GAN-TREE 기반 차량 네트워크 침입 탐지 시스템에 관한 연구

김시온[1], 이선영*[2]
순천향대학교

gu100433@sch.ac.kr[1], *sunlee@sch.ac.kr[2]

A Study on the GAN-TREE Based Vehicle Network Intrusion Detection System

Kim Si On[1], Lee Sun Young*[2]
Department of Mobility Convergence Security, Soonchunhyang Univ.[1]

Department of Information Security Engineering, Soonchunhyang Univ.[2]

요 약

CAN(Controller Area Network)은 차량에 탑재된 전자 장치와 센서를 연결하는 네트워크로, 확장성과 비용의 효율성이 뛰어나 현대까지 사용되고 있다. 이러한 CAN 네트워크는 차량의 ECU 간 통신뿐만 아니라 V2X 기술을 통해 차량 외부의 모든 것과의 상호 작용도 가능하다. 이를 대상으로 하는 공격이 꾸준히 증가하고 있으며, 차량의 오작동을 유발하고 사람과 재산에 손실을 끼칠 수 있기 때문에 차량 네트워크 보안은 중요하다. 본 논문에서는 CAN-bus 를 보호하기 위해 적대적 생성 신경망을 통해 생성한 CAN 데이터와 실제 데이터를 사용한 Tree 기반의 침입 탐지 시스템을 연구하였다. 적대적 생성 신경망을 통해 생성된 데이터는 CAN ID, CAN data 를 포함하여 차량 네트워크에서 수집가능한 패킷들을 바탕으로 생성되었다. Tree 기반의 IDS 는 GAN 을 통해 생성된 가짜 공격 데이터와 실제 데이터를 학습하여, 기존 판별되지 않은 공격을 탐지하는 것이 가능해진다. 최종적으로 Decision Tree F1-score 0.993 과 Random Forest F1-score 0.996 의 결과를 얻었다.

Ⅰ. 서 론

자동차 산업이 발달하며 차량 내부 네트워크의 확대와 연결성 증가를 얻었으나, 차량 보안 취약점을 노린 공격 또한 늘었다. 특히 차량 내부 네트워크를 통한 공격이 증가했다.[1]

이러한 공격을 차단하기 위해 침입 탐지 시스템(IDS)가 연구되고 있다.[2,3,4]

다만, 일반 데이터만으로는 보고되지 않은 새로운 유형의 공격을 IDS 가 탐지하기 힘들다.

따라서, 적대적 생성 신경망(GAN)을 활용하여 새로운 데이터를 생성하고, 기존 데이터와 함께 학습하여 침입을 차단하는 방법을 제안한다.

본 연구에서는 총 4 가지 유형의 공격에 대한 CAN 데이터를 활용하여, 새로운 데이터를 생성하고 Decision Tree 와 Random Forest 모델을 사용한 IDS 로 공격을 탐지하는 방법을 소개한다.

Ⅱ. 본론

A. GAN(Generative Adversarial Networks)

적대적 생성 신경망(GAN)은 생성기(G)와 판별기(D)라는 2 개의 네트워크가 서로 적대적으로 학습하여 목적을 달성한다. 이때, 하나의 네트워크가 학습 중이라면, 다른 네트워크는 고정된 상태로 각각 따로 상태를 업데이트 한다. 생성기(G)는 판별기(D)가 판별하지 못하는 데이터를 생성하도록 D(G(z))값은 높게, 전체 확률은 낮아지도록 업데이트한다. [5]

B. Decision Tree & Random Forest

Decision Tree(결정트리, 의사결정트리)는 특정 조건에 따라 데이터를 구분하는 모델로, 한 분기에 두 개의 변수 영역으로 나뉜다.[6] 대규모 데이터 셋 환경에서 잘 작동하며, 모델에 대한 상황을 모두 관측 가능한 화이트 박스 모델이나, 과적합으로 인한 성능의 한계가 있다.

이러한 한계를 극복하고자 제안된 것이 Random Forest 이다.[7] Random Forest 는 같은 데이터에 대한 Decision Tree 를 여러 개 생성하여, 결과를 종합하는 방식을 사용한다. 각 Tree 들은 서로 조금씩 다른 특성을 가지게 하여, 예측 결과를 다르게 함으로서 일반화 성능을 향상시킨다

C. GAN 을 통한 IDS 학습 능력 향상

기존 데이터는 CAN ID, DATA[0~7], Label 로 구성되어 있다. 수집된 데이터는 공격이 없을 때의 주행 데이터와 Dos 공격을 받았을 때의 데이터, Fuzzy 공격을 받았을 때의 데이터, spoofing 공격(RPM/Gear)를 당했을 때의 데이터를 사용하였다.[8]

적대적 생성 신경망(GAN)을 통해 해당 데이터 중 DATA[0~7]의 Fake 데이터를 생성하여, IDS 학습 데이터로 사용한다.

생성된 데이터와 기존 데이터의 메시지는 그림 2.와 같다. 생성된 데이터는 15,226,945 개로 원본 데이터 17,558,462 개에서 모든 공격 메시지를 제외한 개수와 동일하다.

	Feature_0	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7
1	0.08880324	0.17443177	0.13310447	0.24113323	0.35639203	0.012477976	0.08052052	0.36060518
2	0.1222862	0.45531267	0.24218678	0.2574285	-0.14140299	0.050361075	0.3011894	0.12028416
3	0.34386373	0.04430922	0.25430024	0.08362891	0.47199002	0.3484449	0.39136803	0.19919431
4	0.2298136	0.11361652	-0.112424485	0.65448403	0.43591928	0.1893279	-0.00017510727	0.36441323
5	0.027062193	-0.016988857	0.21744731	0.14528666	0.23969145	0.4353206	0.36594018	0.09212493
6	0.65346086	0.52377546	0.35967636	-0.13871098	-0.13923162	-0.027657822	0.24744937	0.14336711
7	0.2816798	0.3779932	0.31915134	-0.07657975	0.44629905	0.43909332	0.24934758	0.3129217
8	0.062124383	0.0007058418	0.27960432	0.343511	0.399968	0.0066846553	0.29578465	-0.014580884
9	0.2752333	0.23260868	0.33371013	0.24440156	-0.071840346	0.42409685	0.34427494	0.16061573
10	0.46349323	0.37264055	0.008613679	0.12509134	0.41493413	0.36586407	0.03886286	0.12843072
11	0.23962192	0.101884164	0.37590575	0.009345587	0.38147825	0.07441381	0.22032963	0.22534104
12	0.38821843	0.15305974	0.5089818	-0.0758858	0.3373877	0.20155121	0.080223285	-0.17523284
13	0.6112715	0.20489183	0.10062709	0.004933674	0.37933734	0.46777752	0.5056973	0.22616662
14	0.4172639	0.14900608	0.11816244	0.37935063	0.34823558	0.2903459	0.0010053623	0.13381483

그림 1: GAN 을 통해 생성된 Fake 데이터

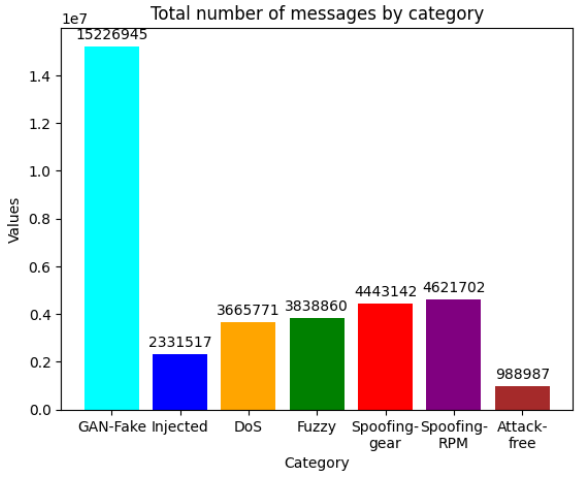

그림 2: 카테고리별 메시지의 총합

그림 3.은 본 연구의 제안 모델 흐름도이다. 최초 입력 데이터인 CAN 데이터를 사용하여, GAN 모델이 데이터를 생성한다. 원본 CAN 데이터와 Fake 데이터를 정규화 한 후, 중요 Feature 를 추출하는 과정을 거친다. 추출된 Feature 를 바탕으로 Decision Tree 모델과 Random Forest 모델을 사용하여 최종적으로 해당 데이터가 정상 CAN 메시지인지 아닌지 판단하는 구조이다.

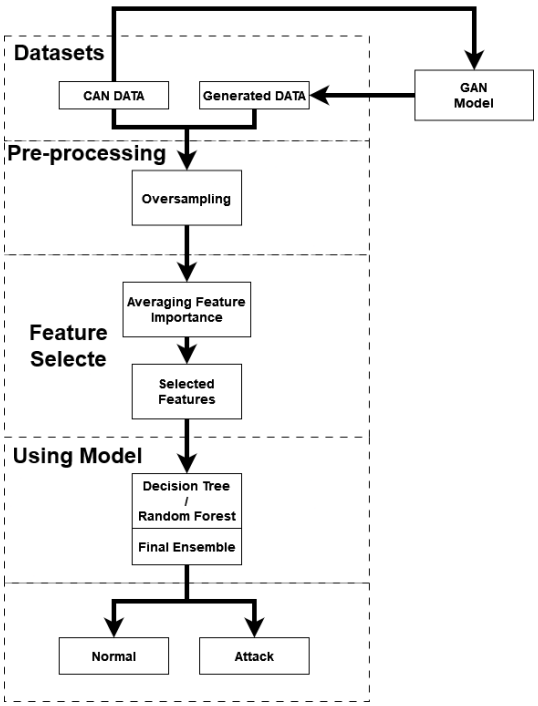

그림 3: 제안 모델의 흐름도

D. 탐지 및 성능

원본 데이터에서 테스트를 위한 데이터를 분리한 뒤, 학습 및 탐지를 진행하였다. 탐지 결과 정상 메시지와 공격 메시지를 판단하는 정확도, 정밀도, 재현율, F1-score 를 기준으로 성능을 확인하였다. GAN 을 통해 생성된 데이터와 실제 데이터를 학습에 사용한 결과, 일반 CAN 메시지와 공격 메시지를 구분하는 것이 가능하다는 결과를 얻었다.

	Decision Tree	Random Forest
Accuracy	0.993	0.996
Precision	0.993	0.996
Recall	0.993	0.996
F1-Score	0.993	0.996

Ⅲ. 결론

IDS 는 실시간으로 시스템 내의 이상 행위를 탐지할 수 있어, 자동차 보안 강화가 가능하다. 본 논문에서는 적대적 생성 신경망인 GAN 을 통해 CAN data 의 데이터를 생성하고, 원본 데이터와 함께 학습하는 Tree 기반의 IDS 를 제안 및 구현하였다.

그 결과, CAN data 를 기반으로 공격과 정상 메시지를 구분하는 실험에서, Decision Tree 모델에서는 0.993, Random Forest 모델에서는 0.996 의 F1-score 를 얻었다.

ACKNOWLEDGMENT

본 연구는 2021 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (NRF-2021R1A4A2001810)

참 고 문 헌

[1] Upstream.” 2023 Global Automotive Cybersecurity Report”. 2023

[2] H. M. Song, et al,” Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network” 2016 international conference on information networking (ICOIN). IEEE, 2016

[3] Müter, Michael, and Naim Asaj. "Entropy-based anomaly detection for in-vehicle networks." 2011 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2011.

[4] Dagan, Tsvika, and Avishai Wool. "Parrot, a software-only anti-spoofing defense system for the CAN bus." ESCAR EUROPE 34 (2016).

[5] Goodfellow, Ian, et al. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

[6] Kamiń ski, Bogumił , Michał Jakubczyk, and Przemysł aw Szufel. "A framework for sensitivity analysis of decision trees." Central European journal of operations research 26, 135-159. 2018

[7] Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." R news 2.3, 18-22. (2002)

[8] https://ocslab.hksecurity.net/Datasets/car-hacking-dataset