLM이 보고서를 분석한다. 보고서의 공격 행위 및 사용된 도구와 관련된 문장을 식별하여 핵심 행위를 추출한다. 추출 행위에 대해 mLLM은 TTP 중, 14가지 Tactic에 매핑을 시도한다. 매핑되지 않는 행위는 '매핑 불가(None)'로 처리되며, 이후 공백을 보완하기 위해 Technique 후보를 확장한다. 즉, 매핑된 Tactic에 적합한 Technique ID 후보를 제시하고 None으로 처리된 행위에 대해서는 사건 보고서의 맥락을 고려하여 발생가능한 Technique ID를 후보군으로 나열한다. 후보군이 생성된 이후 vLLM은 보고서에 기술된 행위와 Technique ID의 정합성을 평가한다. MITRE ATT&CK 데이터셋에 기재된 Technique 설명과 사건 보고서의 공격 행위 문장들을 벡터화한 후, 코사인 유사도를 계산하여 보고서의 공백을 보완할 수 있는 적합한 Technique을 선택한다.

마지막으로 gLLM은 이러한 결과를 종합하여 사건 보고서의 공백을 보완할 Technique을 선정하고 스토리텔링 형식의 시나리오를 구성한다.

상세한 파이프라인은 (그림 1)로 제시하였다.

4. 결론

본 논문은 사이버 사건 보고서를 활용한 LLM 기반 시나리오 생성 방법론을 제안한다. 기존 시나리오 생성 방법이었던 수동 시나리오 생성의 한계를 LLM을 통하여 보완하였다. 보고서에 제시되지 않은 공격 흐름을 합리적으로 추론하여 시나리오를 생성함으로써, 사이버 복원력 측정 및 사이버 훈련 환경에서 다양하게 활용될 것으로 예상된다. 향후 연구에서는 제안한 방법론을 수행하여 시나리오를 생성한 뒤, 정량적 검증 방법을 통해 시나리오의 타당성을 검증할 계획이다.

Acknowledgement

본 연구는 2025년 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학사업 지원을 받아 수행되었음(2025-0-00040)

참고문헌

[1] R. Ross, V. Pillitteri, R. Graubart, D. Bodeau, and R. McQuaid, Developing Cyber Resilient Systems: A Systems Security Engineering Approach (NIST SP 800-160 Vol.2 Draft), Gaithersburg, MD, National Institute of Standards and Technology, 2019.

[2] D. S. W. Gould, G. De Ath, B. Carvell, and N. Pepper, AirTrafficGen: Configurable Air Traffic Scenario Generation with Large Language Models, arXiv preprint arXiv:2508.02269, 2025.

[3] Y. Zhang, T. Du, Y. Ma, X. Wang, Y. Xie, G. Yang, and E. C. Chang, "AttacKG+: Boosting attack graph construction with large language models," Computers & Security, vol. 150, p. 104220, 2025.

[4] M. M. Yamin, E. Hashmi, M. Ullah, and B. Katt, "Applications of LLMs for generating cyber security exercise scenarios," IEEE Access, 2024.

[5] D. Gosmar and D. A. Dahl, Hallucination Mitigation Using Agentic AI Natural Language-Based Frameworks, arXiv preprint arXiv:2501.13946, 2025.