한모코(HMC) 학습일지

이름	서영재	학번	20195178
날짜	2022.11.21.		

학습 계획

- 실험 진행 및 방법론 정리
<CAN 프레임 간격 기반 침입 탐지 방법론>

학습 내용

요약)

- CAN bus의 경량화된 CAN ID 시퀀스와 TIME 인터벌 시퀀스 feature 기반의 침입 탐지 방법론을 제안함
- 제안하는 방법론은 높은 탐지 성능을 나타내며, 4ms의 실시간 탐지 시간을 확보함

추가 실험)

최적의 시퀀스 길이를 확보하기 위해 시퀀스 길이 변화에 따른 탐지 성능 비교

1. 서론

최신 차량들은 CAN bus 탑재를 필수로 하고 있지만 CAN bus는 보안 기능 없이 설계되어서 여러 취약점이 있다. CAN bus상에서 생성 및 전송되는 네트워크 트래픽의 크기가 점차 커짐으로써 많은 양의 네트워크 트래픽을 처리하는 차량 침입 탐지 모듈은 경량화가 요구된다. 기존의 연구들은 실시간성을 요구하므로 딥러닝 알고리즘을 사용하여 IVN(In-Vehicle Network) 침입 탐지에서 단점으로 작용하였다. 따라서 우리는 IVN에서 적용 가능한 경량화된 Sequence feature 및 머신러닝 알고리즘을 사용하여 침입 탐지 방법론을 제시하였다.

2. 배경지식

CAN Data Frame

CAN은 ECU 또는 차량 내 탑재된 다양한 센서 간의 통신을 제공하는 직렬 통신 프로토콜이며, 차량 내부 네트워크로써 센서와 ECU 사이의 중요 정보들을 실시간으로 통신하기 위해 최대 1Mbit/s의 비트 전송률을 가진다. CAN 프로토콜은 Data frame, Remote frame, Error frame, Overload frame 총 4개의 frame으로 구성되며 우리 연구에서는 Data Frame만을 사용하여 침입 탐지에 적용하였다.

공격 시나리오

우리 연구에서는 dos, fuzzy, malfunction 세가지 공격 시나리오를 적용하였다.

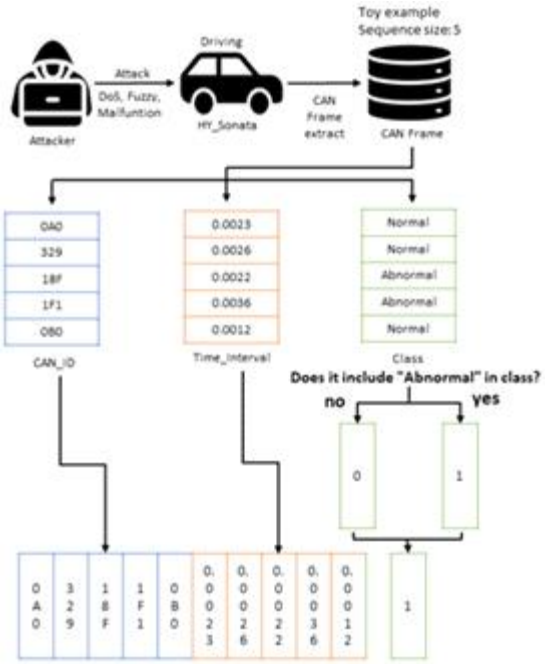
- dos attack : 공격자는 CAN Bus 장악을 위해 충돌 시 높은 우선순위를 가지도록 CAN ID를 '000'으로 설정하여 CAN Bus에 주입한다.
- fuzzy attack : 공격자는 랜덤하게 CAN ID를 선택하고 CAN 메시지를 반복적으로 주입하여 무차별 공격을 수행한다. 공격자는 해당 공격을 통해 차량의 이상 반응을 일으킬 수 있다.
- malfunction attack : 차량의 특정 기능 제어를 위해 공격자는 미리 파악한 CAN ID 및 Data Field를

설정하여 CAN Bus에 공격을 수행한다.

3. 방법론

데이터 전처리

각 패킷당 CAN_ID, DLC, TimeStamp, class를 feature로 가지고 있으며, class는 해당 패킷이 공격 패킷의 유무를 나타내며, Normal은 정상, Abnormal은 공격을 의미한다. 본 연구에서는 CAN_ID와 Time interval를 사용할 예정이며 Time interval은 n번째 패킷과 n-1번째 패킷의 Time stamp차이를 의미한다. 데이터를 특정 시퀀스 크기만큼 추출하고, CAN_ID와 Time interval을 나란히 연결하여 하나의 벡터로 변환하였다.



데이터 라벨링

데이터 전처리 과정에서 추출한 고정 길이의 Sequence Data에서 1개라도 공격 CAN 메시지가 포함되어 있으면 해당 데이터의 class를 1, 그렇지 않으면 0으로 라벨링 하였다.

모델 학습

Random Forest 모델을 사용하였다. 성능의 향상보다는 성능의 차이를 비교하기 위해 하이퍼 파라미터는 디폴트 값으로 설정하였다.

4. 실험 및 평가

데이터셋

HCRL(Hacking and Countermeasure Research Lab)에서 제공하는 "Survival Analysis Dataset for automobile IDS" 데이터셋을 사용하였다. 주행 중인 Hyundai Sonata 차량에서 추출된 데이터셋으로 DoS, Fuzzy, Malfunction 공격이 적용 되었으며 공격이 명확하게 드러나는 DoS, Fuzzy, Malfunction 공격에 대한 데이터를 사용하였다. Sequence 길이는 최적화를 위해 2부터 50까지 설정하고, CAN Sequence 및 Time interval Sequence feature에 성능 비교를 통해 최적의 Sequence 길이 도출하였다.

성능지표

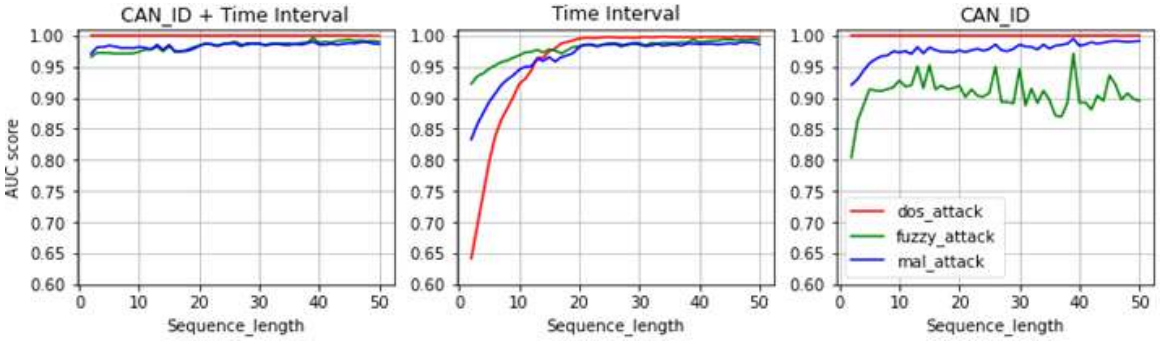
모델 학습에서 사용된 모델의 성능을 평가하기 위해 AUC(Area Under Curve) metrics를 이용하였다. Threshold에 따른 FPR(False Positive Rate), TPR(True Positive Rate)을 통한 ROC-curve의 아래 면적인 AUC metric을 사용하여 성능을 확인하였다.

실험환경

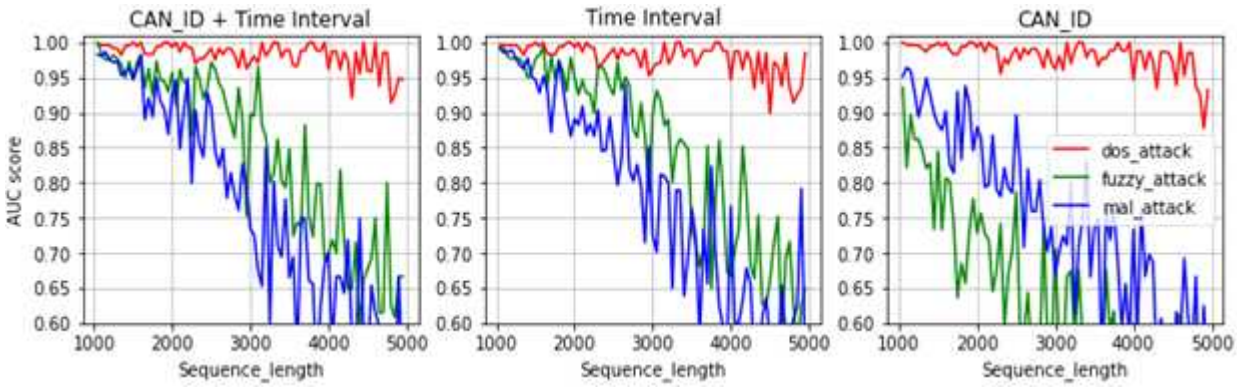
Intel (R) Core (TM) i5-12400F CPU, NVIDIA GeForce RTX 3060, RAM 24.0GB, Windows 10 Pro 64-bit 운영체제 환경에서 수행하였다. 머신러닝 알고리즘 적용을 위해 Python 3.8 IDE를 사용하였고, sklearn, numpy, pandas, matplotlib 등의 라이브러리를 사용하였다.

실험결과

Time interval Sequence, CAN ID Sequence feature들에 대한 AUC Score를 아래 그림에 나타내었다. 본 연구에서는 sequence길이를 2부터 50까지 지정하여 각 길이에 따른 데이터를 구축하였다.



추가로 sequence 길이가 성능에 있어 항상 도움이 되는지 확인하기 위해 1000부터 5000의 sequence 길이를 설정하여 실험을 진행하였다.



Sequence의 길이가 증가에 따른 AUC score의 상관 관계 확인을 위해 Sequence 길이를 1000~5000 까지 변화 간격을 50으로 설정하여 성능 비교하였다. 길이가 1000 이상일 경우 AUC score가 1.0에 수렴하지 않고 기존보다 낮아짐을 확인할 수 있었다. 낮은 Sequence 길이가 길어질수록 높은 AUC score를 가지지만, 길이가 1000이상인 Sequence는 feature로써의 기능을 정상적으로 수행하지 못한다는 사실을 알 수 있었다.

5. 결론

낮은 보안성을 가지는 CAN bus에 대한 침입탐지 방법론 제시하였다. 경량화된 feature로써 CAN ID Sequence 및 Time interval Sequence를 구성하고 머신러닝 알고리즘인 Random Forest를 사용하여 높은 탐지성능 확인하였다. Sequence 길이와 탐지 성능의 수렴 여부를 확인하기 위해, Sequence 길이 변화에 따른 침입탐지 성능 비교하였다. 향후 연구 에서는 자율주행 차량에서의 경량화된 침입탐지 연구를 수행할 예정이다.