Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.33, NO.3, Jun. 2023 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2023.33.3.527

IaC-VIMF: 사이버 공방 훈련을 위한 IaC 기반 가상 인프라 변이 생성 프레임워크*

노 주 영,[1†] 이 세 한,[1] 박 기 웅[2‡]
[1]세종대학교 시스템보안연구실 지능형드론 융합전공 (대학원생), [2]세종대학교 정보보호학과 (교수)

IaC-VIMF: IaC-Based Virtual Infrastructure Mutagenesis Framework for Cyber Defense Training*

Joo-Young Roh,[1†] Se-Han Lee,[1] Ki-Woong Park[2‡]
[1]SysCore Lab., Sejong University (Graduate Student),
[2]Dept. of Computer and Information Security, Sejong University (Professor)

요 약

사이버 침해사고 대응 능력을 갖춘 전문가의 양성을 위해 여러 기관에서 사이버 훈련장을 구축하여 사이버 방호 전략을 갖춘 보안 전문가를 양성하고 있지만, 기존 시스템에서는 가상훈련 시스템 내 자원의 한계, 시나리오 기반의 실습 콘텐츠 개발 및 운영, 비용적인 문제 등에 있어서 어려움을 겪는다. 이를 보완하기 위해 본 논문에서는 각 조직에 맞는 IT 인프라 환경에 대한 유사한 변이 환경을 제공하여 사이버 공방 훈련자가 다양한 경험을 축적할 수 있도록 하는 가상 인프라 변이 생성 프레임워크를 제안한다. 실험 및 평가를 위해 기존의 컨테이너를 IaC(Infrastructure-as-Code) 환경의 컨테이너로 전환하고 코드 내 변이할 수 있는 요소들을 데이터로 추출하여 자연어 처리 모델인 Word2Vec에 학습시켜 구성 데이터를 변이하여 새로운 코드를 생성하고 새로운 컨테이너 환경을 제시한다.

ABSTRACT

To develop experts capable of responding to cyber security incidents, numerous institutions have established cyber training facilities to cultivate security professionals equipped with effective defense strategies. However, these challenges such as limited resources, scenario-based content development, and cost constraints. To address these issues, this paper proposes a virtual infrastructure variation generation framework. It provides customized, diverse IT infrastructure environments for each organization, allowing cyber defense trainers to accumulate a wide range of experiences. By leveraging Infrastructure-as-Code (IaC) containers and employing Word2Vec, a natural language processing model, mutable code elements are extracted and trained, enabling the generation of new code and presenting novel container environments.

Keywords: Cyber Range, Cloud computing, System Orchestration, IaC(Infrastructure as code), Mutation

Received(03. 08. 2023), Modified(04. 19. 2023), Accepted(04. 19. 2023)

* 본 논문은 2022년도 한국정보보호학회 동계학술대회에서 발표한 우수논문을 개선 및 확장한 것임.

* 본 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원(IITP)의 정보통신방송기술 국제공동연구(Project No. RS-2022-00165794, 30%), 국방ICT융합연구(Project No. 2022-0-00701, 10%), 실감콘텐츠핵심기술개발(Project No. RS-2023-00228996, 10%), 정보통신방송혁신인재양성사업(Project No. 2021-0-01816, 10%) 및 한국연구재단(NRF) 중견후속연구사업(Project No. RS-2023-00208460, 40%)의 지원을 받아 수행된 연구임.

† 주저자, hpjoo718@gmail.com

‡ 교신저자, woongbak@sejong.ac.kr(Corresponding author)

I. 서 론

오늘날 정보화 시대는 IT 기술의 발전과 더불어 초연결사회(Hyper-Connected Society)의 형태로 변화하고 있으며 사람과 사람, 사람과 사물, 사물과 사물이 인터넷을 통해 강력하고 밀접한 네트워크를 구성하고 있다. 이러한 발전은 사물인터넷(Internet of Things) 기술의 발전을 토대로 다양한 분야(자율주행 자동차, 스마트 시티, 금융 등)에서 IT 기술에 대한 의존도가 높아지고 있다. 하지만 이와 더불어 사이버 공격의 빈도도 높아지고 있어 사이버 위협에 대한 대응이 필요하다[1].

대표적인 사물인터넷 공격 예시로는 미라이 봇넷(Mirai Botnet)[2]이 있으며, 이는 사물인터넷 기기 내 취약점을 공략하여 악성코드를 감염시키고, 해당 기기와 연결된 다양한 호스트 시스템을 장악하여 대규모 DDoS(Distributed Denial of Service) 공격을 수행하여 IT 인프라를 마비시켰던 사이버 공격 사례 중 하나이다. 또한, 현재는 데이터의 무분별한 활용이나 정보의 공개로 인해 해킹, 위협 등의 사이버 공격이 증가하고 있어, 정부와 기업뿐만 아니라 개인도 피해를 받고 있다. 이에 정부와 기업은 IT 자산을 보호하기 위해 다양한 사이버 공격에 대응하기 위한 사이버 보호 전략을 수립하고, 자체 보안 전문 인력을 양성하기 위해 사이버 공방 훈련 인프라를 구축하고 있다[3].

현재 사이버 공방 훈련장은 일반적으로 여러 사이버 공격 시나리오를 바탕으로 문제를 제시하고 이를 해결하는 방안으로 진행되고 있다[4]. 그러나 이 방법은 새로운 사이버 위협 도구 및 APT(Advanced Persistent Threat) 공격, 인공지능을 이용한 자동화 된 공격 등에 빠르게 대응하기 어려운 단점이 존재하며, 현재 급변하는 IT 기술에 빠르게 대응하기에는 비용 효율적이지 못하다는 단점이 존재한다.

또한, 사이버 공방 훈련을 위한 시나리오의 경우에는 훈련장 운영을 위한 다양한 요구 조건 및 위험부담이 따르기에 사이버 공방 훈련자가 독단적으로는 사용하는 것이 불가하며, 운영 환경의 요건 및 사이버 공방 훈련자의 지식과 기술 수준을 고려하지 않은 공방 훈련 설계는 난이도 조절 실패로 이어지게 되므로 시나리오로서는 활용이 불가하다. 이에 현재 IT 산업 기술 및 사이버 공격에 대한 최신 동향을 바탕으로 능동적인 사이버 공격 및 방어 훈련 시나리오 생성과 사이버 공방 훈련 목적에 따른 체계적인 환경 구성 방안을 갖추어야 할 필요성이 존재한다[5].

본 논문에서는 이와 같은 문제점 및 필요성을 해결하기 위해 다양한 IT 인프라에 대한 사이버 공방 훈련 인프라를 능동적으로 생성하고 각 조직에 맞는 훈련 체계를 구성하는 데 도움을 주기 위해 사이버 공방 훈련을 위한 가상 인프라 변이 생성 프레임워크

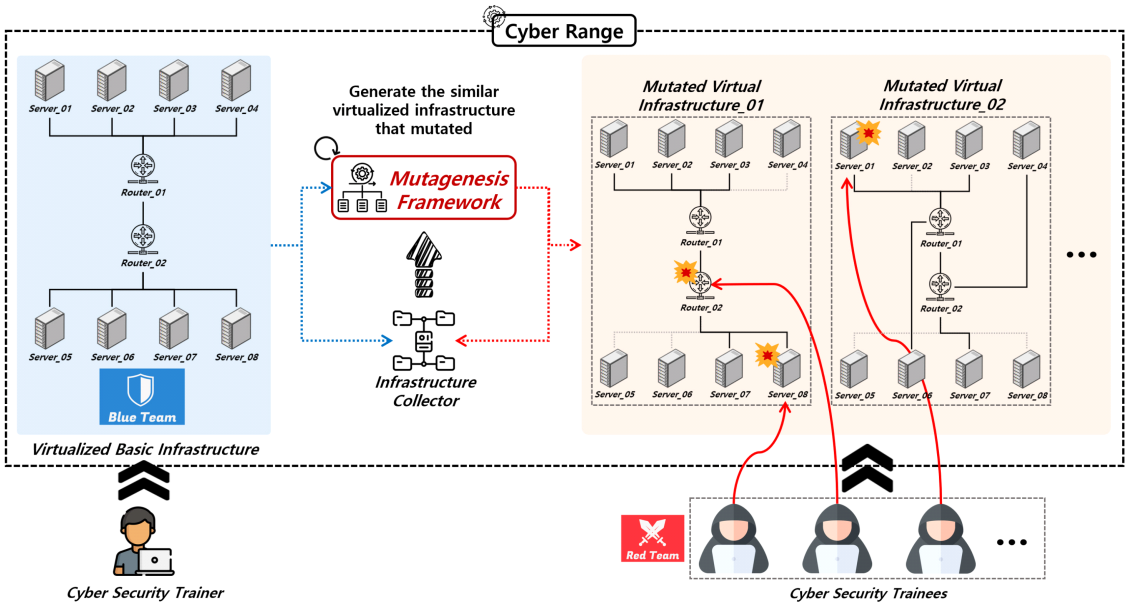

Fig. 1. An overview of virtualized infrastructure mutagenesis framework in cyber range

를 제안한다. 제안하는 프레임워크를 통해 다양한 사이버 공격 유형에 대한 선제적 조치 및 방어 훈련을 수행함으로써 실질적인 사이버 보안 인력에 대한 능동적인 정보보호 수행 능력을 향상하고 사이버 공격에 대한 탐지 및 분석, 대응, 예방 등의 체계적인 사이버 방호 전략을 수립할 수 있다. 또한 사이버 공방 훈련을 순환적으로 실행함으로써 정보보안 전문 인력이 재훈련을 할 수 있는 기회를 제공할 수 있다. 이를 위해 기존 사이버 공방 훈련 인프라를 IaC 환경으로 변환하고 인프라 코드 변이에 있어 딥러닝 엔진을 활용하여 사람이 아닌 컴퓨터가 스스로 학습하여 새로운 코드를 변이 생성하고 새로운 인프라의 IaC 환경구성을 자동화할 수 있도록 프레임워크를 설계하였다.

본 논문의 구성은 다음과 같다. Ⅱ장에서는 현재 사용되고 있는 사이버 공방 훈련장 운영 시스템에 대하여 설명하고 본 논문에서 제안하는 가상 인프라 변이 생성 프레임워크와 관련하여 다양한 시스템 내 구성요소와 유사한 변이 요소 생성에 관한 관련 연구를 기술한다. Ⅲ장에서는 본 논문에서 제안하는 프레임워크에 대한 기본적인 설계 및 IT 시스템 인프라 생성 방안, 해당 인프라에 대한 변이 생성을 데이터 흐름도를 통해 자세히 설명한다. Ⅳ장에서는 본 논문에서 제안하는 프레임워크의 실용성 평가를 위해 테스트 케이스를 생성하고 변이된 데이터의 유사도 결과에 관해 설명한다. 마지막으로 Ⅴ장에서는 결론 및 향후 연구 계획에 대하여 설명한다.

Ⅱ. 관련 연구

2.1 기존 사이버 공방 훈련 운영 시스템

사이버 공방 훈련은 정부 또는 기업에서 운영 중인 IT 인프라 시스템과 유사한 환경에서 사이버 공격과 방어 기술을 습득할 수 있는 경험 중심의 훈련 방안 중 하나이며, 훈련 결과를 실제 환경에 반영하고 개선하여 보안 취약점을 찾아내기도 한다. 사이버 공방 훈련은 사이버전 및 사이버 침해위협 대응 기술 개발을 위한 가상 인프라 환경을 구축하여 실시한다. 이러한 공방 훈련을 통해 사이버 위협 대응 전문가 양성 및 공세적, 그리고 방어적 기술 습득에 중점을 두고 훈련을 수행할 수 있다[6]. 이러한 사이버 공방 훈련과 관련하여 국내외 민·관·군에서의 정보보안 전문가 양성을 위한 사례를 확인하였으며, 그 내용은 다음과 같다.

2.1.1 국내 사이버 공방 훈련 운영 시스템 현황

국내에서는 한국인터넷진흥원(KISA)에서 사이버 침해사고 및 위협 사례를 활용하여 '시큐리티짐(Security-Gym)'[7]을 운영 중이다. 대규모의 실 환경을 가상 사이버 환경으로 모방하여 기존 사건들을 시나리오화하여 보안 인력을 대상으로 기술 훈련부터 방어전략 수립까지 수준에 맞추어 교육을 진행한다.

또한, 국군 사이버 작전사령부에서는 '사이버 공방 훈련장 구축 사업'[8]을 통해 사이버전을 위한 전투준비태세 확립 및 적군에 대한 사이버 공간 대응 능력 향상과 전문 인력을 육성할 수 있도록 마련하였다. 사이버 공방 훈련에는 탐지, 대응, 분석, 예방이 순환적으로 수행되고 공방 훈련장에는 이와 같은 과정이 종합적으로 포함되어 있다. 군에서의 사이버 공방 훈련장은 폐쇄적인 내부 인트라넷 환경과 각종 군사 장비 등 특수한 환경을 반영한 실전 대응훈련을 진행하여 사이버전에 완벽 대비하고 있다.

마지막으로 국가보안기술연구소에서 운영하는 '사이버안전 훈련센터'[3]는 사이버 공방 훈련을 위한 사이버 보안 훈련 프로그램을 운영하며 매년 정보보안 핵심 인력을 양성하고 있다. 주로 정부와 공공기관의 정보보안 교육을 담당하고 있으며, 기존 침해사고 등의 위협 사례를 시나리오로 정의하고 학습하는 방식으로 사이버 공방 훈련장을 조성하고 경보 단계를 기준으로 사이버 공격에 대한 정의와 예방, 실시간 대응, 사후 대응 순으로 대응책을 마련하여 훈련을 시행하고 있다.

2.1.2 국외 사이버 공방 훈련 운영 시스템 현황

국외에서는 미국, 유럽 연합, 이스라엘 등에서 사이버 공방 훈련장을 구축하여 운영하고 있으며, 민·관·군에서 사이버전 및 사이버 방호 전략 수립을 위해 적극적으로 활용하고 있다.

美 국방부(DoD)에서는 'NCR(National Cyber Range)'[9] 을 운영하고 있으며, 이는 폐쇄된 환경의 대형 루프 시스템 구조를 갖춘 인터넷과 유사한 가상의 사이버 환경을 구축한다. 이를 통해 Zero-Day 취약점, 사이버 위협에 대한 공세적 기법 및 방어적 기법과 이를 활용한 사이버 보안 전략 수

립 등에 활용하고 있으며, 해당 시설은 군사적 훈련 목적뿐만 아니라 대학과 연구소에서 연구 목적으로도 활용하고 있다.

유럽 연합(EU)에서는 'Airbus'[10]로 명명된 사이버 공방 훈련 플랫폼을 운영 중이며, 이는 실제 시스템을 쉽게 모델링하고 사이버 공격을 비롯한 다양한 시나리오를 연출할 수 있다. 본 플랫폼에서는 가상화 및 하이브리드 환경에서 실제로 발생하는 사이버 공격 활동을 실시간으로 실행하거나 분석하여 사이버 보안 담당자에 대한 사이버 위협 대응 능력을 향상시킬 수 있으며, VM(Virtual Machine) 및 도커(Docker)를 활용하여 실제 물리적인 인프라 시스템과 가상 시스템, 그리고 다른 외부 장비들과의 인터페이스 공유가 가능하도록 구성되어 있다.

이스라엘의 경우에는 'Cyber Gym'[11]이라 명명된 교육 시설을 구축하여 군대 및 정보 기관, 민간 기업 등에 현실적인 사이버 보안 교육을 제공하고 있다. 각 조직에 맞는 특정 요구 사항을 토대로 다양한 사이버 위협 시나리오를 제공하고 있으며, 위협 분석, 침해사고 대응, 디지털 포렌식 등과 같은 다양한 교육 커리큘럼을 제공한다.

최근에는 인공지능을 활용한 사이버 보안 분야의 기술들도 늘어나고 있다. IT 기술이 점점 발전함에 따라 사이버 공격 또한 더욱 정교하고 다양해지면서 이에 대한 탐지, 방어, 예방 등을 위한 인공지능과 기계학습 기반의 훈련장도 활용되고 있다. 이 외에도 사이버 공방 훈련자들이 보안 침투 테스트 기술을 배울 수 있도록 취약한 환경을 가진 가상머신을 생성하는 'SecGen'[12] 이나, 'APG(Automatic Problem Generation)'[13] 등이 활용되고 있다.

2.2 시스템 내 구성요소 변이 관련 연구

본 절에서는 가상 인프라 변이 생성 프레임워크와 관련하여 시스템 내 다양한 구성요소 중 코드를 변이하는 관련 연구를 소개한다. 현재 코드 변이 기법은 개발자가 돌연변이 테스트를 위해 기존 코드를 변이하여 사용하거나 안티바이러스 및 보안 시스템 탐지 우회를 위해 바이너리 코드를 미묘하게 변경하여 사이버 공격자가 인프라 보안 제어 우회 및 관리자 권한 탈취를 목적으로 사용된다. 이와 관련된 연구 결과에 관한 내용은 다음과 같다.

2.2.1 MART

MART[14]는 LLVM(Low Level Virtual Machine) 기반의 비트 코드(Bit code)를 변이하여 생성된 돌연변이 코드에 대한 실험을 위한 목적으로 만들어진 도구로, 소프트웨어 결함을 찾는 일종의 화이트 박스 테스트 중 하나인 돌연변이 테스트를 수행한다. MART는 고급 프로그래밍 언어에 대한 돌연변이 생성도 가능하며, 중복되어 생성된 돌연변이를 제거하는 In-memory TCE 모듈과 메타 돌연변이 모듈들을 통해 돌연변이의 유형과 정도를 파일로 제작할 수 있으므로 이를 활용하여 사이버 위협 분석에 있어 실용적으로 사용할 수 있다.

Mart에서는 실제 소프트웨어 프로젝트에서 성공적으로 돌연변이를 생성하여 테스트한 사례가 존재하며, 추가적인 연산자 그룹 구현을 통해 더 다양한 프로그램 구문 요소를 처리할 예정이다. 이러한 MART는 소프트웨어 품질 향상을 위한 중요한 연구 사례 중 하나이다.

2.2.2 LittleDarwin

LittleDarwin[15]은 애자일 방법론(Agile Methodology)으로 개발되는 거대한 규모의 Java 소프트웨어 시스템에서 적용이 가능한 돌연변이 테스트 도구로, 소프트웨어 개발 방법론의 변화로 인해 자주 테스트를 수행하게 되는 요즘 추세에 맞추어 테스트 도구의 품질은 매우 중요하다. 이에 따라 소프트웨어 품질을 향상시키기 위한 하나의 방법으로 돌연변이 테스트가 연구되고 있으며, LittleDarwin은 소프트웨어의 특정 부분에 대한 변형 연산자를 적용하여 고차원 돌연변이, null 유형의 돌연변이, 조건문을 입력한 돌연변이 등 다양한 돌연변이를 생성할 수 있다. 이를 통해 변이 테스트 케이스를 생성하고 분석하는 기능을 제공한다. 또한, Python으로 작성되었으며, 오픈소스로 제공하고 있어 여러 시스템 환경과의 호환성이 뛰어나며 새로운 기능을 추가에 용이하다. 이러한 기능들은 LittleDarwin을 소프트웨어 개발의 초기 단계부터 테스트 도구로 적용할 수 있게 하며, 소프트웨어의 안정성과 신뢰성을 높임에 있어 큰 도움을 줄 수 있다.

2.2.3 Mutate-NLP

Mutate-NLP[16]는 자연어 처리 모델에 대한 돌연변이 생성 도구이다. 최근에는 자연어 처리 모델의 성능 향상이 크게 이루어져 왔지만[17], 이러한 모델들도 여전히 일부 데이터에 대해서는 부족한 성능을 보인다. 이를 해결하기 위해 돌연변이 테스트가 적용되고 있다. Mutate-NLP는 기존의 자연어 처리 모델에 변형 연산자를 적용하여 다양한 돌연변이를 생성하고, 이를 이용하여 모델의 성능을 평가하고 개선할 수 있다. 변형 연산자는 단어 대체, 문장 생성, 입력 길이 조정 등 다양하게 적용할 수 있으며, 이를 통해 다양한 유형의 돌연변이를 생성할 수 있다. 또한 Mutate-NLP는 다양한 자연어 처리 모델에 적용할 수 있으며[18], TensorFlow 및 PyTorch 같은 대표적인 딥러닝 프레임워크와 호환된다. 또한 공개 소프트웨어로 제공되어 커뮤니티 기반의 다양한 개발자들이 참여하여 새로운 변형 연산자나 모델에 대한 돌연변이 생성 기능을 추가할 수 있다. 이를 통해 자연어 처리 모델의 성능을 개선하며 추가적인 연구를 진행 중이다.

2.3 기존 연구와의 차이점

기존 변이에 관한 연구들은 경우 제어 연산자, 산술 연산자를 변이하여 수치의 변형을 일으키는 경우가 대부분이었다. 이는 텍스트를 value로 지닌 IaC(Infrastructure as Code) 코드 변이 생성에 적용하는데 제한적이고, 다양한 변이 값을 기대하기에는 한계점이 있었다. 최근 자연어 처리에 대한 관심과 딥러닝 기술의 발전으로 인해 ChatGPT, Copilot 등의 코드 자동 생성 엔진이 개발되면서 단어 유사도를 이용한 어절, 문장 변형 등으로 활용되고 있다. 이러한 딥러닝 엔진을 활용하여 본 논문에서는 IaC 코드의 변형을 통한 가상 인프라 변이 생성 프레임워크를 제안한다. 이는 기존의 단순한 연산에 대한 변이뿐만 아니라, 인공지능의 학습을 통한 새로운 단어 및 수치의 생성으로 이루어질 변이된 인프라를 기대할 수 있다.

III. 가상 인프라 변이 생성을 위한 변이 프레임워크

본 장에서는 능동적이고 효율적인 사이버 공방 훈련장 생성 및 관리를 위해 제안하는 가상 인프라 변이 생성을 위한 변이 프레임워크에 대해 설명한다. 본 논문에서 제안하는 프레임워크에서는 가상의 인프라를 IaC 코드로 환경을 구성하고 인프라를 관리하며 즉시 사용이 가능한 형태로 배치 및 배포하고, 필요할 시 즉시 사용이 가능하도록 가상 환경을 구축한다. IaC는 인프라의 설정을 코드로 작성하여 인프라의 생성, 수정, 삭제를 자동화하는 방법이다. 이를 활용하면 서버, 데이터베이스, 네트워크, 배포 프로세스, 테스트 등을 코드로 관리하고 배포할 수 있다. 이러한 IaC를 활용하는 데 있어 도움을 줄 수 있는 도구로는 앤서블(Ansible), 테라폼(Terraform), 셰프(Chef), 퍼펫(Puppet) 등이 있으며 이를 활용하여 운영 측면의 다양한 요소들을 코드로 대체할 수 있다.

본 논문에서 제안하는 프레임워크의 개념도는 Fig.2와 같으며, 그 설명은 다음과 같다.

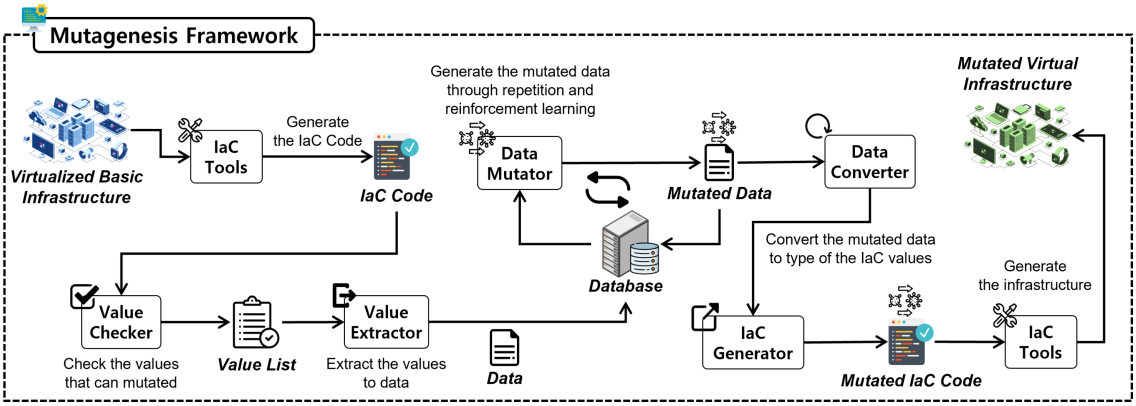

Fig. 2. A diagram of virtualized infrastructure mutagenesis framework

우선 사이버 공방 운영자는 변이를 가할 훈련장에 대하여 IaC 환경의 코드 파일을 작성한다. 작성한 파일은 프레임워크의 'Value Finder'를 통해서 변이가 가능한 요소들을 판단하고 'Value Extractor'를 이용하여 데이터를 추출하고 'Value Pre-Process'를 통해 Value를 'Value Mutator'가 이용이 가능한 데이터로 전처리한다. 전처리 된 데이터를 데이터베이스에 저장하고, 저장된 데이터들을 토대로 'Value Mutator'에 학습시킨다. 충분한 학습이 된 'Value Mutator'는 기존의 Value와 조건 파라미터에 의하여 새로운 변이 데이터를 생성하고 데이터베이스에 저장한다. 이를 반복 및 강화 학습하여 새로운 데이터를 세대를 거듭하여 생성해낸다. 생성된 변이 데이터를 'Data Process'에 전달하여 변이 Value의 형식으로 처리하고 이를 'IaC Generator'가 새로운 IaC 코드 파일을 생성한다. 생성한 파일을 토대로 IaC 도구를 이용하여 기존 인프라와 대조되는 새로운 가상 인프라를 생성한다. 생성한 새로운 IaC 파일을 부모로 삼아 새로운 변이 데이터를 생성하여 순환하는 구조로 프레임워크를 운영할 수 있다.

제안하는 프레임워크를 사용하면 IaC 기반의 IT 인프라에 대한 정보보호 업무를 수행하는 담당자가 실제 사용 중인 인프라 환경과 유사한 새로운 가상 환경에서 사이버 공방 훈련을 진행할 수 있도록 도움을 줄 수 있다.

IV. 실험 및 평가

본 논문에서 제안하는 가상 인프라 변이 생성 프레임워크에 대한 실험 환경은 Table.1.과 같다.

Table 1. An environment for testing

Environment	Environment Config.
Docker Server	Ubuntu 20.04
	Docker v20.10.23
	Docker Compose v2.6.1
Google Colab	Python 3.8
	RAM: 12.7GB
	Disk: 107GB

본 실험에서는 변이 생성을 위해 Gensim[19]의 Word2Vec[20] 자연어 처리(Natural Language Processing) 모델을 사용하였다. Word2Vec 모델을 이용하여 전처리 된 기존 인프라 데이터들을 학습시키고 변이 대상 데이터를 단어 유사도를 기반으로 의미 없는 데이터를 생성하는 것이 아닌 단어 간의 의미를 학습해 기존 가상 인프라 구성 내 코드의 데이터와 유사한 변이 데이터를 생성하여 코드를 제작하였다.

Fig.3은 기존 상용화된 IaC 도구인 테라폼(Terraform)을 이용하여 간단한 AWS 클라우드 인프라를 구성하고 이를 논문에서 제시한 프레임워크를 적용하여 변이한 예시이다.

본 실험에서는 변이로 확인할 수 있는 네트워크 구조와 데이터베이스의 버전 업, 다운을 확인하였으며 실험을 위해 기존 인프라 IaC 코드에서 변이가 가능한 데이터들을 추출하고 이를 Word2Vec 엔진을 활용할 수 있는 데이터의 형식으로 전처리하여 데이터 세트를 구축하였으며, 이를 Word2Vec 모델에 학습하는 방식으로 실험을 진행하였다. 본 실험에서 활용된 Word2Vec 엔진의 반복 학습에 따라 데이터

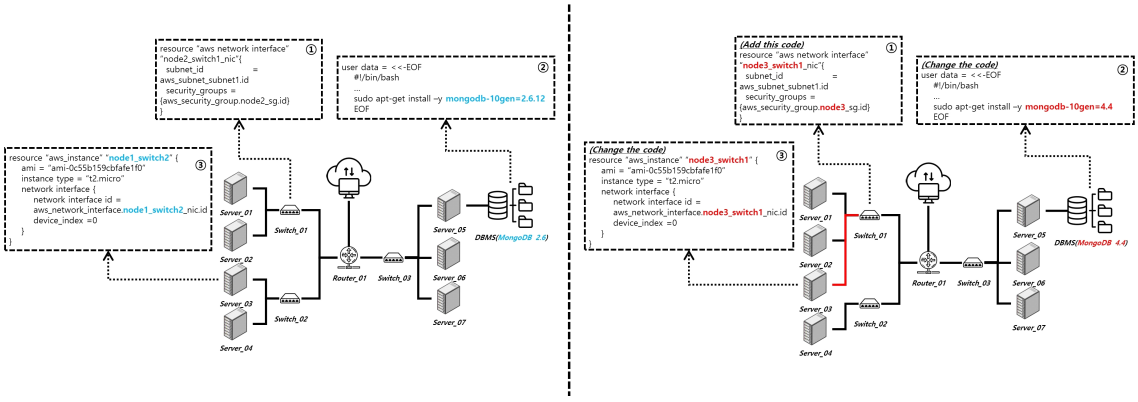

Fig. 3. An example of IaC code variation when using the proposed framework

의 연관성 및 유사도의 변화를 측정하여 데이터가 활용할 데이터가 기존 데이터에 대해 학습을 거듭해 나가며 연관되고 의미 있는 데이터만 분류가 되어 단어를 추천해주는 것을 확인할 수 있다.

Fig.4 은 Word2Vec 엔진의 반복 학습에 따른 데이터의 유사도 변화를 t-SNE[21] 알고리즘을 사용하여 차원 축소 후 시각화하여 나타낸 것이다. X축과 Y축은 데이터 간의 거리에 대하여 나타내며 차원 축소로 인하여 정확한 수치에 대하여 시각화되지는 않으나 학습의 반복 횟수가 증가함에 따라 단어 기억을 개선하고 관련 단어 간의 의미론적 연결을 강화하게 되었다는 것을 그래프를 통해 알 수 있다. 또한 학습의 반복을 통해 모델이 강화되면서 분류되지 않던 의미 없는 단어의 중앙 밀집형 모델의 모형이 알고리즘을 통해 의미 있는 단어의 모임인 군집이 형성된 모형이 된 것을 알 수 있다. 이는 학습을 통해 강화되는 Word2Vec 엔진에서의 단어 유사도 알고리즘을 통해 기존 데이터와 유사한 의미 있는 변이 데이터를 생성하고 이를 IaC 코드에 적용하여 새로운 인프라를 생성하는 데 있어 활용할 수 있다.

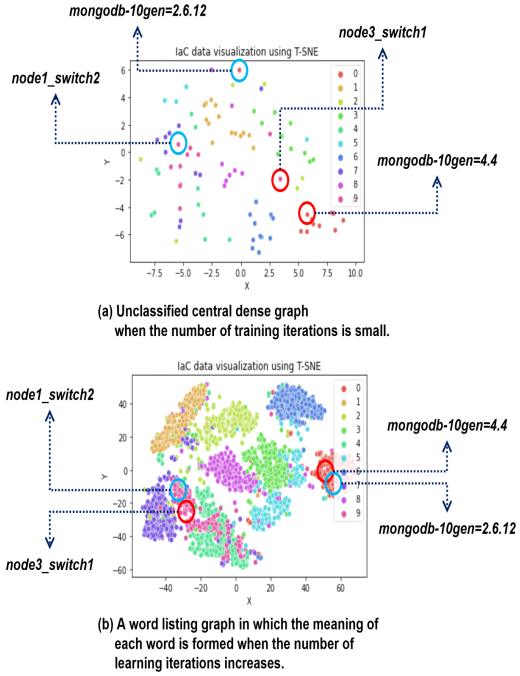

Fig. 4. A visualization of data similarity change according to the number of repetitions of learning using t-SNE

V. 결 론

오늘날 정부와 주요 기업에서는 각 조직에 맞는 효과적인 사이버 공방 훈련을 위한 가상 인프라 구축에 있어서 많은 인력과 자원을 투자하고 있다. 그러나 제한된 자원으로 인한 환경 구축 문제와 급변하는 사이버 공격에 실시간으로 대응할 수 있는 인프라가 충분히 구축되지 않아 훈련에 어려움을 겪고 있다.

본 논문에서는 이러한 어려움을 해결하고자 가상 인프라 변이 생성을 위한 프레임워크를 제안하였고, 이를 재현하고자 실험하였다. 본 실험에서의 한계점으로서 직렬화된 데이터 코드(YAML, XML, JSON 등)에 대한 변이 생성 결과를 확인하였지만, 논리적인 고차원의 코드는 의존성에 대한 문제를 해결할 필요성이 존재함을 확인할 수 있었다. 또한, Word2Vec 엔진에서 측정하는 단어 유사도의 수치를 시각화하는 단계에서 실제 Word2Vec 엔진에서 측정된 유사도와 시각화된 그래프의 수치 차이가 있었는데, 이는 차원 축소를 이용한 t-SNE 알고리즘을 이용하여 고차원 데이터를 임베딩(Embedding)하는 과정에서 생기는 괴리로 확인되었다. 따라서 Word2Vec 엔진에서의 단어 유사도를 수치화 및 시각화하였을 때, 사용자가 대략적인 수치를 통해 단어와 단어 사이의 관계를 직관적으로 확인하는 데 적합하지만, 더욱 면밀한 유사도를 가늠하기에는 부적합하였다[22].

그러나 IaC 환경, 클라우드 네이티브 환경에서처럼 운영 측면이 모두 코드로 대체될 수 있는 환경 구축에서는 본 논문에서 제안한 프레임워크가 적용이 가능하다는 것을 실험을 통해 확인할 수 있었다. 이는 급변하는 사이버 공격에 대응하는 훈련 프로그램의 인프라 구성을 자동화하기 적합하고 기존 사이버 공방 훈련 인프라의 모델을 변이 생성 자동화하여 활용이 가능할 것으로 기대된다.

향후 연구를 통해 본 논문에서 제안한 프레임워크에서 코드 데이터 변이에 적합한 딥러닝(Deep Learning) 엔진을 적용하여 기존 코드 데이터를 학습시키고 새로운 인프라 및 기존 인프라보다 강화된 환경을 구축 및 설계하여 효율적이고 능동적인 IaC 환경의 클라우드 오케스트레이션 도구 및 프레임워크를 적용한 자동화 플랫폼을 개발하고자 한다. 또한 제안한 프레임워크를 현재 운영되고 있는 사이버 공방 훈련 프로그램에 적용하여 기존 프로그램의 수정, 개선 사항을 확인하고 도출하여 이를 개선하여 적용

했을 때의 교육 효과 및 운영의 편리성을 측정하고 검토하는 연구를 진행하고자 한다.

References

[1] Yuchen Yang, Long fei Wu, Guisheng Yin, Lijie Li and Hongbin Zhao, "A Survey on Security and Privacy Issues in Internet-of-Things," IEEE Internet of Things Journal, vol. 4, no. 5, pp. 1250-1258, Oct. 2017.

[2] Manos Antonakakis et al., "Understanding the Mirai Botnet," The Proceedings of the 26th USENIX Security Symposium, pp. 1093-1110, Aug. 2017.

[3] Younghan Choi et al., "Design and Implemntation of Cyber Range for Cyber Defense Exercise Based on Cyber Crisis Alert," Journal of the Korea Institute of Information Security & Cryptology, vol. 30, no. 5, pp. 805-821, Oct. 2020.

[4] Razvan Beuran, Ken-ichi Chinen, Yasuo Tan and Yoichi Shinoda, "Towards Effective Cybersecurity Education and Training," Research Report, JAIST, pp. 1-16, Oct. 2016.

[5] Vincent E. Urias, William M.S. Stout, Brian Van Leeuwen and Han Lin, "Cyber Range Infrastructure Limitations and Needs of Tomorrow: A Position Paper," 2018 International Carnahan Conference on Security Technology (ICCST), pp. 1-5, Oct. 2018.

[6] Dorothy E. Denning, "INFORMATION WARFARE AND SECURITY," Information warfare and security vol. 4, New York: Addison-Wesley, 1999.

[7] Security-Gym, https://academy.kisa.or.kr

[8] J. H. Yu, K. J. Koo, I. K. Kim and D.S. Moon, "Technological Trends in Intelligent Cyber Range," Electronics and Telecommunications Trends, ETRI, vol. 37, no. 4, pp. 36-45, Aug. 2022.

[9] Bernard Ferguson, Anne Tall and Denise Olsen, "National Cyber Range Overview," 2014 IEEE Military Communications Conference, pp. 123-128, Oct. 2014.

[10] CyberRange-Airbus, "CyberSecurity,EU", https://airbus-cyber-security.com/products-andservices/prevent/cyberrange/, June 2023.

[11] Meir Kalech, "Cyber-attack detection in SCADA systems using temporal pattern recognition techniques," Computers & Security, vol. 84, pp. 225-238, July 2019.

[12] Z. Cliffe Schreuders et al., "Security Scenario Generator (SecGen): A Framework for Generating Randomly Vulnerable Rich-scenario VMs for Learning Computer Security and Hosting CTF Events," 2017 USENIX Workshop on Advances in Security Education, Aug. 2017.

[13] Jonathan Burket et al., "Automatic Problem Generation for Caputre-the-Flag Competitions," 2015 USENIX Summit on Gaming, Games, and Gamification in Security Education, Aug. 2015.

[14] Thierry Titcheu Chekam, Mike Papadakis and Yves Le Traon, "Mart: A Mutant Generation Tool for LLVM," Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, pp. 1080-1084, Aug. 2019.

[15] Ali Parsai, Alessandro Murgia and Serge Demeyer, "LittleDarwin: A Feature-Rich and Extensible Mutation

Testing Framework for Large and Complex Java Systems," Fundamentals of Software Engineering, 2017, vol. 10522, pp. 148-163, Oct. 2017.
〔16〕 github, "Mutate-NLP", https://github.com/infinitylogesh/mutate, June 2023.
〔17〕 Kang Min Yoo et al., "GPT3Mix: Leveraging Large-scale Language Models for Text Augmentation," Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 2225-2239, Nov. 2021.
〔18〕 Varun Kumar, Ashutosh Choudhary and Eunah Cho, "Data Augmentation using Pre-trained Transformer Models," Proceedings of the 2nd Workshop on Life-long Learning for Spoken Language Systems, pp. 18-26, Dec. 2020.
〔19〕 Mofiz Mojib Haider et al., "Automatic Text Summarization Using Gensim Word2Vec and K-Means Clustering Algorithm," 2020 IEEE Region 10 Symposium (TENSYMP), pp. 283-286, Jun. 2020.
〔20〕 Kenneth Ward Church, "Word2Vec," Natural Language Engineering, vol. 23, no. 1, pp. 155-162, Jan. 2017.
〔21〕 Laurens van der Maaten and Geoffrey Hinton, "Visualizing Data using t-SNE," Journal of Machine Learning Research, vol. 9, no. 86, pp. 2579-2605, 2008.
〔22〕 Hyungsuc Kang and Janghoon Yang, "Analyzing Semantic Relations of Word Vectors trained by The Word2vec Model," Journal of KIISE, vol. 46, no. 10, pp. 1088-1093, Oct. 2019.

〈저자소개〉

노 주 영 (Joo-Young Roh) 학생회원
2022년 9월~현재: 세종대학교 일반대학원 정보보호학과 석사과정
〈관심분야〉 클라우드 시스템 보안, 시스템 오케스트레이션, 시스템 모니터링

이 세 한 (Se-Han Lee) 학생회원
2017년 2월: 한국기술교육대학교 컴퓨터공학부 학사 졸업
2023년 2월: 세종대학교 일반대학원 정보보호학과 석사 졸업
2023년 3월~현재: 세종대학교 일반대학원 정보보호학과 박사과정
〈관심분야〉 악성코드 분석, 사용자 인증, 공격 시나리오, 임베디드 시스템 보안

박 기 웅 (Ki-Woong Park) 종신회원
연세대학교 Computer Science 학사
KAIST Electrical Engineering 석사
KAIST Electrical Engineering 박사
2009년 10월: Microsoft Research, Graduate Research Fellow
2012년 8월: 국가보안기술연구소 연구원
2016년 8월: 대전대학교 정보보안학과 교수
2016년 9월~현재: 세종대학교 정보보호학과 교수
〈관심분야〉 클라우드 시스템 보안, 초고속 보안 시스템, 시스템 인스펙션, 디지털 포렌식