1D CNN을 활용한 CAN FD 악성 트래픽 분류

김란아, 남승우, 유경민, 김주성, 김명섭
고려대학교
{zufrieden7, nam131119, rudals2710, jsung0514, tmskim*}@korea.ac.kr

CAN FD Malware Traffic Classification by using 1D CNN

Ran-A Kim, Seung-woo Nam, Gyeong-min Yu, Ju-Sung Kim, Myung-sup Kim
Korea Univ.

요 약

본 연구는 CAN-FD 트래픽만 대상으로 한 1D-CNN 기반 침입 탐지 기법을 제안하고, 실제 차량 네트워크 환경에서 발생할 수 있는 다양한 공격 유형을 효과적으로 분류하였다. 제안된 모델은 CAN-FD 프레임의 메타데이터와 페이로드 정보를 직접 활용함으로써 높은 분류 정확도를 달성하였으며, worst-case latency 분석을 통해 실시간 침입 탐지에 적합한 경량 구조임을 입증하였다. 이러한 결과는 자율주행 및 ADAS 환경과 같이 엄격한 지연 요구사항을 갖는 차량 보안 시스템에 본 기법이 실질적으로 적용 가능함을 보여준다.

Ⅰ. 서 론

최근 차량 전자화 및 자율주행 기술의 발전으로 차량 내부 네트워크는 대용량 데이터와 고속 통신을 요구하는 복합 시스템으로 변화하고 있다. 이에 따라 기존 Classic CAN의 한계를 보완한 CAN-FD(Controller Area Network with Flexible Data-rate)가 ADAS 및 자율주행 차량 환경에서 널리 사용되고 있다.[1] 그러나 CAN-FD 역시 메시지 인증 및 암호화 기능이 기본적으로 제공되지 않아 메시지 위조, 재전송, 서비스 거부 공격과 같은 보안 위협에 취약하다.

이러한 문제를 해결하기 위해 다양한 침입 탐지 시스템이 제안되었으며, 최근에는 딥러닝 기반 접근 방식이 주목받고 있다. 기존 연구들은 CAN 또는 CAN-FD 트래픽을 이미지 형태로 변환하여 1D CNN 혹은 2D CNN을 적용하거나, wavelet 변환과 RGB 매핑을 통해 통계적 특징을 추출하는 방법을 사용하였다.[2][3] 이러한 기법들은 높은 탐지 정확도를 보이지만, 복잡한 전처리와 높은 연산 비용으로 인해 실시간 차량 환경에서의 적용에는 한계가 있다. 또한 대부분 평균 추론 시간에 초점을 맞추고 있어, 실시간 시스템에서 중요한 worst-case latency에 대한 분석은 충분하지 않다.

이에 본 연구에서는 CAN-FD 트래픽만을 대상으로 한 1D CNN 기반 침입 탐지 기법을 제안한다. 제안된 방법은 CAN-FD 프레임의 타임스탬프, arbitration ID, DLC와 같은 메타데이터와 페이로드 정보를 직접 활용함으로써 복잡한 전처리 없이도 높은 탐지 성능을 달성한다. 또한 입력 길이를 고정한 경량 구조를 통해 단일 입력 기준의 worst-case latency를 측정하고, 이를 통해 실시간 차량 보안 시스템에의 적용 가능성을 실험적으로 검증한다.

Ⅱ. 본론

2.1 데이터셋

본 연구에서는 CAN-FD 기반 차량 네트워크 침입 탐지 성능을 평가하기 위해 공개된 CAN-FD Intrusion Dataset을 사용하였다. 해당 데이터셋은 실제 차량 네트워크 환경에서 수집된 트래픽으로 구성되어 있으며, 정상 통신과 함께 다양한 공격 시나리오를 포함하고 있어 차량 보안 연구에 널리 활용되고 있다. 데이터셋에는 Classic CAN과 CAN-FD 메시지가 혼재되어 있으나, 본 연구에서는 프레임 구조와 페이로드 길이의 차이를 고려하여 CAN-FD 메시지만을 대상으로 분석을 수행하였다.

CAN-FD 메시지는 DLC(Data Length Code)가 8을 초과하는 프레임으로 정의하여 Classic CAN 메시지를 필터링하였다. 각 CAN-FD 프레임은 타임스탬프, arbitration ID, DLC, 그리고 최대 64바이트의 페이로드로 구성되며, 원본 CSV 파일에서 각 행은 하나의 CAN-FD 메시지를 나타낸다. 라벨은 정상(Normal)과 공격 트래픽으로 구분되며, 공격 유형은 Flooding, Fuzzing, Malfunction의 세 가지로 구성된다.

모델 입력 차원의 일관성을 확보하기 위해 페이로드 길이가 64바이트 미만인 경우 zero padding을 적용하였으며, 모든 프레임을 고정 길이의 1차원 벡터로 변환하였다. 이후 입력 값에 정규화를 적용하여 학습 안정성을 향상시켰다. 전체 데이터는 80%를 학습 데이터로, 20%를 테스트 데이터로 분할하여 4-class 분류 실험을 수행하였다. 이러한 데이터 구성은 CAN-FD 환경에서 발생 가능한 대표적인 공격 유형을 반영하며, 제안 모델의 성능을 평가하는 데 적합하다.

2.2 데이터 전처리

본 연구에서는 CAN-FD 트래픽을 1D CNN 모델의 입력으로 활용하기 위해 일관된 데이터 형식과 안정적인 학습을 위한 전처리 과정을 수행하였다. 먼저 원본 CSV 데이터에서 Classic CAN 메시지를 제거하고, DLC 값이 8을 초과하는 CAN-FD 프레임만을 선별하였다. 이후 각 프레임의 페이로드 길이를 64바이트로 고정하기 위해 페이로드가 64바이트 미만인 경우에는 zero padding을 적용하였다. 이를 통해 입력 데이터의 길이를 통일하고 모델 구조 설계의 복잡성을 최소화하였다.

각 CAN-FD 프레임은 타임스탬프, arbitration ID, DLC, 페이로드 정보를 순차적으로 결합한 1차원 벡터로 변환되었다. 또한 입력 값의 스케일 차이에 따른 학습 불안정을 방지하기 위해 각 필드에 대해 정규화를 수행하였다. 타임스탬프와 arbitration ID는 8비트 범위로 스케일링하였으며, DLC는 최대 길이를 기준으로 정규화하였다. 페이로드 바이트는 0부터

255 범위로 정규화하여 모든 입력 특징이 동일한 범위에 위치하도록 하였다.

2.3 1D CNN 모델 아키텍처

본 연구에서는 CAN-FD 트래픽의 순차적 특성과 바이트 단위 구조를 효과적으로 학습하기 위해 1차원 합성곱 신경망(1D Convolutional Neural Network, 1D CNN)을 기반으로 한 침입 탐지 모델을 설계하였다. CAN-FD 프레임은 시간 순서에 따라 연속적으로 전송되는 바이트 데이터로 구성되며, 이러한 특성은 공간적 관계를 가정하는 2D CNN보다 1D CNN 구조에 더 적합하다. 따라서 본 연구에서는 복잡한 이미지 변환이나 차원 확장을 수행하지 않고, 원본 프레임 정보를 유지한 채 직접적인 특징 추출이 가능한 1D CNN을 채택하였다.

제안된 모델은 입력으로 고정 길이의 1차원 벡터를 받아 두 개의 1D 합성곱 계층을 통해 특징을 추출한다. 첫 번째 합성곱 계층은 비교적 작은 커널 크기를 사용하여 국소적인 패턴을 학습하며, 두 번째 계층에서는 보다 추상적인 고수준 특징을 추출한다. 각 합성곱 계층 이후에는 ReLU(Rectified Linear Unit) 활성화 함수를 적용하여 비선형성을 부여하고, 기울기 소실 문제를 완화함으로써 안정적인 학습을 유도하였다. 이후 풀링 계층을 통해 특징 맵의 길이를 축소함으로써 연산량을 감소시키고 모델의 일반화 성능을 향상시켰다.

합성곱 및 풀링 계층을 통해 추출된 특징은 평탄화(flatten) 과정을 거쳐 완전 연결 계층으로 전달된다. 완전 연결 계층은 추출된 특징을 종합하여 최종 분류에 적합한 표현으로 변환하며, 출력 계층에서는 Softmax 함수를 통해 Normal, Flooding, Fuzzing, Malfunction의 네 가지 클래스에 대한 확률 값을 산출한다. 본 모델은 비교적 단순한 구조를 유지함으로써 높은 분류 정확도와 함께 낮은 추론 지연 시간을 달성하도록 설계되었다. 이를 통해 제안된 1D CNN 기반 침입 탐지 모델은 실시간 처리가 요구되는 차량 내부 네트워크 환경에 적합한 경량 IDS로서의 적용 가능성을 실험적으로 입증하였다.

2.4 실험 및 평가

본 연구에서는 제안된 1D CNN 기반 침입 탐지 모델의 성능을 이진 분류 및 다중 분류 환경에서 평가하였다. 먼저 이진 분류 실험에서 1D CNN 모델은 1 epoch 학습 시 99.90%의 정확도를 보였으며, 2 epoch 이후에는 100%의 분류 정확도를 달성하였다. 이는 CAN-FD 트래픽에서 정상과 공격 패턴이 구조적으로 명확히 구분됨을 의미한다.

다중 분류 실험에서는 Normal, Flooding, Fuzzing, Malfunction의 4-class 분류를 수행하였으며, 1 epoch 학습만으로도 99.96%의 높은 정확도를 기록하였고, 50 epoch 학습 시에는 100%의 분류 정확도를 달성하였다. 반면, 페이로드 정보만을 사용한 4-class 분류 실험에서는 25 epoch 학습 기준 약 86.58%의 정확도를 보였으며, 이는 공격 탐지에 있어 메타데이터 정보의 중요성을 보여준다.

또한 실시간 적용 가능성을 평가하기 위해 추론 지연 시간을 측정한 결과, 평균 latency는 0.2275 ms로 매우 낮았으며, worst-case latency는 182.3739 ms로 측정되었다. 이러한 결과는 제안된 모델이 높은 탐지 성능과 함께 실시간 차량 보안 시스템에 적용 가능함을 입증한다.

	1epoch Accuracy(%)	2epoch Accuracy(%)
1D CNN Binary	99.90	100

표 1 1D CNN 이진 분류 정확도

Ⅲ. 결론

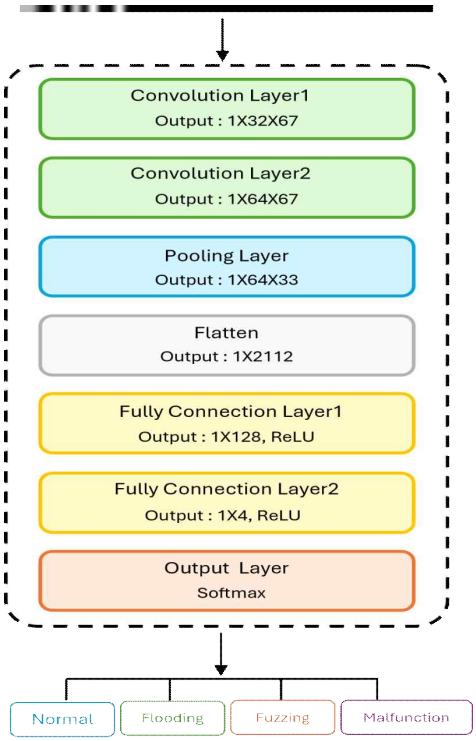

그림 1 1D CNN 아키텍처

	1epoch Accuracy(%)	50epoch Accuracy(%)
1D CNN 4Class	99.96	100

표 2 1D CNN 4-Class 분류

본 연구에서는 CAN-FD 트래픽을 대상으로 한 1D CNN 기반 침입 탐지 기법을 제안하고, 높은 분류 정확도와 낮은 추론 지연 시간을 통해 실시간 차량 보안 시스템에의 적용 가능성을 입증하였다. 실험 결과, 제안된 모델은 이진 및 4-class 분류 환경에서 모두 우수한 성능을 보였으며, 특히 메타데이터를 포함한 전체 프레임 정보를 활용할 경우 탐지 성능이 크게 향상됨을 확인하였다. 또한 worst-case latency 분석을 통해 자율주행 및 ADAS 환경에서도 활용 가능한 경량 침입 탐지 모델임을 검증하였다.

ACKNOWLEDGMENT

이 논문은 2023년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원과(00235509, ICT융합 공공 서비스•인프라의 암호화 사이버위협에 대한 네트워크 행위기반 보안관제 기술 개발), 정부(중소벤처기업부)의 재원으로 중소기업기술정보진흥원(TIPA)의 창업성장기술개발사업(TIPS)사업의 지원을 받아 수행된 연구임(RS-2025-25466990, 5G/6G 네트워크 Cross-domain Observability Engineering Orchestrator 기술 및 표준 개발).

참 고 문 헌

[1] Bosch, R. (2012). CAN with flexible data-rate (CAN FD) specification. Robert Bosch GmbH, Version 1.0.

[2] Aung, Y. L., Park, J., & Hong, K. S. (2025). CANDIDS: CAN/CAN-FD deep learning-based intrusion detection systems. IEEE Transactions on Intelligent Transportation Systems

[3] Zhou, J., Zhang, Z., & Wang, Y. (2020). A novel intrusion detection method for CAN bus based on 1D convolutional neural network. IEEE Access, 8, 182435 - 182445.