KNOM Review '23-02 Vol.26 No.02
https://doi.org/10.22670/knom.2023.26.2.73

원격 제어 차량 운영 환경에서의 딥러닝 기반 중간자 공격 탐지 방법

백 의 준◆, 박 지 태*, 최 정 우* 김 명 섭°

DL-based Detection and Prediction of Man-in-the-Middle Attacks in Remote Control Vehicle Environments

Ui-Jun Baek◆, Jee-Tae Park*, Jeong-Woo Choi*, Myung-Sup Kim°

요 약

원격 제어 차량 기술은 자율 주행 자동차, 드론, 로봇 등 다양한 분야에서 활용되며 나날이 발전하고 있으나 이를 대상으로 한 보안 위협 및 사이버 범죄 또한 늘어나는 추세이다. 차량 내 이상 및 공격 탐지를 목표한 많은 연구들이 수행되고 있으나 원격 제어 차량 운영 환경에서 발생할 수 있는 공격 시나리오를 고려한 연구는 없으며 데이터셋 또한 없는 실정이다. 이에 본 논문은 LSTM 기반의 이상 탐지 모델을 통해 원격 제어 차량 운영 환경에서 발생할 수 있는 중간자 공격을 탐지하는 방법을 제안한다. 또한, 제시한 5개의 중간자 공격 시나리오에 따라 LSTM을 활용하여 정상 차량 상태 데이터셋에 공격 데이터를 합성하는 방법을 제안하고 합성된 데이터를 통해 탐지 모델을 평가한다. 제안한 이상 탐지 방법은 제시한 5가지 공격 유형 중 3개의 시나리오에서 90% 이상의 높은 정확도를 나타내었다.

Key Words : Remote control vehicle, Man-in-the-Middle, Anomaly detection, Long short term memory

ABSTRACT

The remote control vehicle technology is widely used in various fields such as autonomous driving cars, drones, and robots, and it is constantly advancing. However, there is also a growing trend of security threats and cybercrimes targeting these technologies. While many studies aim to detect anomalies and attacks within vehicles, there is a lack of research considering attack scenarios that can occur in remote control vehicle operating environments, and datasets are also non-existent. In this paper, we propose a method for detecting man-in-the-middle attacks that can occur in remote control vehicle operating environments using an LSTM-based anomaly detection model. Additionally, we propose a method for synthesizing attack data into normal vehicle state datasets using LSTM based on five suggested man-in-the-middle attack scenarios and evaluate the detection model using the synthesized data. The proposed anomaly detection method achieved an accuracy of over 90% in three out of the five attack scenarios presented.

※본 연구는 20203년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No. RS-2023-00230661, 하이브리드 양자키분배 방법 및 망 관리 기술 표준개발) 및 산업통산자원부의 재원으로 한국산업기술진흥원의 지원 (No. P0016151, 원격 제어 차량을 위한 AI 기반 사이버 보안 기술 개발)을 받아 수행 된 연구임
◆ First Author : Korea University Department of Computer and Information Science, pb1069@korea.ac.kr
° Corresponding Author : Korea University Department of Computer and Information Science, tmskim@korea.ac.kr
* Korea University of Department of Computer and Information Science, {pjj5846, choigoya97}@korea.ac.kr
논문번호 : KNOM2023-02-10, Received November 25, 2023; Revised December 3, 2023; Accepted December 12, 2023

Ⅰ. 서 론

RCV(Remote Control Vehicle) 기술은 차량을 원격에서 조작하고 모니터링하기 위한 기술 및 시스템으로 주로 차량의 움직임, 동작 및 기능을 조작 및 모니터링한다. RCV 기술은 초기 군사적 용도로 개발된 이후 산업 및 상업 응용으로 확장되었으며 2000년대 이후 RCV 기술은 자동화 및 센서 기술의 발전을 통해 혁신을 이루었으며 현재에는 인공지능 및 기계 학습 기술의 발전으로 자율 주행 자동차, 드론, 로봇 및 무인 항공기와 같은 다양한 분야에서 중요한 역할을 하고 있다. RCV 기술은 자동차 운전 환경에서의 효율성, 안전성 편의성을 향상하는 데 이바지한다. 특히, 최근 V2X(Vehicle to Everything) 기술과의 통합으로 상승효과가 극대화되고 있으며, 예를 들어, V2X 통신을 통해 다른 차량의 정보를 수신하고, 그 정보를 바탕으로 RCV를 사용하여 차량을 원격으로 조작하여 교통 상황을 관리하거나 비상상황에 차량을 안전하게 제어할 수 있다. 또한, 최근 급성장 중인 자율주행기술과 연계하여 테스트, 안전 검증 및 원격 관리 등 다양하게 활용되는 추세다.

RCV 기술은 나날이 발전하고 있으나 이를 대상으로 한 보안 위협 및 사이버 범죄 또한 늘어나고 있으며 정교해지고 있다. AIEDGE LABS에서 보고한 바에 따르면 2021년에 전 세계적인 사이버 범죄로 인한 피해는 추정치로 6조 달러에 달했으며, 2025년까지 연간 10조 달러까지 증가할 것이라고 예상된다. 또한, 자동차 산업에서의 사이버 공격 또한 2019년부터 2022년까지 3년간 225% 증가하였고 2024년까지 5,050억 달러의 손실이 예상된다 [1]. 이에, 연구자들과 기업에서는 사전에 보안 취약점 분석을 통해 공격으로 인한 피해를 최소화하고자 노력하고 있다. 예를 들어, Fiat Chrysler Automobiles 사가 자사의 “Jeep Cherokee” 모델이 내장된 엔터테인먼트 시스템의 보안 취약점으로 인해 제어할 수 있다는 검증 결과에 따라 약 백만 대의 차량을 리콜한 대표적인 사례가 있다 [3]. 이후 차량에 대한 사이버 보안을 위한 국제표준들이 제정되었으며 대표적으로 UNECE WP.29(World Forum for Harmonization of Vehicle Regulations) UNR.155(Cyber security and cyber security management system) [3], ISO/SAE 21434(Road vehicles - Cybersecurity Engineering) [4] 등이 있다. 다른 사례로는 사이버 보안 대회 Pwn2Own 2023에서 Tesla model3의 소프트웨어 시스템이 해킹된 사례이다 [5]. 대회 우승자는 차량의 인포테인먼트와 게이트웨이 네트워크로 나눠진 시스템을 모두 탈취하고 제어하는 것에 성공했다.

상기한 보안 위협 및 취약점을 해결하기 위하여 많은 연구가 수행되고 있으나 RCV 기술은 아직 많은 잠재적인 보안 문제와 취약성을 가지고 있다. RCV 기술에서 고려해야 할 주요 보안 문제는 다음과 같다:

1) 인증 및 접근제어 : 원격으로 차량을 조작하려는 사용자의 신원을 확인 후 권한이 부여되어야 하며 부정한 접근 방지를 위해 강력한 인증 및 접근제어 메커니즘이 필요하다.
2) 데이터 보호 : RCV 시스템에서 전송되는 데이터는 보호되어야 하며 암호화와 같은 기술을 사용하여 데이터 무결성 및 기밀성을 보장해야 한다.
3) 원격 공격 방지 : 공격자가 원격으로 차량 시스템에 침투하여 조작하려는 시도를 방지해야 하며 이를 위해 침입 탐지 시스템 및 방화벽과 같은 보안 메커니즘이 필요하다.
4) 펌웨어 및 소프트웨어 업데이트 검증 : RCV 기술 시스템 업데이트 시 업데이트 프로세스는 안전해야 하며, 악의적인 소프트웨어 업데이트를 방지해야 한다.
5) 물리적 보안 : 원격 조작 시스템은 물리적으로 안전하게 보호되어야 하며 적절한 물리적 보안 조치가 필요하다.
6) 긴급 대응 계획 : 보안 사고가 발생할 경우를 대비하여 신속한 대응 및 손상 최소화를 위한 긴급 대응 계획이 마련되어야 한다.

제시된 6개 중 1) ~ 5)는 보안 사고를 예방하는 것이며 6)은 보안 사고가 발생한 이후에 대한 계획을 말하며 모두 중요하다. 그러나, 만약 보안 사고가 발생한다면 인명 피해를 포함하여 막대한 피해를 입게 되므로 공격이 발생한 타이밍에 이를 탐지하고 긴급 대응할 방법에 관한 연구는 차량 보안 분야의 필수적인 최후의 보루 구축과 같다. 이와 관련하여 차량 내 ECU(Electronic Control Unit)에서 수집되는 센서 값들을 모니터링하고 이에 대한 이상을 탐지하는 연구가 수행되고 있으나 최근 차량의 보안 위협으로 인해 발생한 이상 발생 사례가 계속 보고되고 있는 만큼 보안 위협으로 인한 이상 발생에 대한 구체적인 시나리오에 대한 정의와 깊

이 있는 분석이 필요하다.

본 논문에서는 제어 센터와 원격지 차량이 단대단 통신하는 환경에서 제어 메시지가 중간 공격자에 의해 탈취 및 변조된 시나리오를 제시하고 이에 대한 탐지 방법을 제안한다. 현재 원격제어차량 운영 환경에서 발생가능한 시나리오를 고려한 연구는 없으며 이와 관련된 공개데이터 세트 또한 없다. 이에, 본 논문에서는 해당 시나리오에 따라 발생할 수 있는 보안 위협을 반영한 가상의 공격 데이터 세트 생성 방법을 제안하고 이를 통해 평가를 수행한다. 또한, 딥러닝 기반 이상 탐지 모델을 통해 가상의 데이터 세트에 포함된 공격을 탐지하는 방법을 제안하며 우리가 제안한 방법은 5개 중 3개의 공격 유형에 대하여 90% 이상의 높은 정확도로 공격을 탐지했다.

본 논문은 서론에 이어 관련 연구로 차량 내 가능한 중간자 공격에 대한 요약과 차량 내 이상 탐지에 관한 연구들을 제시하고 분석한다. 본론에서 공격 시나리오를 반영하는 가상의 데이터 세트 생성 방법과 이를 탐지하는 방법을 소개한다. 마지막으로 결론에서 제안한 방법에 대한 요약, 한계점 그리고 향후 연구를 제시하고 논문을 마친다.

Ⅱ. 관련 연구

2.1. 중간자 공격

중간자 공격은 사용자의 인터넷 서버와 해당 인터넷 트래픽의 목적지 사이에 끼어들어 데이터 전송을 가로채는 공격이다. Verizon의 데이터 조사 보고서 [6]에 따르면 중간자 공격이 가장 흔한 유형의 보안 공격 중 하나임을 보였으며 Frankel 등 [7]은 중간자 공격을 네트워크 보안에 대한 주요 위협 중 하나로 설명하였다. Netcraft [8]에 따르면 중간자 공격 HTTPS 서버의 95%에 위협을 가할 것으로 보고했으며 F5의 “The State of Application Strategy in 2022” 보고서 [9]에 따르면, 중간자 공격의 모든 공격 중 50% 이상이 로그인 자격 증명 및 은행 정보와 같은 민감한 정보의 가로채기를 포함한다고 보고했다.

중간자 공격은 실제 안전한 연결에 연결되었는지 또는 비슷한 비 안전한 연결에 연결되었는지를 이해하기 어렵게 만드는 방식으로 작동한다. 사용자가 네트워크와 연결을 설정하려고 할 때, 사용자는 먼저 사용자 장치에 대한 정보를 포함하는 패킷을 해당 네트워크로 보내고 네트워크는 암호화된 연결 키와 사용자 장치 주소를 포함하는 디지털 인증서를 생성한다. 연결 초기화 중에 전달되는 인증서가 안전하지 않기 때문에 공격자는 쉽게 디지털 인증서에 액세스하고 인증서 내의 정보를 수정할 수 있으며, 이로써 인증서의 승인을 사용자에게 남긴다. 많은 사용자는 위조된 및 중복된 인증서의 진위를 확인하고 그에 따른 위협을 파악하는 충분한 지식이 없으므로 이러한 인증서를 수용하고 안전하지 않은 네트워크에 연결을 허용하며, 이로써 공격자들이 공격을 실행할 가능성을 제공한다 [10].

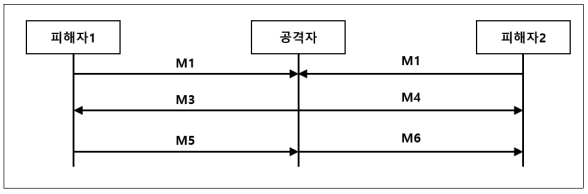

그림 1. 일반적인 중간자 공격의 메시지 교환
Fig. 1. Exchanged messages in a typical MITM attack

중간자 공격의 일반적인 시나리오는 그림 1과 같다. 두 피해자와 하나의 공격자가 있을 때, 공격자는 두 피해자 사이의 통신 채널에 접근하고 이들의 메시지를 조작한다. 피해자들이 서로 공개키(메시지 M1, M2)를 보내서 안전한 통신을 초기화하려고 할 때, 공격자는 메시지 M1, M2를 가로채고 자신의 공개 키를 피해자들에게 보낸다(메시지 M3, M4). 그 후 피해자 1은 공격자의 공개키로 메시지를 암호화하고 피해자2에게 보내며(메시지 M5) 공격자는 메시지 M5를 가로채 알려진 개인 키를 사용하여 해독한다. 마지막으로 공격자는 피해자2의 공개키로 평문을 암호화하여 피해자2에게 전송한다 (메시지 M6).

2.2. 차량 내 이상 탐지 연구

Wang 등은 HTM(hierarchical temporal memory)를 사용한 분산 이상 탐지 시스템을 제안했으며 HTM 모델은 이전 학습 상태에 따라 실시간으로 상태를 예측할 수 있다는 강점을 가진다. 또한, 데이터 필드에서 필드 수정 및 리플레이 공격을 수동으로 합성하고 이를 기반으로 제안한 모델을 평가했다 [11]. Zhang 등은 CAN(Controller Area Network) 주기성 또는 프로토콜에 의존하지 않으며 프레임 페이로드를 검사하는 화이트리스트 기반 이

표 1. 차량 내 이상 탐지 연구 요약
Table. 1. Summary of Research on Anomaly Detection in Vehicles

논문 번호	게재 년도	데이터셋				
		수집 환경	공격 데이터 타입	공격 유형 개수	공격 유형	명령-상태 쌍 여부
[11]	2018	Impreza 차량 CAN bus	합성	1	replay	X
[12]	2023	차량 CAN bus	합성	5	random injection, targeted injection, replay, random tampering, targeted tampering	X
[13]	2023	YF Sonata, Soul 등 차량 CAN bus	합성	4	spoofing, fuzzing, sniffing, replay	X
[15]	2023	시뮬레이션 환경	합성	3	fuzzing, replay, injection	X
[16]	2023	Lincoln MKZ 차량 ROS 네트워크	합성	5	hijiack, bias, injection, DoS, replay	X
[17]	2022	HCRL-car hacking dataset	모의공격	3	DoS, fuzzing, spoofing	X
[18]	2022	차량 CAN bus	모의공격	4	spoofing, fuzzing, sniffing, replay	X
[19]	2021	UTU dataset	합성	5	delete, single injection, random injection, insert message, overwrite observed message	X
Proposed		원격 제어 차량 운영 환경 내 CAN bus	합성	5	hijiack, bias, injection, DoS, replay	O

상 감지 방법을 제안했다. 5개 가능한 공격 유형에 따라 평가를 수행했으며 공격 유형은 Random Injection, Targeted Injection, Replay, Random Tampering, Targeted Tampering으로 이루어져 있다. 제안한 방법은 Replay 공격 유형에서 20% 향상된 악성 프레임 감지율을 보였으며 변조 공격 상황에서 20% 향상된 악성 패킷 감지율 향상을 보였다 [12]. Kim 등은 차량 네트워크 내 ECUs 간 브로드캐스트 통신 환경에서 발생 가능한 공격을 탐지하는 LSTM(Long Short Term Memory)-Autoencoder 모델을 제안했다 [13]. 제안한 모델은 Spoofing Attack, Fuzzing Attack, Sniffing Attack, Replay Attack에 대한 공격 시나리오에 따라 모델을 평가했으며 사용한 데이터셋 [14]은 데이터 챌린지 대회에 사용되었다. Stabili 등은 in-vehicle 통신 환경에서 n-gram 분석 기반의 이상 탐지 알고리즘 DAGA(Detecting Attacks to in-vehicle networks via n-Gram Analysis)를 제안했다 [15]. 제안하는 알고리즘은 페이로드 또는 다른 필드의 내용을 요구하지 않으며 CAN 메시지 ID만을 사용해 탐지하며 간단한 마이크로컨트롤러에서도 낮은 성능에서 실행될 수 있다는 강점을 가진다. 제안하는 방법은 Replay Attack에 대한 3가지 유형, Fuzzing Attack에 대한 2가지 유형, DoS Attack에 대한 2가지 유형에 대하여 평가가 수행되었으며 데이터셋 또한 공개되어 있다. Wang 등은 차량 내 네트워크를 위한 CWT(Continuous Wavelet Transform)과 CNN(Convolutional Neural Network)를 통합한 이상 탐지 모델은 제안했다. 제안한 모델은 입력을 CWTS(Continuous Wavelet Transform Scalogram)으로 변환 후 CNN을 통해 공간 특징을 추출하여 이상 탐지를 수행한다. 제안한 모델은 hijack, bias, injection, dos, replay 시나리오에 대하여 평가가 수행되었다 [16]. Aksu 등은 특징 선택과 분류자를 기반으로 한 새로운 침입 탐지 프레임워크를 제안했다. 저자는 MGA(Modified Genetic Algorithm)을 포함하는 메타 휴리스틱 알고리즘 m-feature를 기반으로 차원 축소를 수행하고 이를 SVC(support vector classifier), LRC(logistic regression classifier), DTC(decision tree classifier), KNC(k-nearest neighbors classifier) and LDAC(linear discriminant analysis)를 통해 분류한다. 제안된 방법은 HCRL-car hacking dataset, UNSW-NB15, CIC-IDS2017에서 평가되었다 [17]. Kim 등은 CAN 버스 프로토콜을 사용하는 차량 내 네트워크를 보호하기 위한 새로운 이상 감지

프레임워크를 제안했다. 제안된 프레임워크는 네트워크의 정상성을 나타내기 위해 여러 개의 HMM(Hidden Markov Model)을 사용하며, 이 모델들은 전송 시간 간격 및 페이로드 데이터 변경 두 가지 유형의 네트워크 정보를 사용하여 생성된다. 제안한 모델은 현대 아반떼(CN7)에서 수집된 CAN 메시지 및 flood, spoofing, fuzzing과 같은 공격이 포함된 데이터셋에서 평가되었다 [18]. Donmez는 메시지 식별자 시퀀스를 분석하여 차량용 CAN 버스 트래픽에서 이상 현상을 감지하는 침입 탐지 시스템을 제안했다. 제안된 방법은 대형 트럭에서 수집된 CAN 버스 데이터를 통하여 평가되었다 [19].

표 1은 최근 연구가 수행되었던 차량 내 이상 탐지 연구를 정리하여 나타낸 것이다. 표 1에 따르면 대부분의 연구에서는 차량 내 CAN bus에서 수집된 데이터셋을 사용했으며 자율 주행 차량 환경에서 주로 사용하는 ROS(Robot Operating System) 환경에서 수집된 데이터셋도 있다 [16]. 대다수의 데이터셋은 정상 데이터셋의 일부를 가상 공격 시나리오에 따라 수정한 것이며 일부 데이터셋의 경우 주행 시 실제 공격을 수행하여 수집된 경우도 있다 [17, 18]. 공격 유형은 주로 replay, injection, spoofing, fuzzing, DoS(denial of service), sniffing으로 분류되며 injection의 경우 삽입되는 값이 무작위인지, 과거의 상태 기록에서 추출해오는지(observed), 사용자가 설정한 임의의 값인지에 따라 추가적인 분류를 한다. 또한, 지금까지 수행되었던 연구들은 모두 차량 내 CAN bus에서 수집된 차량 상태에 대한 데이터만 이루어져 있는 것을 확인할 수 있다.

Ⅲ. 제어센터-원격차량 환경의 이상 탐지

3.1. 데이터셋

3.1.1. 데이터셋 수집 및 전처리

본 항에서는 연구를 위하여 사용된 데이터셋에 대한 소개 및 가상 시나리오에 따른 공격 유형 합성 방법을 소개한다.

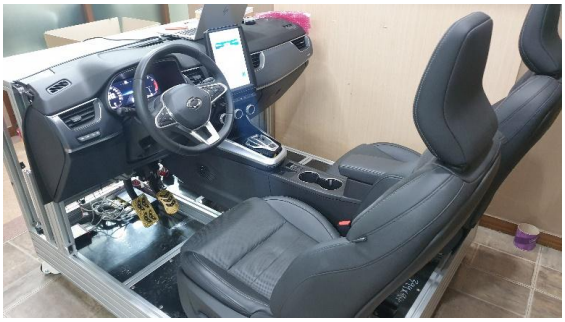

그림. 2. 원격 제어 환경 구성
Fig. 2. Setting up a remote control environment

데이터셋은 자체적으로 구성한 차량용 게이트웨이 및 원격 제어 기술을 통해 제어 센터와 원격 차량이 인터넷을 통해 통신하는 환경에서 수집된 것으로 통신 메시지는 제어 메시지-차량 상태 메시지의 쌍으로 구성되어 있다. 실제 제어 센터에 구성된 시스템은 그림 2와 같으며 실제 조향 및 가속을 수행할 때 원격 차량에서는 같은 동작을 하도록 설계되어있다.

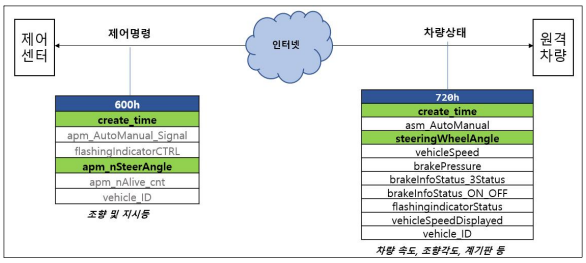

그림. 3. 수집 환경 내 간략한 통신 과정
Fig. 3. A brief communication process within a data collection environment

그림 3은 데이터셋 수집 환경 내 통신 과정에 대한 간략한 과정을 나타낸다. 제어 명령은 차량 제어를 수행하는 모듈 600h에서 생성한 데이터로 구성되어 있으며 주요 데이터는 조향각을 제어하는 apm_nSteerAngle이 있다. 제어 명령은 인터넷을 통해 원격 차량으로 전송되며 제어 센터에서 수행한 같은 동작을 수행하게 된다. 원격 차량에서는 동작을 수행하며 720h 모듈 내 실제 조향 센서 값인 steeringWheelAngle을 수집한다.

apm_nSteerAngle과 steeringWheelAngle은 제어 명령-차량상태로 구성된 데이터 쌍이지만 통신 환경 또는 차량 내 센싱 주기의 차이로 인해 시간 동기화가 반드시 필요하다. 따라서, 해당 데이터 쌍은

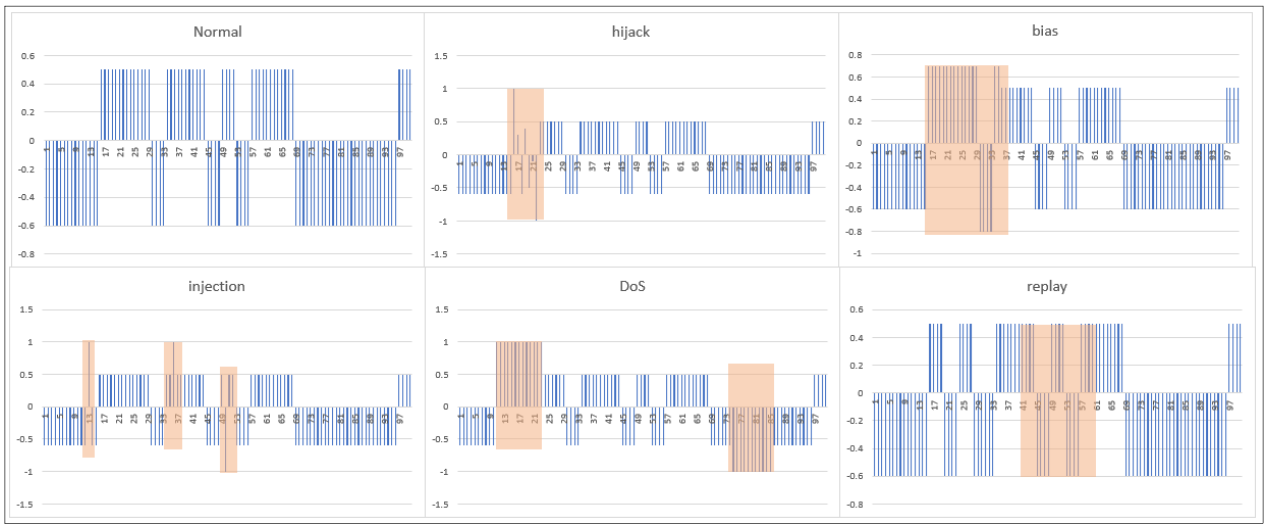

그림. 5. 공격 유형 별 데이터셋 시각화
Fig. 5. Data set visualization by attack type

데이터 수집 시간인 create_time 필드를 통해 시간 동기화를 수행하게 되며 0.2초 간격으로 보간 및 샘플링된다. 시간 동기화를 마친 데이터셋은 딥러닝 모델 입력에 적합하도록 표준화(Standardization)가 수행된다. 전처리를 마친 데이터는 그림 4와 같다.

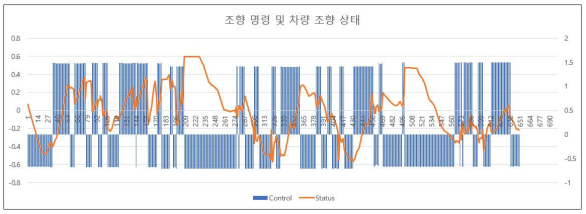

그림. 4. 제어명령-차량상태 데이터 쌍
Fig. 4. Control commands-vehicle status data pairs

3.1.2. 공격 시나리오 및 합성

데이터셋은 차량이 정상일 때만을 고려하여 수집되었으므로 공격을 가정한 시나리오에 따라 가상의 공격 데이터가 합성되어야 한다. 본 논문에서는 통신 과정에서 제어 센터에서 원격 제어 차량으로 전송되는 메시지가 탈취 및 변조되었다는 공격 시나리오 아래 공격 데이터를 합성했다.

[16]에서 제시한 공격 유형에 따라 공격 데이터를 합성한다. 공격 유형은 다섯 가지로 구성되어 있고 합성 예시는 그림 5에 나타내었으며 다음과 같다:

1) Hijack : 공격자는 특정 연속적인 구간의 값을 임의의 값으로 변조
2) Bias : 공격자는 특정 연속적인 구간의 값을 일괄적으로 축소 또는 확대하여 변조
3) Injection : 공격자는 좁은 구간의 값을 임의의 값으로 변조
4) DoS : 공격자는 특정 연속적인 구간의 값을 같은 값으로 변조
5) replay : 공격자는 특정 연속적인 구간의 값을 이전 정상적인 구간에서 관측되었던 값으로 변조

3.1.3. 변조된 상태 메시지 생성

공격 데이터 합성은 제어 명령에 대해서만 수행되었으므로 차량 상태 메시지 또한 공격 시나리오에 따라 적절하게 변조되어야 한다. 본 논문에서는 이를 위해 제어 명령을 차량 상태 메시지로 변환하는 딥러닝 모델을 설계하고 이를 기반으로 상태 메시지를 변조한다. 이때, 제어 명령-차량 상태 쌍으로 이루어진 데이터의 규모가 충분히 클 때 정상 제어 명령에 따른 정상 차량 상태를 출력하는 모델 M_n(normal)과 변조 제어 명령에 따른 변조 차량

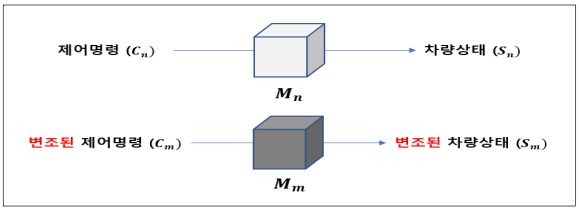

그림. 6. 모델 M_n과 모델 M_m의 입출력
Fig. 6. In/output of model M_n and M_m

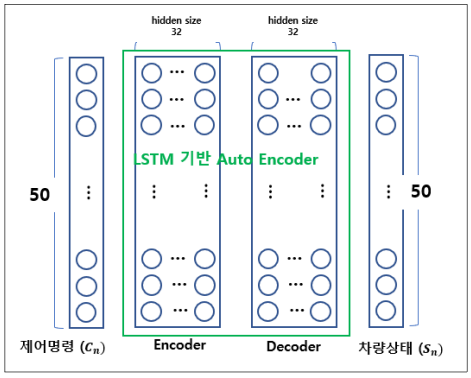

그림. 7. 딥러닝 기반 복원 모델 M의 구조
Fig. 7. Structure of DL model M

상태를 출력하는 모델 M_m(malicious)은 근사한다는 가정을 따르며 이는 그림 6과 같다.

그림 7은 딥러닝기반 복원 모델 M의 구조를 나타낸다. 딥러닝 모델은 입력으로 길이 50의 정상 제어 명령 C_n을 입력 받아 길이 50의 복원된 정상 차량 상태 S_n'을 출력하도록 학습되며 크기가 같은 LSTM 레이어로 이루어진 오토 인코더 형태로 설계되었다. 정상 데이터를 학습한 모델을 통해 복원한 차량 상태 S_n'과 실제 차량 상태 S_n는 그림 8에 나타나 있으며 제어 명령에 따른 차량 상태의 경향이 잘 드러나 있는 것을 확인할 수 있다.

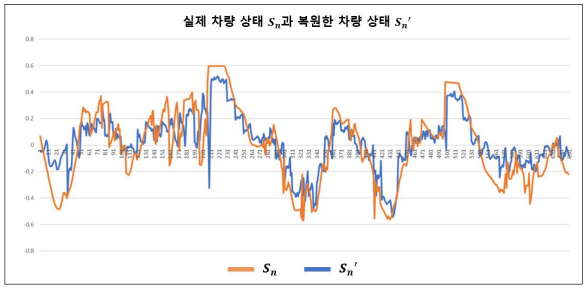

그림. 8. 실제 차량 상태 S_n과 복원 차량 상태 S_n'
Fig. 8. real vehicle status S_n and restored vehicle status S_n'

앞서 제시한 공격 시나리오 5개에 따라 합성된 제어 명령 데이터 C_m (C_{hijack}, C_{bias}, $C_{injection}$, C_{DoS}, C_{replay})는 정상 제어 명령-차량 상태 쌍을 통해 생성된 모델 M_n에 입력되어 변조된 제어 명령 S_m (S_{hijack}, S_{bias}, $S_{injection}$, S_{DoS}, S_{replay})를 생성하며 이를 통해 모델의 성능 평가를 수행한다.

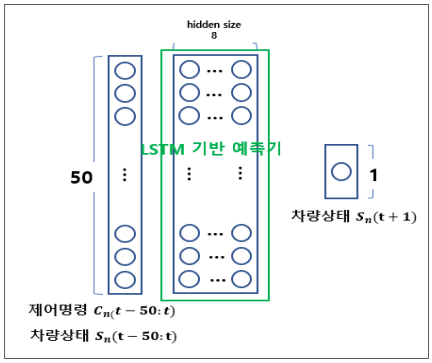

그림. 9. LSTM 기반 상태 예측 모델
Fig. 9. LSTM-based status prediction model

3.1.4. 이상 탐지 모델

이상 탐지 모델은 LSTM 기반의 상태 예측 모델로 그림 9와 같다. 이상 탐지 모델은 복원한 정상 차량 상태를 바탕으로 예측을 수행하는 모델과 수신한 차량 상태를 바탕으로 예측을 수행하는 모델로 나뉘어진다. 복원한 정상 차량 상태를 바탕으로 예측을 수행하는 예측 모델 P_n은 입력은 앞서 공격 데이터 생성에 사용했던 M_n의 출력 $S_n'_{(t-50:t)}$과 실제 제어 명령 $C_{n(t-50:t)}$을 사용하여 정상 차량 상태 $S_n'_{(t+1)}$를 예측한다. 수신한 차량 상태를 바탕으로 예측을 수행하는 모델 $P_{predicted}$은 수신한 차량 상태 $S_{m(t-50:t)}$과 실제 제어 명령 $C_{n(t-50:t)}$을 통해 복원 차량 상태 $S_{predicted}'_{(t+1)}$를 예측한다. 최종적으로 두 개의 예측 모델에서 출력된 $S_n'_{(t+1)}$과 $S_{predicted}'_{(t+1)}$의 차이가 사전 설정한 임계치를 초과할 경우 이상이라고 판단하며 출력된 $S_n'_{(t+1)}$과 $S_{predicted}'_{(t+1)}$의 예시는 그림 10과 같다. 이때, 회색 구역은 실제 변조가 수행된 구역을 나타낸다.

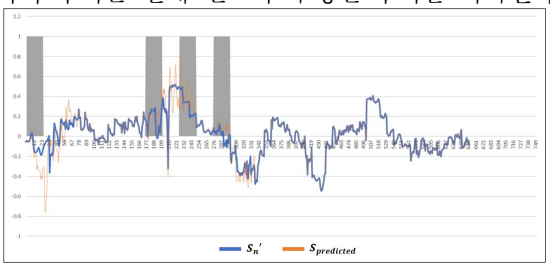

그림. 10. S_n'과 $S_{predicted}$의 차이 예시
Fig. 10. A Example of difference between $S_n'$and $S_{predicted}$

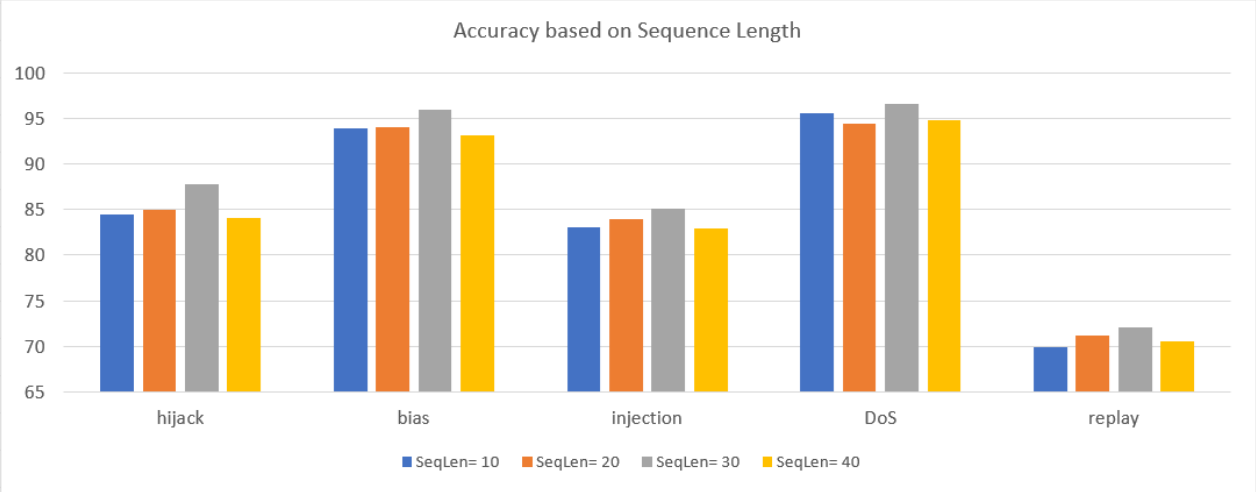

그림. 11. Sequence Length에 따른 정확도 비교
Fig. 11. Accuracy based Sequence Length

Ⅳ. 실험 결과 및 평가

4.1. 데이터셋 분할 및 평가 지표

평가를 위해 제어센터-원격차량에서 수집된 3,250개의 데이터셋을 사용하였으며 데이터셋 내 공격은 500개가 포함되어 있다. 또한, 5-fold를 통해 교차검증 하였으며 평가 지표로는 정확도를 사용하였다.

4.2. 공격 유형 별 정확도

표 2. 공격 유형 별 정확도 비교
Table. 2. Comparison of accuracy by attack type

공격 유형	정확도 (%)	편차 (%)
hijack	87.74	3.57
bias	95.99	2.12
injection	85.12	7.11
DoS	96.58	1.51
replay	72.14	9.23

표 2는 공격 유형 별 정확도 및 5-fold에 따른 편차를 나타낸다. 표 2에 따르면 가장 높은 정확도를 보이는 공격 유형은 DoS이며 그 뒤로 bias 공격이 뒤를 이었다. 가장 낮은 정확도를 보이는 공격 유형은 replay 공격으로 이전에 발생했던 패턴을 답습하여 공격하기에 모델이 정확하게 예측을 수행하지 못하는 것으로 추정된다. 또한 DoS, bias, hijack과 같이 연속적인 구간에서 발생한 공격의 경우 비교적 높은 정확도와 낮은 편차를 보이나 간헐적으로 좁은 구간에 공격이 발생한 경우에 낮은 정확도와 높은 편차를 보이는 것으로 나타났다.

4.3. 복원 모델 M의 매개변수에 따른 정확도

그림 11은 복원 모델 M의 매개변수 Sequence Length에 따른 정확도를 나타낸다. 매개 변수는 복원 모델과 예측 모델에서 공유하는 매개변수 sequence length에 대해 10부터 70까지 정확도를 바탕으로 비교하였다. 비교 결과, 매개변수 변화에 따라 큰 차이를 보이진 않으나 sequence length가 50일 때, 가장 높은 정확도를 보였으며 70일 경우에는 소폭 감소하는 것을 확인했다.

4.4. 임계치 ε에 따른 정확도

그림 12는 이상 탐지 임계치인 ε에 따른 정확도를 나타내며 1e-01부터 1e-04 범위 내에서 비교된다. 비교 결과, 임계치로 1e-03가 설정되었을 때 가장 높은 정확도를 보였으며 1e-03까지 낮아질 땐 정확도가 상승하는 것이 확인되었으며 1e-04로 더 낮아진 경우 정확도가 소폭 하락하였다. 이는 실험 환경 및 모델이 이상을 탐지하는 데 있어 적합한 임계치가 1e-03임을 나타낸다.

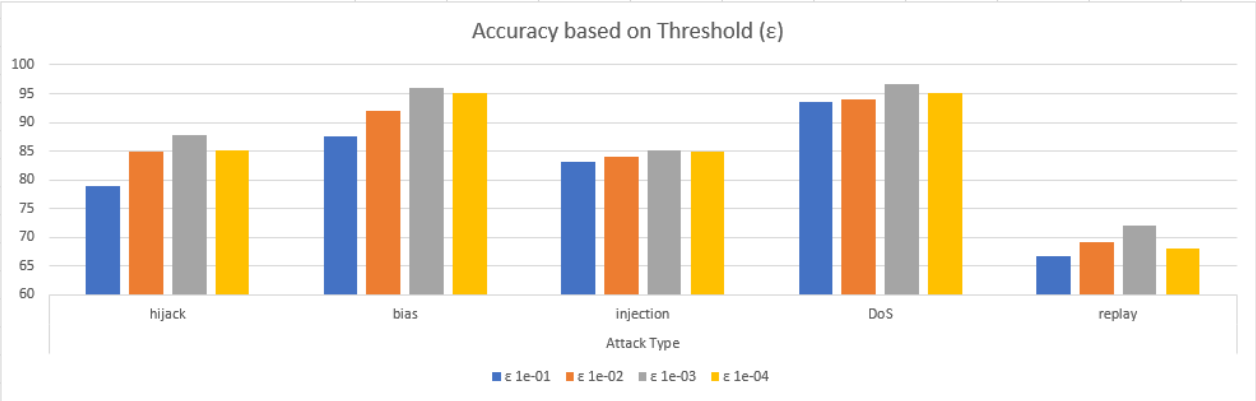

그림. 12. 임계치(ε)에 따른 정확도 비교

Fig. 12. Accuracy based threshold (ε)

V. 결 론

본 논문은 제어 센터-원격 차량으로 구성된 원격 제어 차량 운영 환경에서 중간자 공격으로 인해 발생할 수 있는 이상을 LSTM 기반으로 탐지하는 방법을 제안한다. 제안하는 방법은 원격 제어 차량 운영 환경에서 수집 가능한 제어 명령-차량 상태의 쌍을 활용하여 정상 차량의 상태 및 수신되는 차량의 상태의 차이가 일정 임계치를 초과하면 이상이라고 판단한다. 또한, 논문에서는 원격 제어 차량 운영 환경에서 발생할 수 있는 중간자 공격 데이터 합성을 위하여 딥러닝 기반의 복원 모델을 제시하였으며 모델을 통해 생성된 5가지 공격 시나리오에 따라 이상 탐지 모델을 평가하였다. 임계치 및 딥러닝 모델의 매개변수는 다양한 조건에서 실험되어 최적의 값으로 조정되었으며 제시한 5가지 공격 유형 중 replay를 제외한 다른 공격 유형에서 높은 성능을 나타내었고 replay 공격 유형에서는 다소 아쉬운 정확도를 나타내었다. 제안한 방법은 원격 제어 차량 운영 환경에서 오프라인 분석을 통해 기존 차량에서 공격이 발생하였는지 탐지하는데 활용할 수 있다. 또한, 데이터 샘플링 간격을 줄여 실시간에 가깝게 탐지하도록 모델 구조를 수정한다면 차량 운전자가 공격 발생을 인지하기 전에 탐지를 수행하는 등 예측에 가까운 기능 또한 수행할 수 있을 것이라 기대한다.

본 논문에서는 향후 연구로 데이터셋 생성 방법 고도화와 이상 탐지 모델 고도화를 제시한다. 원격 제어 차량 운영 환경에서 발생 가능한 공격에 대한 시나리오가 고려하여 공격 데이터셋을 생성하였으나 실제 세상의 공격을 완전히 모방할 순 없으므로 실제 공격에 대한 심도 있는 분석에 기반한 데이터셋 생성 방법이 요구된다. 두 번째로 현재 제안한 이상 탐지 모델은 조향에 대한 명령 및 상태 값만을 통해 이상을 탐지하고 있다. 하지만 다른 기존 연구에서 활용하는 시계열 특징 및 다른 명령 및 센서 값들을 활용하고 또한 딥러닝 모델 또한 최신 기법을 통해 고도화한다면 다소 낮은 탐지 정확도를 개선할 수 있을 것이라 기대한다.

References

[1] Top Automotive Cyberattacks in 2021 & 2022 (2021), accessed on Dec 13 2023, http://edgelabs.ai/blog/edge-computing-top-cyber-attacks-in-2021-2022-for-the-automotive-industry

[2] Greenberg, A. Wired, Hackers Remotely Kill a Jeep on the Highway－With Me in It (2015), accessed on Dec 13 2023, https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway

[3] UNECE, UN Regulation No. 155 - Cyber Security and Cyber Security Management System (2021), accessed on Dec 13 2023, https://unece.org/transport/documents/2021/03/standards/un-regulation-no-155-cyber-security-and-cyber-security.

[4] ISO, ISO/SAE 21434:2021, accessed on Dec 13 2023, https://www.iso.org/standard/70918.html

[5] VICONE, Two Tesla Hacks Triumph at Pwn2Own Vancouver 2023 (2023), accessed on Dec 13 2023, https://vicone.com/blog/two-tesla-hacks- triumph-at-pwn2own-2023

[6] Baker, W. et al. , Data Breach Investigations Report, VERIZON, 2021.

[7] Frankel, S.E. et al., A Guide to IEEE 802.11i; 0 ed., National Institute of Standards and Technology: Gaithersburg, 2007.
[8] Netcraft, Mutton, P. 95% of HTTPS Servers Vulnerable to Trivial MITM Attacks (2023), accessed on Dec 20, https://www.netcraft.com/blog/95-of-https-servers-vulnerable-to-trivial-mitm-attacks
[9] F5, 2022 State of Application Strategy Report (2022), accessed on Dec 20 2023, https://www.f5.com/go/report/2022-state-of-application-strategy-report
[10] Pateriya, P. et al., Analysis on Man in the Middle Attack on SSL. International Journal of Computer Applications in Technology, vol 45 (23), May 2012.
[11] Wang, C. et al., A Distributed Anomaly Detection System for In-Vehicle Network Using HTM. IEEE Access vol 6, p. 9091-9098 Jan 2018.
[12] Zhang, G. et al., Anomaly Detection for in-Vehicle CAN Bus Using Binary Sequence Whitelisting. Computers & Security 2023, vol 134, 103436, Nov 2023.
[13] Kim, T. et al., An Anomaly Detection Method Based on Multiple LSTM-Autoencoder Models for In-Vehicle Network. Electronics (Switzerland), vol 12(17), Aug 2023.
[14] HCRL, In-Vehicle Network Intrusion Detection Challenge, accessed on Dec 13 2023, https://ocslab.hksecurity.net/Datasets/datachallenge2019/car
[15] Stabili, D. et al., M. A Multidisciplinary Detection System for Cyber Attacks on Powertrain Cyber Physical Systems, Future Generation Computer Systems, vol 144, pp. 151 – 164, Mar 2023.
[16] Wang, L. et al., Anomaly Detection for Automated Vehicles Integrating Continuous Wavelet Transform and Convolutional Neural Network. Applied Sciences (Switzerland), 13.9(2023):5525, Apr 2023.
[17] Aksu, D. et al., MGA-IDS: Optimal Feature Subset Selection for Anomaly Detection Framework on in-Vehicle Networks-CAN Bus Based on Genetic Algorithm and Intrusion Detection Approach. Computers & Security vol 118, 102717, Apr 2022.
[18] Kim, Y.N. et al., T. Hidden Markov Model Based Anomaly Detection Method for In-Vehicle Network. Journal of Internet Services and Information Security 2022, 12, pp.115 – 125, May 2022.
[19] Donmez, T.C.M. Anomaly Detection in Vehicular CAN Bus Using Message Identifier Sequences, IEEE Access, vol 9, pp.136243 – 136252, Oct 2021.

백 의 준 (Ui-Jun Baek)

2018년: 고려대학교 컴퓨터정보학과 학사
2018년~현재: 고려대학교 컴퓨터정보학과 박사과정
<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

박 지 태 (Jee-Tae Park)

2017년: 고려대학교 컴퓨터정보학과 학사
2017년~현재: 고려대학교 컴퓨터정보학과 석사과정
<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

최 정 우 (Jeong-Woo Choi)

2022년: 고려대학교 컴퓨터정보학과 학사
2022년~현재: 고려대학교 컴퓨터정보학과 석사과정
<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석

김 명 섭(Myung-Sup Kim)

1998년: 포항공과대학교 전자계산학과 학사
2000년: 포항공과대학교 전자계산학과 석사
2004년: 포항공과대학교 전자계산 학과 박사
2006년: Dept. of ECS, Univ of Toronto Canada
2006년~현재: 고려대학교 컴퓨터정보학과 교수
<관심분야> 네트워크 관리 및 보안, 트래픽 모니터링 및 분석, 멀티미디어 네트워크