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.2023.33.3.549

5G 모바일 에지 컴퓨팅에서 빅데이터 분석 기능에 대한 데이터 오염 공격 탐지 성능 향상을 위한 연구*

옥 지 원,[1†] 노 현,[1] 임 연 섭,[2] 김 성 민[2‡]
[1,2]성신여자대학교 (대학원생, 교수)

A Study on Improving Data Poisoning Attack Detection against Network Data Analytics Function in 5G Mobile Edge Computing*

Ji-won Ock,[1†] Hyeon No,[1] Yeon-sup Lim,[2] Seong-min Kim[2‡]
[1,2]Sungshin Women's University (Graduate student, Professor)

요 약

5G 네트워크의 핵심 기술로 모바일 에지 컴퓨팅(Mobile Edge Computing, MEC)이 주목받음에 따라, 모바일 사용자의 데이터를 기반으로 한 5G 네트워크 기반 에지 AI 기술이 최근 다양한 분야에서 이용되고 있다. 하지만, 전통적인 인공지능 보안에서와 마찬가지로, 에지 AI 핵심 기능을 담당하는 코어망 내 표준 5G 네트워크 기능들에 대한 적대적 교란이 발생할 가능성이 존재한다. 더불어, 3GPP에서 정의한 5G 표준 내 Standalone 모드의 MEC 환경에서 발생할 수 있는 데이터 오염 공격은 기존 LTE망 대비 현재 연구가 미비한 실정이다. 본 연구에서는 5G에서 에지 AI의 핵심 기능을 담당하는 네트워크 기능인 NWDAF를 활용하는 MEC 환경에 대한 위협 모델을 탐구하고, 일부 개념 증명으로써 Leaf NWDAF에 대한 데이터 오염 공격 탐지 성능을 향상시키기 위한 특징 선택 방법을 제안한다. 제안한 방법론을 통해, NWDAF에서의 Slowloris 공격 기반 데이터 오염 공격에 대해 최대 94.9%의 탐지율을 달성하였다.

ABSTRACT

As mobile edge computing (MEC) is gaining attention as a core technology of 5G networks, edge AI technology of 5G network environment based on mobile user data is recently being used in various fields. However, as in traditional AI security, there is a possibility of adversarial interference of standard 5G network functions within the core network responsible for edge AI core functions. In addition, research on data poisoning attacks that can occur in the MEC environment of standalone mode defined in 5G standards by 3GPP is currently insufficient compared to existing LTE networks. In this study, we explore the threat model for the MEC environment using NWDAF, a network function that is responsible for the core function of edge AI in 5G, and propose a feature selection method to improve the performance of detecting data poisoning attacks for Leaf NWDAF as some proof of concept. Through the proposed methodology, we achieved a maximum detection rate of 94.9% for Slowloris attack-based data poisoning attacks in NWDAF.

Keywords: 5G Network, Data Poisoning Attack, Feature Selection, NWDAF, MEC

Received(03. 16. 2023), Modified(04. 12. 2023), Accepted(04. 13. 2023)

* 본 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2021R1G1A100632611, NRF-2022R1G1A1006174)과 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0 사업(IITP-2022-RS-2022-00156310)의 연구결과로 수행되었음.

† 주저자, 220224011@sungshin.ac.kr
‡ 교신저자, sm.kim@sungshin.ac.kr(Corresponding author)

I. 서 론

모바일 에지 컴퓨팅(Mobile Edge Computing, MEC)은 5G 네트워크에서 데이터를 수집한 단말 근처에서 바로 데이터를 처리하고 연산을 적용하는 기술이다. 이러한 MEC에 기반하여 데이터 수집 지점에서 바로 인공지능 학습 및 추론을 실행하는 에지 AI 기술이 제안되고 있다. 에지 AI 기술은 인공지능 기술에 에지 컴퓨팅을 적용함을 통해 통신 및 처리 지연의 감소, 효율적인 핸드오버 등의 장점을 제공하며[1], 이러한 장점을 바탕으로 사물인터넷, 스마트 그리드, 자율 주행과 같은 5G 킬러 콘텐츠 및 서비스의 핵심 기술로 주목받고 있다. 최근 국내에서도 지능형 에지 네트워크 플랫폼 구축을 통해 차세대 분산 지능 인프라 환경을 제공하기 위한 연구가 수행되고 있다[2].

에지 AI 기술의 성능 및 보안성을 평가하고 개선 및 고도화하기 위해서는 5G 네트워크 인프라에 대한 이해가 필요하다. 5G 네트워크는 소프트웨어 정의 네트워크 (Software Defined Network, SDN) 및 네트워크 기능 가상화(Network Function Virtualization, NFV) 기술을 바탕으로 한 네트워크 구성요소의 소프트웨어화를 통해 다양한 서비스 모델을 탄력적으로 제공할 수 있다[3]. 또한, 물리적 네트워크를 여러 가상 네트워크로 분할하고 독립적인 네트워크 자원을 확보할 수 있는 네트워크 슬라이싱을 통해 서로 다른 응용 별로 격리성을 보장하는 여러 개의 전용망을 만들 수 있다. 이와 같은 기능들은 기존 LTE망과 혼용되어 사용되던 5G 코어망과 무선망에 대한 SA(Standalone) 방식으로의 전환이 이루어지고 있다.

SA 방식과 기존 혼용 방식(Non-Standalone, NSA)의 특징적인 차이점으로는 새롭게 추가된 네트워크 기능(Network Function)의 등장이 있다. 이동통신 시스템 표준화 기구 3GPP(3rd Generation Partnership Project) Rel-16에 따르면 [4], 5G 네트워크의 자동화를 위한 NWDAF(Network Data Analytics Function), 네트워크 기능간 상호 연동을 위한 NRF(Network Repository Function), 최적의 네트워크 슬라이스를 선택하도록 정보를 제공해주는 NSSF(Network Slicing Selection Function) 등이 이에 해당한다. 이러한 네트워크 기능은 5G 네트워크 기반 에지 AI의 핵심적인 역할을 담당하며, 실시간 네트워크 자동 탐지, 인증 등의 지능형 보안 솔루션을 제공한다[5].

전통적인 인공지능 분야에서의 보안 이슈와 마찬가지로, 5G 네트워크에서의 에지 AI에서 사용되는 모델에 대한 적대적 공격은 고려해야 할 중요한 요소이다. 그중에서도 데이터 오염 공격(data poisoning attack)은 MEC와 같은 분산 인프라 환경에서 빈번하게 발생하며, MEC와 중앙집중형 클라우드가 협업하는 연합학습 시나리오에서는 모델의 성능에 직접적인 영향을 끼친다. 하지만 LTE 망에서의 데이터 오염 공격에 대한 다양한 연구[6, 7]에 비교해서 5G SA 망에서의 연구는 부족한 실정이다. 예를 들어, 선행연구[6]의 경우 RAN(Radio Access Network)에 대한 오염 공격만을 대상으로 하였으며 NWDAF, NRF와 같은 코어망 내 네트워크 기능에 대한 데이터 오염 공격 시나리오는 충분히 탐구가 이루어지지 않았다. 또한, LTE와 5G SA 모드는 다른 네트워크 프로세스를 가지고 있어 4G에서의 오염 공격 탐지 기술이 5G 환경에서 유효하지 않다.

데이터 오염 공격의 구체적인 대상에는 5G 네트워크 내 물리적 또는 가상 엔티티의 액세스 권한을 제어하기 위한 AI 기반 인증 및 권한 서비스가 있다. 5G 네트워크에서의 스몰 셀 밀도화로 인해 빈번한 인증이 발생하기 때문에, 짧은 지연시간을 가지는 효율적인 인증 메커니즘을 위해 AI를 사용한다[8]. 이때, 5G 네트워크 통신 중의 네트워크 패킷 등 데이터들에 대한 오염 공격 가능성이 존재한다[9].

본 연구에서는 5G SA 모드에서 에지 AI 기술에 관여하는 코어망 내 표준 네트워크 기능에 대한 데이터 오염 공격을 탐구한다. 공격 시나리오 도출을 위해, 5G 네트워크에서 AI와 결합할 수 있는 5G MEC 아키텍처 내 네트워크 기능을 분석하고 공격 표면을 분석한다. 이때, NWDAF 내 학습 과정에서 데이터 오염 공격의 예시로, 패킷 헤더에 대한 오염을 통한 서비스 거부 공격인 Slowloris 공격을 활용하였다. 또한, 5G MEC환경에서 분산 NWDAF 아키텍처 관련 연구[10]의 성능 검증을 바탕으로 데이터 오염 위협 모델에 대해 탐구하고, 제안한 배치 시나리오의 일부 개념 증명으로서 leaf NWDAF에서의 데이터 오염 공격 탐지 성능을 향상하기 위한 특징 선택(feature selection) 방법을 제안한다.

제안한 방법론의 성능 평가는 5G 네트워크 공개 데이터 셋인 5G-NIDD[11]을 활용하여 진행하였다. 결정트리, 선형회귀, SVM의 특징 선택 방법에 따른

분류 성능을 조사하고 최적의 특징 선택의 필요성을 분석하였다. 제안한 특징 선택 방법을 적용한 실험 결과, NWDAF의 Slowloris 공격 유형의 데이터 오염 공격을 최대 94% 탐지 가능함을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 5G 네트워크의 NWDAF와 데이터 오염 공격에 관한 배경 지식을 설명한다. 3장에서는 5G 네트워크의 데이터 오염 공격과 MEC 환경에서의 NWDAF 아키텍처 관련 연구를 분석한다. 4장에서는 5G MEC 환경 Slowloris 공격 시나리오에서 데이터 오염에 따른 특징 선택을 포함한 NWDAF 위협 모델을 제시하며, 5장에서 5G Slowloris 공격 데이터 셋 분석 및 특징 선택 여부에 따른 분류 실험을 수행한다.

II. 배경 지식

2.1 5G 네트워크 기능

3GPP에서 정의한 5G 표준 내 시스템 기준 구조(Reference Architecture)에 따르면, 5G 서비스 기반 아키텍처의 주요 핵심 구성 요소로 NWDAF 및 NRF가 새로운 네트워크 기능으로 추가되었다. NWDAF는 5G 네트워크의 자동화 및 지능화를 위한 코어 네트워크 시스템 내 새로운 표준 기능이다. 인공지능 알고리즘을 사용하여 분석을 도출하고 이러한 분석을 다른 네트워크 기능에 제공한다. 구체적으로, 3GPP Rel-17에서는 인공지능 모델 학습 및 학습된 모델을 NF 사용자에게 제공하기 위해 NWDAF 내에 MTLF(Model Training Logical Function)와 관련된 인터페이스를 정의한다[12]. 현재 규격은 AI/ML을 위한 데이터 수집 인터페이스 및 절차에 관해서만 규정하고 있으며, 세부적인 알고리즘이나 구체적 적용 조건에 관해서는 규정하지 않고 있다.

NRF는 네트워크 기능 서비스 프레임워크 구성 요소로, 변화하는 5G 코어 네트워크 기능의 서비스 상태 모니터링과 연동 정보 관리를 통해 이들 간 상호 연동을 지원한다. 5G 코어 네트워크에서 사용 가능한 네트워크 기능 인스턴스 및 지원되는 서비스의 프로필을 유지하고, 다른 인스턴스가 특정 유형의 새로운 인스턴스의 NRF 등록을 구독하고 알림을 받도록 허용한다. 또한, 각 인스턴스에서 NF 발견 요청(Discovery Requests)을 수신하고 특정 기준을 충족하는 사용 가능한 인스턴스의 정보를 제공하는 서비스 검색 기능을 지원한다[13].

2.2 이동통신망에서의 데이터 오염 공격

데이터 오염 공격이란, 공격자가 학습 단계에서 잘못된 학습데이터를 제공하거나 학습데이터에 노이즈를 추가하는 등 학습 알고리즘을 변조하여 학습 결과에 악영향을 미치는 것을 말한다. 변조할 학습데이터 표본을 선택하기 위해, 공격자는 먼저 유추된 대리 모델을 통해 표본을 실행한 다음 레이블을 변경하고 딥러닝의 결과가 대리 모델에서의 기준 영역에서 멀리 떨어져 있는 경우 잘못된 레이블이 지정된 표본을 공격할 대상 분류기로 학습데이터로 전송하며 오염 공격이 이루어진다[14].

선행 연구[9]에서는 인지 무선 통신 네트워크의 스펙트럼 데이터 오염 공격을 다루고 있다. 공격자가 채널이 유휴상태일 때 짧은 시간 동안 오염된 스펙트럼 데이터를 전송하고, 인공지능 및 기계 학습 모델을 재학습시킬 때 오염 데이터를 이용할 수 있게 한다. 결과적으로, 오염된 모델은 채널이 비어 있을 때도 채널이 사용 중이라고 판단하도록 하여 데이터를 전송하지 않는다는 잘못된 결정을 내리도록 함으로써 데이터 오염 공격이 이루어진다. 공격자들은 심층 신경망을 사용한 오염 공격으로 데이터 및 제어 평면 통신을 모두 조작하여 결과적으로 기존 사용자와 5G 네트워크 간의 효율적인 스펙트럼 공유를 막을 수 있으므로 시스템이 데이터 오염 공격에 노출된다[14].

2.3 Slowloris 공격

Slowloris(Slow HTTP Header DoS) 공격이란, HTTP 헤더의 정보를 비정상적으로 조작하여 웹서버가 완전한 헤더 정보가 올 때까지 기다리도록 하는 서비스 거부 공격이다[15]. 이때, 웹서버가 연결 상태를 유지할 만큼의 가용자원은 한계가 있으므로 임계치를 넘어가면 다른 정상적인 접근을 거부하며 공격이 이루어진다. HTTP 프로토콜에서는 헤더의 끝을 '/r/n'이라는 개행문자로 구분을 하는데, 공격자는 마지막 개행문자를 보내지 않고 지속해서 의미 없는 변수를 추가하며 연결을 유지하기 위한 가용자원을 소진하게 만들며 공격이 동작한다.

III. 관련 연구

Chafika Benzaid et al.[9]의 선행 연구는 인공지능이 5G 네트워크 보안에 어떤 영향을 미칠 수 있는지 탐구하였다. 5G 네트워크에서 모든 인공지능 위협을 해결하기 위한 올인원 솔루션이 없어 부분적으로 나누어 방어기법이 적용되어야 함을 강조한다. 이에 따라 5G 네트워크에서 인공지능 및 기계 학습을 위한 ITU-T 통합 아키텍처를 나타내고, 이 아키텍처에서의 데이터 오염 공격, 회피 공격, API 기반 공격에 대한 공격 표면을 지정한다. 또한, 각 표면에서 활용될 수 있는 방어기법도 제시한다. 하지만 제안된 공격 및 방어기법은 일반적 관점에서 데이터 주입(injection), 조작(manipulation), 논리 손상(logic corruption) 공격만을 다루고 있어 5G 네트워크에 특정된 것이라고 보기 어렵고, 담론적인 기법이라는 한계점이 있다. 또한, 제시한 아키텍처 역시 5G 네트워크 기능들이 데이터들을 어떻게 수집, 전처리 및 송수신하는지 등의 프로세스에 대한 고려가 충분히 이루어지지 않았다.

Yalin E et al.[14]의 선행 연구는 5G 시스템에서의 무선 통신에 대해 적대적 기계 학습의 새로운 공격 표면과 해당 공격을 식별하며 적대적인 기계 학습에 대한 5G 시스템의 주요 취약성을 논의하였다. 구체적으로, 5G와 사용자 단말 간의 스펙트럼을 공유하는 5G 딥러닝 패턴을 학습하고 데이터 및 제어 신호를 재밍(jamming)하여 5G 통신을 방해하는 시나리오와 공격자가 5G gNodeB에서 딥러닝 기반 물리 계층 인증 시스템을 통과하기 위한 스푸핑 공격 기반 적대적 공격 시나리오를 고려했다. 이때, 공격자는 5G 네트워크 슬라이싱 응용 프로그램에 대해 액세스하기 위해 유사한 신호를 전송하여 권한을 얻으며 적대적 공격을 수행한다. 즉, 공격자는 합성 데이터를 무선으로 학습시켜 스푸핑 신호를 생성하고 이를 전송하여 5G 신호 인증에 침투할 수 있다.

해당 논문의 경우 5G 시스템에서 의도적인 오류를 만들어 사용자가 부정확한 모델을 학습하도록 하고, 성능에 부정적인 영향을 끼치는 적대적 공격 가능성에 대해 증명한 점에서 기여점이 있다. 하지만 해당 논문의 경우, 5G 네트워크의 핵심 패러다임인 에지 클라우드에 대한 고려가 이루어지지 않았다. 5G 네트워크상에서 수많은 단말 장치가 방대한 양의 데이터를 생성하여 MEC에서 분산 학습이 이루어지고, 중앙 클라우드 서버와 연계하여 수행하는 연합학습이 주목받기에 이에 대한 공격 영역 식별도 고려되어야 한다[15]. 구체적으로, 에지 클라우드를 고려한 연합학습 시나리오에서 로컬 모델의 입력으로 악성 데이터를 사용할 수 있는 잠재적인 데이터 오염 공격이 발생할 수 있다. 또한, 무선 액세스 네트워크를 대상으로 했다는 점에서 코어 망에서의 데이터 오염 공격을 고려한 본 연구와 차이점이 있다.

IV. 5G MEC 환경에서의 NWDAF 위협 모델 및 배치 시나리오

본 연구에서는 네트워크 기능에 대한 데이터 오염 공격이 발생할 수 있는 5G 네트워크 환경에서의 NWDAF 아키텍처 및 시나리오를 선행 연구를 바탕으로 고안하고, 5G MEC 환경에서 인공지능을 활용하는 NWDAF 아키텍처 배치 시나리오를 설명한다. 먼저 데이터 오염 공격의 예시로 본 연구에서는 Slowloris 공격을 대상으로 선정하였는데, Slowloris 공격은 헤더 정보에 대한 오염을 통해 웹 서버는 트래픽이 아직 전송 중이라고 잘못 인식하게 한다는 점에서 데이터에 잡음(noise)을 주는 오염 공격의 한 예로 활용 가능하다고 볼 수 있다.

5G MEC 환경에서의 네트워크 기능 배치 시나리오의 경우, 선행 연구[10, 17]에서 제안된 아키텍처를 바탕으로 한다. Fig. 1.은 5G MEC 환경에서의 분산 NWDAF 아키텍처를 나타낸다. 기본적인 5G 네트워크 구성요소인 gNodeB, UE, 코어망 내 NRF, AMF(Access Management Function) 및 SMF(Session Management Function)와 같은 네트워크 기능들이 존재하며, 학습을 위한 NWDAF가 존재한다. 이때, 연합학습에 적합한 분산형 NWDAF 구조를 위해 중앙집중형 클라우드에는

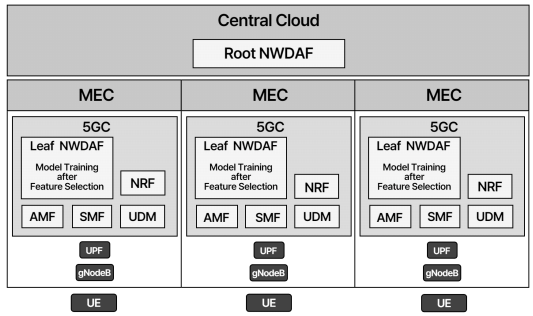

Fig. 1. A NWDAF Architecture for distributed and federated learning in 5G MEC Environment

Root NWDAF가, 각각의 MEC에는 Leaf NWDAF들이 존재한다. 이와 같은 아키텍처는 단일 장애점(single point of failure) 측면에서 보안성을 향상시키며, 네트워크 자원 사용량을 효율적으로 관리할 수 있다[10].

이와 같은 아키텍처에서 단말로부터 공격 데이터가 유입되어 NRF에 오염된 데이터가 존재하는 위협 모델을 가정한다. 이때, 오염된 데이터는 의도적으로 라벨이나 특징 값들을 조작하여 인공지능 학습 모델의 정확도를 떨어뜨리는 데이터를 의미한다. 공격자는 네트워크에서 악의적인 공격 트래픽과 정상 트래픽과의 구분을 회피하고자 한다. 이러한 악의적인 트래픽이 정상 트래픽이라고 인식된 상태에서 학습에 사용되면 오염된 데이터로 간주할 수 있다. 예를 들어, Slowloris 공격으로 HTTP 헤더 정보가 변경된 악의적인 트래픽이 정상으로 인식되어 학습에 사용될 경우, 이러한 데이터는 오염된 데이터로 간주할 수 있다. 이러한 오염된 데이터가 NWDAF에서 활용될 경우, NWDAF의 학습 모델의 정확도를 낮추고 오류를 발생시킬 수 있다. 정리하면, 제안한 시나리오에서는 탐지에 따른 라벨을 정상과 악성으로 구분하고, 오염된 데이터가 식별되었을 경우 이를 악성 트래픽으로 라벨링한다.

정상 및 오염 데이터에 따른 NWDAF 아키텍처의 구체적인 동작 과정은 Fig. 2.와 같다. 이때 MEC마다 1개의 단말이 연결되는 식으로 gNodeB는 단말로부터 정상 및 오염 데이터를 받고, AMF 및 SMF에 데이터를 전달하기 위해 UPF(User Plane Function)를 거친다. 본 논문에서는 각 MEC에서 모델 성능을 높이기 위해 Leaf NWDAF에서 특징 선택을 수행하는 시나리오를 고려한다. 전체적인 동작 과정은 크게 3가지 프로세스로 구분된다. 이때, 설명의 편의를 위해 하나의 MEC 환경을 기준으로 실행 흐름을 설명한다.

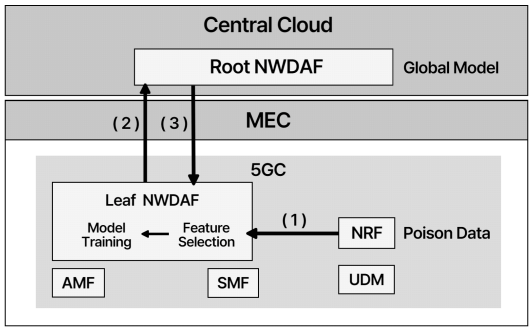

Fig. 2. Operation process of NWDAF architecture including feature selection in 5G MEC environment

(1) 5G 네트워크를 사용하는 악의적인 사용자로부터 오염 데이터가 포함된 네트워크 트래픽이 발생한다. 5G NF 간 상호 연동을 지원하는 NRF에 AMF, SMF, UDM 등 여러 NF를 거쳐 단말로 유입된 정상 및 오염 데이터가 저장된다. 이후 Leaf NWDAF는 NRF에 존재하는 정상 및 오염 데이터를 전달받고, 전처리 및 특징 선택을 수행한다. 마지막으로, Leaf NWDAF 내의 지역 인공지능 모델로 학습을 수행한다. 해당 과정에서 오염 데이터로 인해 적대적 교란으로 인한 오분류가 발생 가능하며, 학습 성능이 떨어진 채로 각 Leaf NWDAF의 데이터 셋에 대해 학습 정보인 가중치(weight)가 생성된다.

(2) 각 MEC에서의 Leaf NWDAF는 학습 결과의 가중치들을 Root NWDAF로 전달한다. 가중치들을 집계하여 Root NWDAF가 배치된 중앙집중형 클라우드에서 전역 모델을 생성한다. 연합학습의 특성으로 각 MEC 별로 데이터를 공유하지 않고도 전역 모델 생성을 허용한다.

(3) 집계 서버인 Root NWDAF에서 데이터 오염 공격의 영향을 받은 전역 모델을 생성하여 Leaf NWDAF로 전달한다. Leaf NWDAF는 적대적 교란의 영향을 받은 전역 모델을 지역 모델의 업데이트에 활용되며, 따라서 각각의 MEC는 데이터 오염 공격의 영향을 받게된다.

본 연구에서는 이러한 MEC 기반 5G NWDAF 아키텍처 동작 과정에서, 데이터 오염 공격의 영향과 이를 탐지하기 위한 특징 선택을 통한 성능 향상을 분석하고 실험을 통해 평가하고자 한다.

V. 특징 선택을 통한 데이터 오염 공격 탐지 성능 향상

본 장에서는 설명한 5G MEC 배치 시나리오에서 기계학습 기반 데이터 오염 공격 탐지 성능 향상을 위한 특징 선택 기법을 탐구한다. 실제 트래픽에 기반한 5G 네트워크 데이터는 기본적으로 통신 사업자의 데이터로, 프라이버시 이슈로 인해 얻기 힘들고 공개된 오픈 소스 데이터 셋이 충분하지 않다. 이중 5G 네트워크에서의 침입 탐지 데이터를 제공하는 5G-NIDD 내 Slowloris 공격 데이터를 4장의 배치 시나리오에서 언급한 바와 같이 악의적인 트래픽을

정상으로 오분류하도록 유도하는 오염된 데이터로 가정한다. Slowloris 공격에 대해 특히 데이터의 라벨 뿐만 아니라, 특징 값 자체에도 오염을 유발하는 예로 보며, 5.2장에서 주요 특징을 살펴보면서 설명하고자 한다. 전체 연합 학습 프로세스 중 지역 모델에 해당하는 Leaf NWDAF에서의 데이터 오염 탐지를 통해 전체 아키텍처 중 일부에 대한 개념 증명(proof of concept)으로 성능을 평가하고자 한다.

본 제안에서는 데이터 오염 공격 탐지 성능 향상을 위해 예측의 성능을 저하시키는 특징들을 제거함으로써 과소적합(underfittng)에 대응할 수 있도록 특징 선택을 수행한다. 이때, 결정 트리(decision tree), 선형 회귀(logistic regression), SVM 분류기를 이용하여 특징 선택 여부에 따른 분류 성능을 비교함으로써 제안한 특징 선택 방법론의 실효성을 평가하고자 한다.

5.1 5G-NIDD 데이터 셋

5G-NIDD[10] 데이터 셋은 5G 무선 네트워크를 통해 생성된 포괄적인 네트워크 침입 탐지 데이터를 제공한다. 구체적으로, 해당 데이터 셋은 AI/ML 관련 솔루션을 개발하고 테스트하는 사용자들을 위해 5G 테스트베드에서 구축되고 추출한 데이터가 포함되며 레이블이 지정된 데이터 셋이다. 5G 테스트베드는 핀란드 오울루 대학교의 네트워크에 연결되어 있으며, 데이터는 공격자 노드, 5G 사용자가 있는 두 기지국에서 추출된다[18]. 공격자 노드는 5GTN MEC 환경에 구축된 서버를 공격하며, 서비스 거부와 포트 스캔이 공격 시나리오에 포함된다. 본 연구에서는 이중 서비스 거부 공격 중 하나인 Slowloris 공격 데이터를 학습 데이터 셋으로 사용한다.

Slowloris 공격 데이터 셋 관련 내용은 Table 1.

Table 1. 5G-NIDD dataset overview (Slowloris Attack)

Number of features	113 (SrcID, Flgs, RunTime, Sum, Rate, Label etc)
Total instances	31015
Values of 'Label' feature	Benign, Malicious
# of 'Benign' instances	24808
# of 'Malicious' instances	6207

과 같으며, 이는 30분 세션 중 약 10분 동안 특정 웹 서버에 Slowloris 공격을 수행하여 리소스 고갈을 발생시킨 네트워크 트래픽으로부터 추출한 정보이다. Label 항목에서 정상(Benign), 악성(Malicious)이라는 값을 확인할 수 있으며, 이때 악성인 데이터가 학습 결과에 오분류를 야기시키는 오염 데이터로 간주한다.

5.2 특징 선택 수행

특징 선택(feature selection)은 모델을 구성하는 특징들을 수정하지 않고, 중요한 특징을 선택하는 것이다[19]. 라벨과 상관관계가 낮은 특징은 학습 과정에서 모델의 성능을 낮출 가능성이 존재하고, 많은 특징으로 모델을 학습하면 과적합 이슈가 발생할 수 있다. 따라서 데이터 오염 공격 탐지를 위해, Slowloris 공격 데이터 셋에 존재하는 다수의 특징 중 주요 특징을 선별하여 개선된 학습모델을 만들기 위해 특징 선택을 수행했다. 우선, 특징의 value가 모두 0인 것처럼 분포의 패턴이 없거나, 결측값(Null)이 너무 많거나, 중복되거나 패턴파악이 불가능한 문자열 값을 가진 특징들을 제거하여 113개에서 30개로 줄이는 전처리 작업을 수행했다.

특징 선택의 방법으로는 크게 Filter, Wrapper, Embedded 방법이 있는데 그 중 Filter와 Wrapper 방법의 장점들을 결합한 Embedded 방법론을 사용하였다[19]. Filter 방법은 특징 간 관련성을 측정하는 방법이고, Wrapper 방법은 특징 서브셋의 유용성을 측정하는 방법이다. Embedded 방법은 Filter와 Wrapper 방법의 장점들을 결합한 방법으로, 특징 서브셋의 유용성을 측정하면서 내장 매트릭을 사용하여 각각의 특징을 직접 학습하고, 학습 절차를 최적화할 수 있다. 이것을 이용하여 모델의 정확도에 기여하는 특징을 뽑고자 했으며, 그중 결정 트리 기반 알고리즘을 이용해서 주요한 특징을 선정했다. 결정 트리 기반 알고리즘은 각각의 범주에 속하는 빈도에 기초한 데이터들의 분리를 통해 분류 트리를 구성하여 분류와 예측을 보다 쉽게할 수 있다[20].

특징 중요도 분석을 위해 데이터 셋 내 전처리된 30개의 특징들에 대해 DecisionTreeClassfier를 사용한 결정 트리 모델 알고리즘[20]을 이용하였다. Slowloris 공격이 일어날 경우, 패킷 헤더 내 마지막 개행 문자에 대한 변조로 인해 학습 과정에서 사

용되는 데이터 인스턴스 내 특징 중 하나인 Rate 값이 영향을 받게 된다. 이때, Rate 특징은 에코 요청에 따른 응답률로, Slowloris 공격에 의해 특정 웹 서버에 HTTP 헤더를 지속적으로 전송하여 연결 완료를 방해함으로써 정상 트래픽과 다르게 응답률이 낮아질 수밖에 없다. 웹서버는 완전한 헤더 정보가 올 때까지 에코 요청에 응답하지 않고 기다리며, 그 결과 공격을 받아 변조된 데이터로부터 추출된 Rate 특징의 값은 정상 데이터와 다르게 0이 된다. Slowloris 공격은 기본적으로 가용성을 침해하는 서비스 거부 공격의 일종이기는 하나, 코어망에서 데이터를 수집한 후 학습을 수행하는 NWDAF의 관점에서는 학습 데이터의 특징 값에 대한 변조를 야기시킴으로써 학습 결과를 왜곡시키는 데이터 오염 공격으로 간주할 수 있다. 따라서 Rate 특징이 Slowloris 공격의 Label을 결정짓는 핵심적인 특징이므로 Label 값을 Rate 특징 기준으로 분리하는 것을 목적으로 한다. 데이터 셋의 학습과 테스트 비율을 7:3으로 나눈 뒤 결정 트리 모델을 생성하여 확인한 결과, 테스트데이터 셋의 정확도는 93.7%를 가진다. 이에 대해 파이썬 라이브러리인 feature_importance_를 사용하여 Rate를 기준으로 각 특징의 중요도를 뽑았다. 그 결과 Sum이 약 43%, SrcRate가 약 8%, TotPkts가 7%의 순으로 확률값이 높은 것으로 보아 중요한 특징임을 확인하였다.

Table 2.는 Slowloris 공격 데이터 셋 특징 총 113개에서 전처리된 30개의 특징들 중 중요도가 높은 특징들에 대한 설명을 한다. 네트워크 흐름에 대해 Argus 툴[21]로 수집한 내용이며, 공격 여부를 나타내는 Label 특징과 중요도가 높은 Rate, Sum, TotPkts, SrcRate 특징들의 정보를 각각 보여준다.

Table 2. Information on the main features of the slowloris dataset

Main features	Information
Label	(integer) Slowloris attack or not
Rate	(float) Response rate per echo request
SrcRate	(float) Source to Destination packets per second
Sum	(float) total duration of aggregated record
TotPkts	(integer) Total transaction packet count

검증을 위해 메모리를 초기화하여 중요도가 높은 Rate, Sum, TotPkts, SrcRate 4개의 특징들을 제외한 모든 특징들에 대해 같은 방법으로 알고리즘을 실행해본 결과, Fig. 3.에서 테스트데이터 셋의 정확도가 93.7%에서 91.7%로 약 2%가 감소한 것을 확인할 수 있다.

이는 앞서 제외한 4개의 특징이 중요하다는 것을 입증하는 바이다. 마찬가지로, 중요한 특징들을 제외하여 테스트데이터 셋의 정확도를 확인한 결과는 Fig. 3.과 같다. 결론적으로, Label을 구분짓는 핵심 특징인 Rate를 포함, 관련하여 중요도가 높은 Sum, TotPkts, SrcRate 4개의 특징을 선택하여 학습한 뒤 5G 데이터 오염 공격 중 하나인 Slowloris 공격 탐지 성능을 최종 평가하고자 한다.

```
* Rate, Sum, TotPkts, SrcRate 특성 제외
테스트 셋 정확도: 0.9165
학습 셋 정확도: 1.0000
* Rate, Sum, TotPkts 특성 제외
테스트 셋 정확도: 0.9199
학습 셋 정확도: 1.0000
* Rate, Sum 특성 제외
테스트 셋 정확도: 0.9376
학습 셋 정확도: 1.0000
```

Fig. 3. Accuracy of test dataset and training dataset

5.3 분류 실험 결과

특징 선택의 효율성 평가를 위해, 3가지의 지도학습 기반 기계 학습 분류기인 결정 트리, 선형 회귀, SVM을 이용하여 주어진 데이터 셋이 Label이라는 클래스에 따라 분류가 이루어지는지 확인하였다. Slowloris 공격 데이터 셋에서 Label을 제외한 29개 특징과 앞서 Slowloris 공격 데이터 셋에 대한 특징 선택으로 중요도가 높았던 데이터 특징 4개 (Rate, Sum, TotPkts, SrcRate)를 나누어서 비교 분석하였다. 이때, 데이터 셋의 학습률을 0.7, random_state는 42로 설정하였다.

실험 환경 구축을 위한 기계 학습 툴체인의 경우, Google Colab을 활용하였다. GeForce RTX 3090 GPU를 사용하고, Ubuntu 22.04 버전이 설치된 호스트에 nvidia-driver-525와 CUDA

Table 3. Comparison of classification performance according to feature selection

	Type (# of features)	Decision Tree	Logistic Regression	SVM
Accuracy	ALL features (29)	0.917	0.789	0.801
	proposed approach (4)	0.938	0.833	0.949
Precision	ALL features (29)	0.920	0.725	0.636
	proposed approach (4)	0.942	0.736	0.966
Recall	ALL features (29)	0.916	0.080	0.798
	proposed approach (4)	0.938	0.251	0.773
F1	ALL features (29)	0.916	0.150	0.708
	proposed approach (4)	0.937	0.375	0.859

version 12.0을 설치하여 실험을 수행했다.

Table 3.은 전체 데이터 셋의 30%에 해당하는 테스트 데이터 셋에 대한 분류 Accuracy, Precision, Recall, F1 score이며, 소수 넷째 자리에서 반올림하여 나타낸 결과이다. 3가지 분류기 모두 전처리된 전체 특성을 모두 포함하여 학습을 진행하는 것보다 데이터 오염 공격 탐지 성능 향상을 목적으로 특징 선택을 한 경우가 분류 성능이 개선됨을 확인할 수 있다. 주목할 점은 특징 중요도 분석 시 활용된 결정 트리 분류기 성능 뿐만 아니라, 선형 회귀 및 SVM 모델에서도 성능이 향상된다는 것이다. 이는 분류 성능 향상을 위한 특징 선택의 필요성을 반증한다.

선형 회귀 분류기의 경우, Recall과 F1 score가 다른 지표에 비해 낮은 결과를 갖는다. 이것은 서로 다른 클래스를 구분하기 위한 경계선을 생성하는 방식이 두 분류기와 다르기 때문이다. 선형 회귀의 경우 선형 경계선의 결정 임계값에 따라 Recall, F1 score 값이 바뀌므로 이 값을 얼마로 지정하느냐에 따라 분류 성능이 결정된다. 반면에, 결정 트리와 SVM의 경우 비선형 경계선까지 다룰 수 있으며 클래스를 구분하기 위한 최적화된 방향으로 경계선을 자동 지정하기 때문에 선형 회귀에 비해 분류 성능 결과가 우수하다.

5.4 논의 및 고찰

본 논문에서 실험한 데이터 셋의 경우 정상과 악성 라벨의 데이터에 대해 약 1:4의 비율로 학습을 수행했다. 하지만 정상과 악성 라벨의 비율이 달라진다면, 데이터 셋에 대한 성능 평가 결과가 달라질 수 있다. 이러한 오염 비율의 변화는 5G MEC 아키텍처 내 연합 학습 시나리오에서 각 Leaf NWDAF 마다 다양하게 나타날 수 있다. 따라서 본 제안에 관한 후속 연구로 데이터 오염 비율을 다양하게 조정을 하며 데이터 편향성에 대해 실험해볼 계획이다.

또한, 연합학습 시나리오에서 중앙집중형 클라우드는 각 MEC로부터 데이터를 수집한다. 이때, MEC의 엔티티로부터 받은 데이터에 대한 신뢰성을 완벽히 보장할 수 없는 상태로 각 데이터가 센트럴 클라우드의 집계함수에 악영향을 끼칠 수 있다. 본 논문에서는 지역 모델에서의 학습에 대해서만 성능 평가를 수행했기 때문에 연합학습 시 MEC에 입력되는 오염 비율에 따라 전역 모델에 어떤 영향을 끼치는지에 대해서 후속 연구를 수행하고자 한다.에 어떤 영향을 끼치는지에 대해서 후속 연구를 수행하고자 한다.

VI. 결 론

5G MEC 환경이 AI와 결합하여 확대됨에 따라 사이버 보안 위협 영역 증가하고 있지만, 해당 영역 식별에 대한 연구가 부족한 실정이다. 특히, 5G 네트워크 통신 과정에서 데이터 오염 공격 가능성이 있지만, 5G SA 방식의 네트워크 프로세스는 LTE 망과 차이가 존재하므로 기존의 데이터 오염 공격에 관련 연구를 적용할 수 없다는 한계점이 존재한다. 본 논문에서는 5G SA 모드에서 에지 AI 기술이 결합된 NWDAF 위협 모델에 대해 탐구하고, 제안한 배치 시나리오에서 NWDAF에 대한 데이터 오염 공격 가능성을 확인하였다. 또한, NWDAF에 대한 데이터 오염 공격 탐지 성능을 높이기 위해 특징 선택 방법을 제안하며 3가지 분류기를 통해 성능을 비교하며

특징 선택의 필요성을 입증하였다. 향후 논의 및 고찰에 따라 5G MEC 환경에서의 NWDAF 아키텍처에서 에지 AI에 대한 데이터 오염 비율에 따른 NWDAF 아키텍처 성능 비교를 하고자 한다.

References

[1] A. C. Chen Liu, O. M. K. Law, J. Liao, J. Y. C. Chen, A. J. En Hsieh and C. H. Hsieh, "Traffic Safety System Edge AI Computing," 2021 IEEE/ACM Symposium on Edge Computing (SEC), pp. 01-02, Dec. 2021.

[2] S.W.Hong et al, "Technologies of Intelligence Edge Computing and Networking," Electronics and Telecommunications Trends, 34(1), pp.23-35, Feb. 2019.

[3] M.K.Shin, "Development of Network Data Analytics Function (NWDAF) and Intelligence Technology Standards for 5G Network Automation," TPKO 202100009073, Electronics and Telecommunications Research Institute, 2021.

[4] 3GPP, "Release 16 Description; Summary of Rel-16 Work Items," TR 21.916, 2020.

[5] S. -M. Senouci, H. Sedjelmaci, J. Liu, M. H. Rehmani and E. Bou-Harb, "AI-Driven Cybersecurity Threats to Future Networks," IEEE Vehicular Technology Magazine, vol. 15, no. 3, pp. 5-6, Sep. 2020.

[6] S. Hussain, O. Chowdhury, S. Mehnaz, and E. Bertino, "LTEInspector: A Systematic Approach for Adversarial Testing of 4G LTE," Network and Distributed Systems Security (NDSS) Symposium, Jan. 2018.

[7] Chuan Yu, Shuhui Chen, Fei Wang, and Ziling Wei, "Improving 4G/5G air interface security: A survey of existing attacks on different LTE layers," Computer Networks: The International Journal of Computer and Telecommunications Networking, vol. 201, no. C, DOI:10.1016/j.comnet.2021.108532, Dec. 2021.

[8] H. Fang, X. Wang and S. Tomasin, "Machine Learning for Intelligent Authentication in 5G and Beyond Wireless Networks," IEEE Wireless Communications, vol. 26, no. 5, pp. 55-61, Oct. 2019.

[9] C. Benzaïd and T. Taleb, "AI for Beyond 5G Networks: A Cyber-Security Defense or Offense Enabler?," IEEE Network, vol. 34, no. 6, pp. 140-147, Dec. 2020.

[10] Y. Jeon, H. Jeong, S. Seo, T. Kim, H. Ko and S. Pack, "A Distributed NWDAF Architecture for Federated Learning in 5G," 2022 IEEE International Conference on Consumer Electronics (ICCE), pp. 1-2, 2022.

[11] Sehan Samarakoon et al, "5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network," IEEE Dataport, Dec. 2022.

[12] 3GPP, "Architecture enhancements for 5G System (5GS) to support network data analytics services," TS 23.288, 2021.

[13] 3GPP, "5G; 5G System; Network function repository services; Stage 3," TS 29.510, 2019.

[14] E. Yalin, E. Sagduyu, and Yi. Shi, Adversarial Machine Learning for 5G Communications Security, IEEE Game Theory and Machine Learning for Cyber Security, pp. 270-288, Jan. 2021.

[15] D. Moustis and P. Kotzanikolaou,

"Evaluating security controls against HTTP-based DDoS attacks," IISA 2013, pp. 1-6, Jul. 2013.

〔16〕 Y. Liu, J. Peng, J. Kang, A. M. Iliyasu, D. Niyato and A. A. A. El-Latif, "A Secure Federated Learning Framework for 5G Networks," IEEE Wireless Communications, vol. 27, no. 4, pp. 24-31, Aug. 2020.

〔17〕 S.Y.Lee et al, "Federated learning over private 5G networks: demo", Proceedings of the Twenty-Third International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '22). Association for Computing Machinery, pp. 295 - 296, Oct. 2022.

〔18〕 E. Piri, P. Ruuska, T. Kanstren, J. Mäkelä, J. Korva, A. Hekkala, A. Pouttu, O. Liinamaa, M. Latva-Aho, and K. Vierimaa, "5GTN: A Test Network for 5G Application Development and Testing," 2016 European Conference on Networks and Communications (EuCNC), pp. 313 - 318, Jun. 2016.

〔19〕 Girish Chandrashekar, Ferat Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16-28, Nov. 2014.

〔20〕 N. M. Tahir, A. Hussain, S. A. Samad, K. A. Ishak and R. A. Halim, "Feature Selection for Classification Using Decision Tree," 2006 4th Student Conference on Research and Development, pp. 99-102, Jun. 2006.

〔21〕 Argus, "Argus: System + Network Monitoring," https://jaw0.github.io/argus5docs/docs/, 2023. 03. 30.

〈저자소개〉

옥 지 원 (Ji-won Ock) 학생회원
2022년 8월: 성신여자대학교 융합보안공학과 학사
2022년 9월~현재: 성신여자대학교 미래융합기술공학과 석사과정
〈관심분야〉 정보보호, 인공지능, 통신공학, 클라우드

노 현 (Hyeon No) 학생회원
2022년 8월: 성신여자대학교 융합보안공학과 학사
2022년 9월~현재: 성신여자대학교 미래융합기술공학과 석사과정
〈관심분야〉 정보보호, 클라우드 컴퓨팅, 무선 이동통신망 보안, 신뢰 실행 환경

임 연 섭 (Yeon-sup Lim) 정회원
2007년 2월: 서울대학교 컴퓨터공학부 학사
2009년 2월: 서울대학교 전기컴퓨터공학부 석사
2017년 2월: Ph.D., Computer Science, University of Massachusetts Amherst,
〈관심분야〉 네트워크, 이동통신, 연합학습, 클라우드 컴퓨팅

김 성 민 (Seong-min Kim) 종신회원
2012년 2월: 한국과학기술원 전기 및 전자공학과 졸업
2014년 2월: 한국과학기술원 전기 및 전자공학과 석사
2019년 2월: 한국과학기술원 정보보호대학원 박사
2019년 9월~2020년 8월: 삼성전자 삼성리서치 Staff Engineer
2020년 9월~현재: 성신여자대학교 융합보안공학과 조교수
〈관심분야〉 신뢰 실행 환경, 클라우드 컴퓨팅, 시스템 보안