AN INTELLIGENT SECURED FRAMEWORK FOR CYBERATTACK DETECTION IN ELECTRIC VEHICLES' CAN BUS USING MACHINE LEARNING

2022.08.23 서영재

한림대학교 소프트웨어융합대학
(Data-driven Cybersecurity Research Lab)

HALLYM UNIVERSITY

목차

요약

HALLYM
UNIVERSITY

요약

■ CAN Bus

□ 전기 자동차의 CAN Bus는 차량 내 네트워크(IVN) 통신을 위한 legacy 프로토콜 역할을 함

□ 특징 : 실시간의 단순성, 견고성 및 적합성

□ 프로토콜 자체에 메시지 인증 메커니즘이 없어 다양한 사이버 공격에 취약

- 공격자가 네트워크에 침투할 수 있는 상황을 조성

요약

- **본 논문은 CAN traffic에서 단일 클래스 지원 벡터 머신을 기반으로 한 새로운 효과적인 이상 탐지 모델을 제안한다**
  - 오프라인 훈련에서 가장 정확한 구조를 찾기 위해 수정된 bat 알고리즘으로 알려진 개선된 알고리즘 사용

- **제안된 모델 효과 평가 방법**
  - CAN traffic은 정상 작동중인 수정되지 않은 라이선스 전기 차량으로부터 기록하여 공격없이 각 메시지와 해당 발생 빈도에 대한 데이터셋을 생성
  - Isolation Forest와 고전적인 단일 클래스 지원 벡터 머신과 같은 서로 다른 두 개의 유명한 CAN Bus 이상 탐지 알고리즘을 비교
    - 성능과 제안된 방법의 우월성을 측정하기 위해, 공격을 포함하는 테스트셋에 각 방법에 대한 **ROC**를 제공

- **실험 결과**
  - 제안된 방법이 서로 다른 두 개의 알고리즘과 비교한 이상 탐지에 대해 가장 높은 양성 비율(TPR)과 가장 낮은 거짓 양성 비율(FPR)을 달성했다고 나타냄
  - 다른 데이터셋에도 적용될 수 있음을 보여주기 위해 CAN BUS traffic 이상 탐지 범위에서 최근 인기있는 두 개의 공개 데이터셋을 사용
    - 다른 CAN 데이터셋으로 적응할 수 있는 각 메시지 ID 및 데이터 필드의 의미로부터 제안된 방법의 독립성을 입증

요약

■ Bat 알고리즘(박쥐 알고리즘)

- □ 2010년 Xin-She Yang이 개발한 method
- □ 박쥐가 초음파를 활용하는 방법에서 영감을 받아 만들어진 전역 최적화 방법
- □ 박쥐는 초음파 이용한 반향 탐지 기술을 통해 주변의 3차원 공간을 구축할 수 있음
  - 이는 도플러 효과로도 활용하여 대상의 유형과 이동속도까지 탐지가 가능

Bat Algorithm

Objective function $f(\mathbf{x})$, $\mathbf{x} = (x_1, ..., x_d)^T$
Initialize the bat population $\mathbf{x}_i$ $(i = 1, 2, ..., n)$ *and* $\mathbf{v}_i$
Define pulse frequency f_i *at* $\mathbf{x}_i$
Initialize pulse rates r_i *and the loudness* A_i
while ($t <$ *Max number of iterations*)
Generate new solutions by adjusting frequency,
and updating velocities and locations/solutions [equations (2) to (4)]
  if $(rand > r_i)$
  Select a solution among the best solutions
  Generate a local solution around the selected best solution
  end if
  Generate a new solution by flying randomly
  if $(rand < A_i \ \& \ f(\mathbf{x}_i) < f(\mathbf{x}_*))$
  Accept the new solutions
  Increase r_i *and reduce* A_i
  end if
Rank the bats and find the current best $\mathbf{x}_*$
end while
Postprocess results and visualization

요약

■ ROC(Receiver operating characteristic)

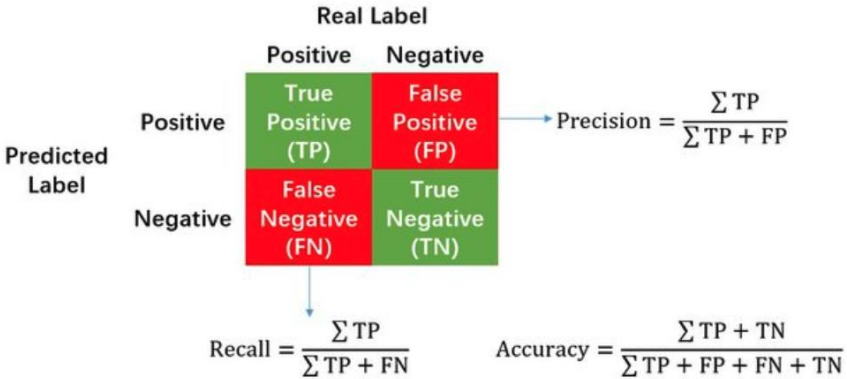
□ 분류 검정의 민감도 대를 도표화하여 모형 예측의 정확도를 평가하기 위한 방법

□ TPR : True Positive Rate (= 재현율 / 민감도, Recall)

- 양성인 케이스에 대해 양성으로 잘 예측한 비율.(Ex. 암환자를 진찰해서 암이라고 진단)

□ FPR: False Positive Rate (=1-TNR, false accept rate)

- 음성인 케이스에 대해 양성으로 잘못 예측한 비율. (Ex. 정상환자를 암이라고 진단)



서론

HALLYM
UNIVERSITY

서론

■ ECUs(Electronic Control Units)

□ 현대 전기 차량은 정교한 소프트웨어 구성요소에 의해 제어되는 전자 제어 장치(ECUs)로 알려진 많은 하드웨어 모듈로 구성되어 있음

□ 다양한 센서에 의해 측정된 데이터를 읽고, 보행자 감지, 경로 계획 및 자동 주차 등 다양한 목적을 위한 처리를 수행함

□ 센서와 액츄에이터 값을 차량 내 네트워크(IVN) 프로토콜을 통해 다른 ECUs에 전송되어 하드웨어와 소프트웨어의 하위모듈이 매우 복잡한 네트워크를 형성함

- CAN, CAN FD, LIN, FlexRay 및 MOST(Media Oriented Systems Transport)

서론

■ 차량내 네트워킹(In-vehicle networking, IVN) 프로토콜

- □ 현재는 보급형 차량에도 수십 개의 ECU가 기본적으로 포함되어 있어, 차량 주변으로 라우팅 해야 하는 신호가 수백 개 또는 수천 개에 이름

- □ 현대식 차량 주변 모든 전력 및 데이터 신호를 라우팅 할 수 있는 와이어링 하니스와 연결된 커넥터들은 너무 커지고, 무겁고 복잡해지고, 가격이 높아진다면 이것은 실제 적용할 수 없는 것이 되어 버림

- □ 각각의 와이어 하니스는 차량 제품군 내 단일 모델에 완전히 맞춤화 되어야 함

> IVN은 이러한 문제를 해결함

서론

■ CAN Bus

□ 항공우주, 농업, 의료기기, 심지어 일부 가정 및 상업용 가전제품과 같은 다른 산업에서 다양한 응용 분야를 찾아 단순한 자동차 네트워크 이상의 분야에 적용되어 옴

□ Ethernet은 다음과 같은 이유로 차량 내 네트워크 통신을 위한 CAN Bus를 완전히 대체할 수 없음

1) CAN Bus는 어려운 실시간 환경에 완벽하게 적용되도록 설계되었으며, 최소한의 시간 지연으로 통신을 보장함
2) CAN Bus 프로토콜에는 우선순위가 낮은 메시지가 높은 우선순위의 메시지를 방해하지 않는 우선순위 지정 방법이 있음
   – 엔진 컨트롤 또는 에어백 컨트롤 메시지 > 도어 또는 실내 온도 조절 메시지
3) CAN Bus 프로토콜을 차량 내 네트워크 통신의 backboned로 모든 현대 차량에서 사용됨

완전히 다른 프로토콜을 대체하려면?

전체 차량 네트워크 아키텍처를 재설계하고 CAN프로토콜을 기반으로 실행되는 차량 소프트웨어를 엄청나게 변경해야 함

> 다른 프로토콜은 CAN Bus의 역할과 기능을 완전히 대체하지 않고, 오히려 CAN Bus를 증가시킬 것임

서론

■ CAN Bus

- □ Robert Bosch GmbH가 CAN Bus 프로토콜의 발명하는 동안, 차량은 외부 환경과 통신하지 않는 고립된 환경이라고 간주됨
- □ 따라서 CAN Bus는 설계상 데이터 암호화와 메시지 인증을 포함하는 인증 및 보안 기능의 부족으로부터 고통받음
  - 상대방이 전송 제어 프로토콜/인터넷 프로토콜(TCP/IP)와 같이 다른 프로토콜을 사용할 때보다 더 쉽게 네트워크에 침투하고 악의적인 활동을 시작할 수 있는 길을 열어줌

- □ 예를 들어, 효과적인 메시지 인증 방법의 부족으로 인해 공격자는 악의적인 메시지를 주입하고 공격을 재생하여 ECU를 손상시킬 수 있음

> **데이터마이닝 기술의 발전은 이러한 공격의 유형은 연구자들에 의해 모든 비정상적인 통신 트래픽 활동을 탐지하고 무시할 수 있는 방식으로 해결됨**

서론

□ 최근 현대 차량은 폐쇄 루프 시스템으로 간주될 뿐만 아니라 외부 세계와 여러 유형의 통신 수행

- 폐쇄 루프 시스템 : 출력시스템의 입력은 입력에 따라 다른 시스템
  - 폐쇄 루프 시스템은 입력과 출력의 차이인 오류 신호를 생성

- 공격자는 OBD-II 포트 차량 내, 근거리 무선 액세스, 장거리 무선 액세스, 텔레매틱스 제어 장치(TCU) 등 다양한 내부 및 외부 인터페이스를 통해 CAN 트래픽에 악의적인 메시지를 침투하고 주입할 수 있음
  - 근거리 무선 액세스 (예 : 블루투스)
  - 장거리 무선 액세스 (예 : Wi-Fi)

- 예를 들어, OTA(Over-The-Air) 업데이트를 채택하여 ECU를 원격으로 재프로그래밍할 수 있으므로 차량 소유자와 딜러에게 더 많은 편안함과 편리함을 제공할 수 있습니다. 하지만, 이러한 인터페이스는 공격자가 악성 메시지를 사용하여 ECU를 손상시키는 데 도움이 될 수 있는 더 많은 원격 공격 표면을 도입함
  - OTA 업데이트 : 무선 소프트웨어 업데이트로, 무선 통신을 활용하여 소프트웨어를 업데이트
    - 차량의 성능 및 안정성을 높여줌

서론

■ 이상 탐지

□ 데이터 셋에서 예상한 정의된 동작을 따르지 않는 패턴을 찾는 문제

□ CAN 버스 프로토콜에서의 이상 탐지

- 기계 학습(ML) 알고리즘을 사용하여 ECU 간의 통신 트래픽을 모니터링하고 트래픽의 이상 동작을 식별하는 프로세스

□ 데이터가 CAN 버스 프로토콜에서 암호화되지 않기 때문에 공격자는 각 CAN 패킷을 해석하는 리버스 엔지니어링 절차를 수행하여 재생 공격을 시작하고 악성메시지를 네트워크에 주입할 수 있다. (침입 기반 공격 수행)

- 위 목표를 달성하기 위해 공격자는 버스의 모든 메시지에 사용되는 중재 메커니즘을 이기기 위해 매우 높은 빈도로 메시지를 보내야 함

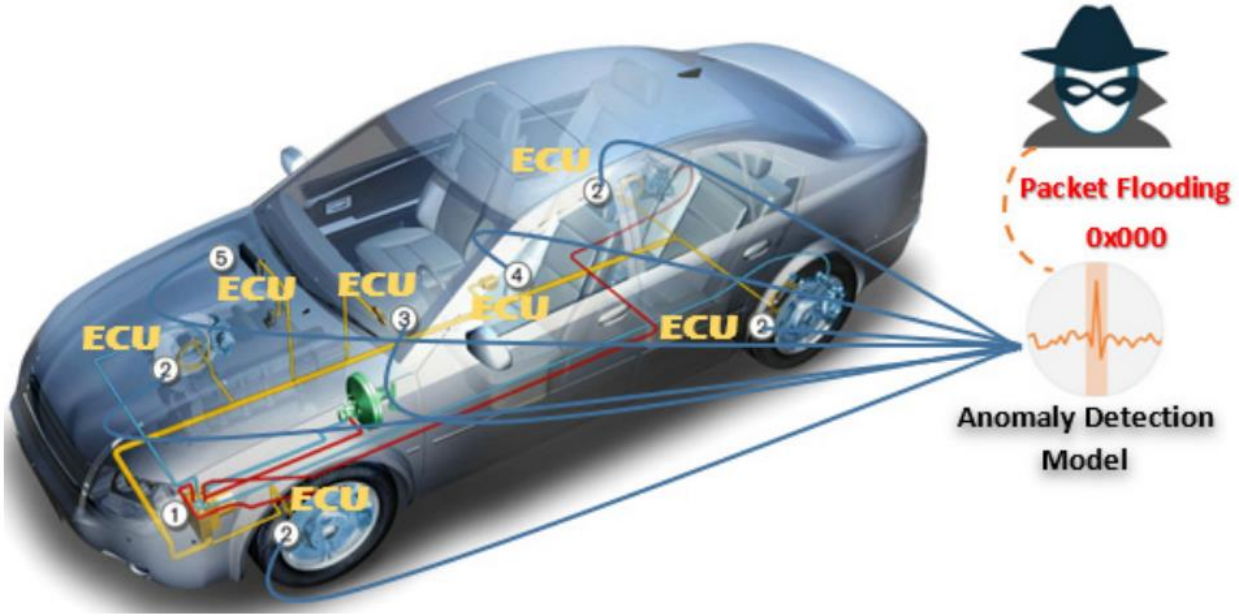
□ 메시지 삽입 절차

- CAN 버스 프로토콜에서 이상 탐지 방법을 개발하여 탐지할 수 있는 통신 트래픽에서 일부 이상 동작을 생성

■ 본 논문에서는 차량 네트워크 프로토콜에서 비정상적인 트래픽을 감지할 수 있는 예측 모델을 제안

서론

■ 차량 네트워크 트래픽에 대한 이상 탐지 방법의 적용



관련 연구

HALLYM
UNIVERSITY

MOCSVM

- 제안된 모델은 최고의 보안성과 정확성을 제공하기 위해 수정된 단일 클래스 지원 벡터 머신(OCSVM)을 기반으로 구성
  - OCSVM : one class support vector machine

- OCSVM에서 SVM 모델은 "nomal" 클래스라고 하는 클래스가 하나만 있는 데이터에 대해 학습됨
  - 정상적인 경우의 속성을 강조하고 이러한 속성에서 정상적인 사례와 다른 예제를 예측할 수 있다.
  - 훈련 예제가 부족한 경우, 비정상적으로 정의하기 때문에 이는 비정상 감지에 매우 실용적이다.

- CAN 버스 데이터 세트의 높은 복잡성과 비선형성으로 인하여 수정된 박쥐 알고리즘(MBA)라고 하는 새로운 최적화 알고리즘을 사용하여 제안
  - 오탐률을 줄이고 비정상적인 메시지 탐지의 전체 적중률을 개선

- 제안된 모델은 수정되지 않은 면허 차량에서 수집한 실험 데이터를 사용하여 높은 정확도와 만족스러운 성능을 보임

- 독립성 증명
  - CAN 버스 이상 탐지 영역에서 유명한 두 개의 다른 데이터 세트를 사용하여 제안한 방법 수행

연구 기여 요약

1. **차량 내 네트워크 통신 CAN 버스 프로토콜을 강화하기 위해 고급 머신 러닝을 기반으로 하는 지능형 이상 탐지 모델을 제안**

2. **박쥐 알고리즘을 기반으로 수정된 새로운 단일 클래스 지원 벡터 머신(MOCSVM) 개발**
   - 제안된 박쥐 알고리즘은 사이버 공격에 대한 모델의 효율성과 성능을 극대화하기 위해 이상 탐지 모델 매개변수를 최적으로 조정

3. **조기 수렴을 피할 때 인구 다양성을 증가시키기 위해 박쥐 알고리즘에 대한 효과적인 2단계 수정 접근 방식을 도입**
   - 유전자 알고리즘에서 차용한 교차 연산자와 돌연변이 연산자를 기반으로 구성되며, 알고리즘이 전체 검색 공감에서 최적의 솔루션을 찾는데 도움이 됨

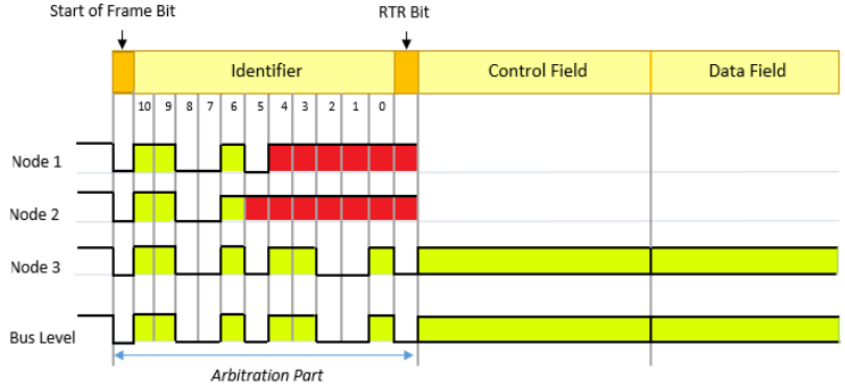
CAN BUS 프로토콜

HALLYM
UNIVERSITY

CAN 버스 프로토콜과 요구 사항

□ 실시간 시스템 기반 요구사항을 충족하기 위해
- CAN 버스 프로토콜의 각 메시지에는 메시지 우선순위를 정의하는 데 사용
- 모든 ECU에서도 사용되는 고유 식별자(ID) 프레임이 할당됨

□ 메시지 식별 값이 낮고 우선 순위가 높으면 버스 액세스 권한을 얻음



바이너리

Node 1 : 11001011111

Node 2 : 110011111111

Node 3 : 110010110010

- 3개의 노드가 동시에 메시지 전송을 시도하는 상황
- 버스 충돌을 방지하기 위해 ID가 가장 낮은 노드(Node 3)가 정보를 전송함
- 공격자는 이 기능을 악용하여 ID가 가장 낮은 악성 메시지를 매우 높은 빈도로 전송하여 악성 메시지가 항상 중재에서 이기고, 다른 메시지 전송을 허용하지 않는 상황을 발생시킬 수 있음 (Denial-of-Service 공격)

CAN 버스 프로토콜과 요구 사항

- 제안된 방법에서 DOS와 같은 공격은 시스템이 버스의 정상적인 트래픽 동작을 학습하여 탐지할 수 있으므로 비정상적인 트래픽 동작을 탐지할 수 있음
  - 비정상적인 트래픽 동작의 예 : 동일한 메시지를 높은 빈도로 전송

- 메시지 데이터 페이로드는 여러 센서에서 생성된 서로 다른 값을 보유하고 특정 ECU에서 관리하고 데이터베이스 컨테이너(DBC) 파일 사양을 기반으로 인코딩됨
  - DBC : 특정 차량 구성에 대한 모든 ECU, CAN 메시지, 신호, 메시지 ID, 메시지 주파수 및 데이터 페이로드의 사양을 보유하는 차량 제조업체 독점 형식의 데이터베이스 파일

제안된
이상 탐지 모델 및 공식

HALLYM
UNIVERSITY

CAN 버스 이상 탐지 방법 제안

■ 기본 아이디어

□ 전송되는 메시지 ID에 반복되는 패턴을 포함하는 일반 CAN 버스 트래픽을 기반으로 모델을 설정

1. 자동차에서 가져온 여러 추적 분석
2. 기록된 추적에서 일부 반복 메시지 ID 패턴 식별
   - 즉, 모든 메시지 ID 뒤에 특정 반복 메시지 ID 하위 집합이 옴
3. 정상적인 트래픽에서 패턴을 식별하는 모델을 개발
   - 패턴이 벗어나면 악성 활동으로 간주될 수 있으며, 이상 행동으로 탐지할 수 있음

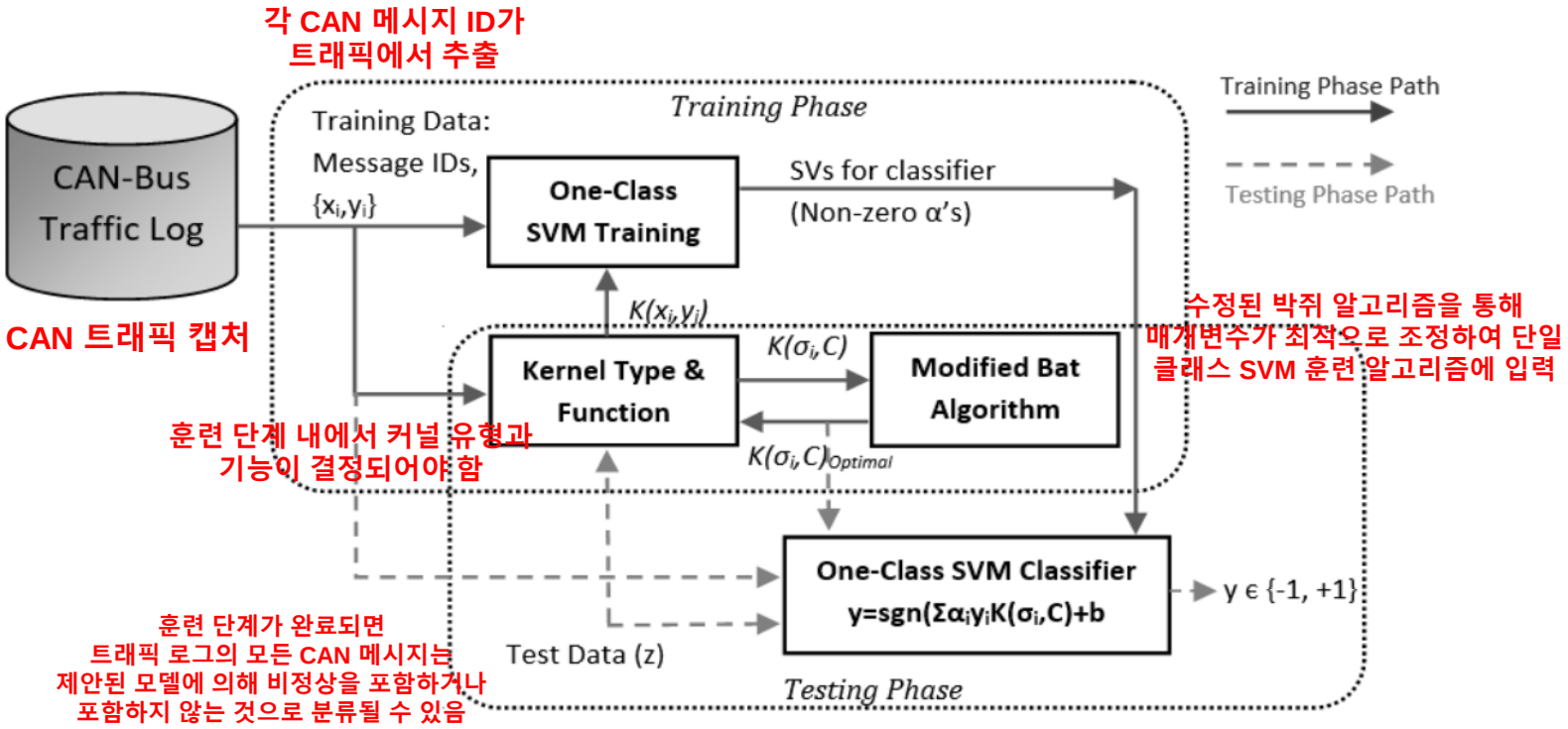
■ 제안된 방법은 훈련 단계와 테스트 및 평가 단계로 구성

□ 교육 단계
- CAN 버스 트래픽의 정상적인 동작은 수정되지 않은 라이선스에서 기록됨
- 정상 작동 상태에서 차량을 공격하지 않고 연속 ID 간에 가능한 모든 전환을 생성

□ 훈련 단계
- 개발된 모델을 참조로 사용하여 공격자가 시작한 CAN 트래픽의 이상 동작을 감지

CAN 버스 이상 탐지 방법 제안

■ 제안된 CAN 버스 이상 탐지 방법의 블록도



CAN 버스 이상 탐지 방법 제안

■ Bat Algorithm

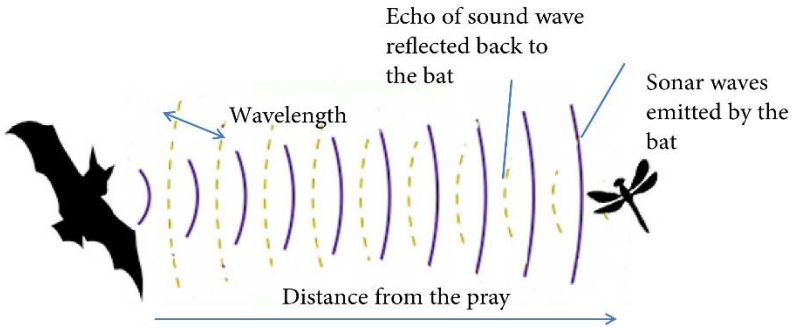
□ 노이즈(소리)와 맥박이라는 주요 사항이 있음

- 노이즈(소리)는 근접성을 감지하며, 먹이에 가까울수록 노이즈 감소
- 맥박주는 박쥐가 사물에 대해 가지고 있거나 소나 방출의 반사에서 얻은 신호로, 먹이에 가까울수록 맥박 상승

- 일반적으로 짧은 펄스를 전송하는데, 먹이를 만나면 펄스 전송 속도와 빈도 수가 증가한다

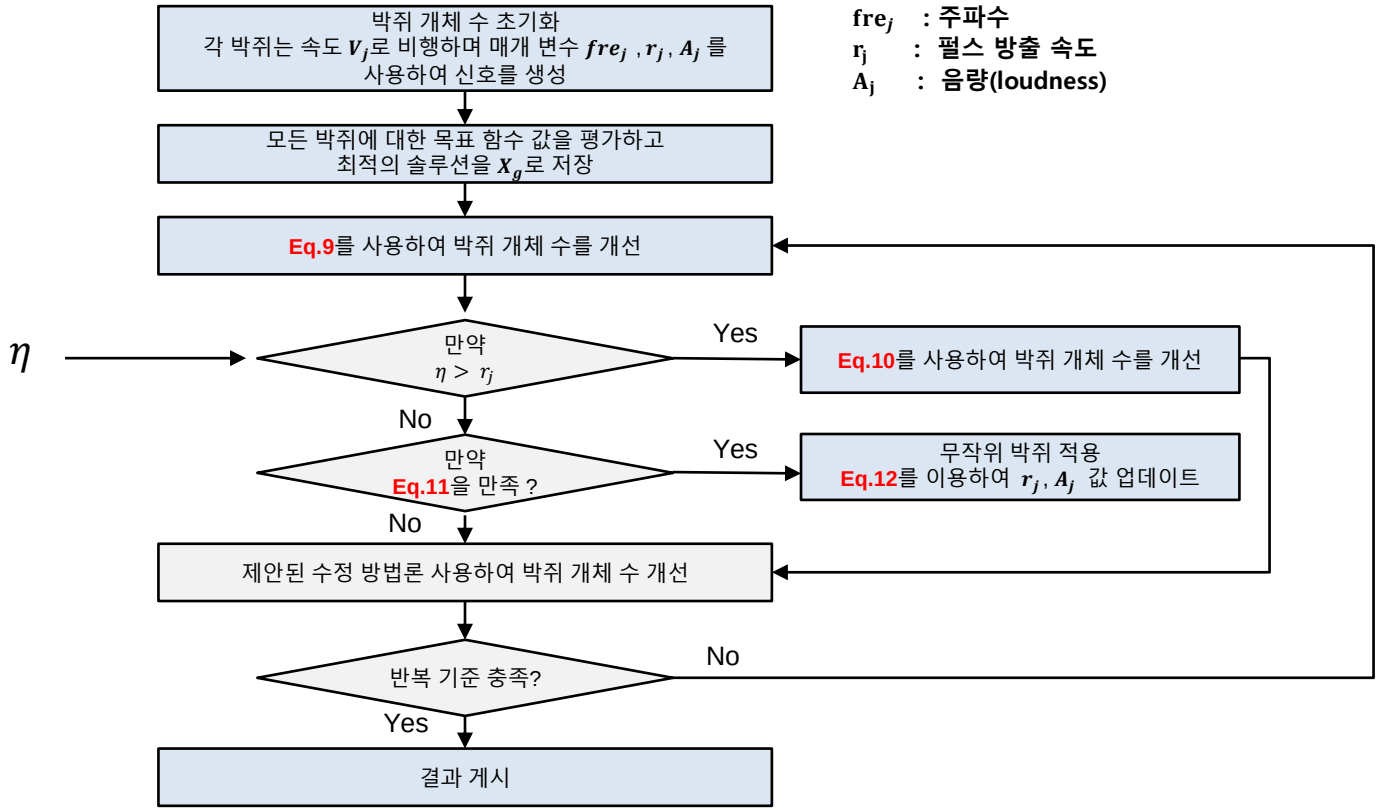
➢ 주파수 증가는 주파수 조정을 의미하며, 반향위치 측정 시간이 단축되고 위치 정확도가 상승한다.

□ Bat Algorithm으로 얻은 x 값이 이전 값보다 작지 않으면 적합도 값을 변경하지 않으므로 기계학습 성능이 향상됨



CAN 버스 이상 탐지 방법 제안

■ 수정된 박쥐 알고리즘 순서도



수정된 박쥐 알고리즘 MBA

■ BA(Bat Algorithm)

1. 각 박쥐에 대해 위치, 속도 및 매개변수 초기화

$$f_j = f_{\min} + \theta_1 (f_{\max} - f_{\min}) \quad \forall j \in \Omega^{bat}$$

범위 (0,1]의 임의의 값

- 위의 수학식과 같이 주파수를 무작위로 생성

2. 각 박쥐의 속도와 위치 업데이트

$$V_{j,k+1} = V_{j,k} + f_j \left(X_g - X_{j,k}\right) \quad \forall j \in \Omega^{bat}$$
$$X_{j,k+1} = X_{j,k} + V_{j,k+1} \quad \forall j \in \Omega^{bat}$$

3. 각 박쥐에 대해 난수를 생성(-1 < ε < 1)
   - $\varepsilon < r_j$ 일 경우, 새로운 방법론 $\boldsymbol{X_j}$ 가 무작위 생성됨

$$X_{j,k+1} = X_{j,k} + \varepsilon A_{j,k}$$

   - 위의 수학식을 이용하여 해당 박쥐의 적합도 값을 계산하고 위치를 업데이트한다.

수정된 박쥐 알고리즘 MBA

■ BA(Bat Algorithm)

4. 각 박쥐에 대해 난수를 생성 $(0 < \eta < 1)$

- $\eta < A_{j,k}$ 이고, $f(X_{j,k}) < f(X_{best})$ 일 때,

$$[\eta < A_j] \& [f(X_j^{new}) < f(X_{best})]$$

두 개의 상수 매개변수

$$A_{j,k+1} = \lambda A_{j,k}$$
$$r_{j,k+1} = r_j^0 \left[1 - \exp(-\gamma k)\right]$$

- 위의 수학식을 통해 $A_{j,k}$와 $r_{j,k}$를 업데이트한다.

5. fitness 값을 기준으로 각 개인을 정렬하고 가장 좋은 위치 저장
6. 조건 충족 시 알고리즘 종료하고, 충족하지 않는다면 2단계로 이동

CAN 버스 이상 탐지 방법 제안

- 저자는 초기 분석에서 프레임 순서, 시간 및 발생 빈도를 포함하여 가능한 모든 특성을 고려함
- 적절한 기능 선택 절차를 통해 적절하고 신뢰할 수 있는 이상 탐지 모델에는 주파수와 프레임 ID만 있으면 충분함을 알 수 있었음

■ **즉, 다른 기능을 고려하면 분류 모델에 개선 사항이 추가되지 않고 이상 탐지 모델의 복잡성만 증가시킴**

■ **본 연구에서는 가장 유익한 특징만을 선택하기 위해 퍼지 기반 특징 선택 방법을 적용함**

단일 클래스 지원 벡터 머신 OCSVM

■ **OCSVM(One Class Support Vector Machine)**

- □ 데이터의 미리 정의된 부분을 포함하는 입력 공간의 최소 부분 집합을 추정하는 분류 알고리즘
- □ 데이터들을 N차원의 좌표축으로 뿌린 후, 원점과의 거리를 기준으로 선(Hyper Plane)을 그려 분류하는 것

- □ OCSVM은 (1) 과 같은 수식에 의해 Hyper Plane이 얻어짐

데이터 포인트의 수

$$\min_{w,\rho} \frac{1}{2}\|w\|^2 + \frac{1}{\upsilon N}\sum_{i=1}^{N}\xi_i - \rho$$

$$s.t.\ w.\Phi(x_i) \geq \rho - \xi_i$$

$$\xi_i \geq 0 \qquad (1)$$

결정 경계를 정의

데이터 포인트에 대한 여유 변수 세트
: 주어진 데이터 포인트가 결정 경계 외부에 위치하도록 허용

이상치의 비율에 대한 상한과 지원 벡터의 비율에 대한 하한을 나타내는 트레이드 오프 매개변수

단일 클래스 지원 벡터 머신 OCSVM

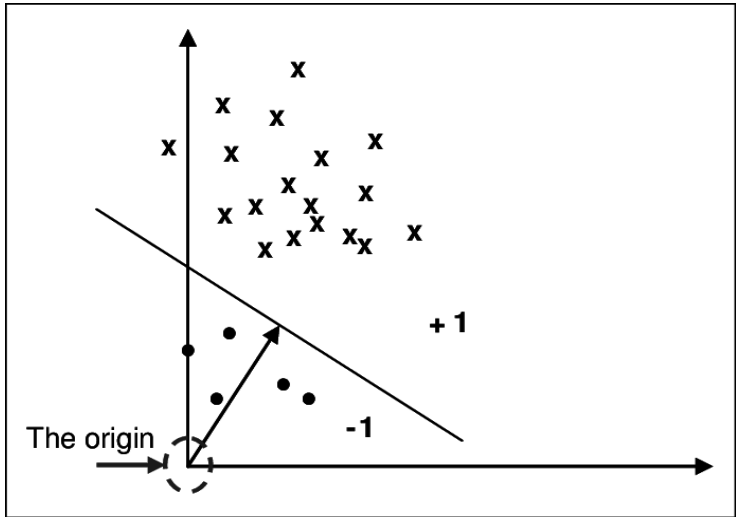
■ **OCSVM(One Class Support Vector Machine)**

□ 결정 경계 함수 정의

$$f(x) = w \cdot \Phi(x) - \rho \qquad (2)$$

목표

→ 주어진 데이터 포인터가 정상적인 CAN 트래픽 내에 속할 때 +1 반환 , 그렇지 않으면 -1 반환



https://limitsinx.tistory.com/147

단일 클래스 지원 벡터 머신 OCSVM

■ **OCSVM(One Class Support Vector Machine)**

□ 결정 경계 함수 정의

$$L(w,\xi,\rho,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \frac{1}{\upsilon N}\sum_{i=1}^{N}\xi_i - \rho - \sum_{i=1}^{N}\alpha_i(w.\Phi(x_i) - \rho + \xi_i) - \sum_{i=1}^{N}\beta_i\xi_i \quad (3)$$

$$w = \sum_{i=1}^{N}\alpha_i\Phi(x_i) \quad (4)$$

$$\alpha_i = \frac{1}{\upsilon N} - \beta_i \qquad \sum_{i=1}^{N}\alpha_i = 1 \quad (5)$$

$$\min_{\alpha}\alpha^T H\alpha$$

$$s.t.\ 0 \le \alpha_i \le \frac{1}{\upsilon N}$$

$$\sum_{i=1}^{N}\alpha_i = 1 \quad (6)$$

$$H_{ii} = K(x_i,x_i) = \Phi(x_i).\Phi(x_i) \quad (7)$$

$$\rho = \frac{1}{n_s}\sum_{i=1}^{n_s}\sum_{j=1}^{N}\alpha_j K(x_i,x_j)\alpha \quad (8)$$

수정된 박쥐 알고리즘 MBA

■ BA(Bat Algorithm)

□ 박쥐가 먹이를 탐지하는 데 사용하는 반향 위치 확인 프로세스에서 영감을 얻은 메타 발견적 최적화 알고리즘

- 각 박쥐는 각각의 방출 속도, 주파수 및 크기를 사용하여 큰 신호를 대기 중으로 보냄
- 반향 정위 소리를 들음으로써 박쥐는 먹이를 기준으로 속도와 위치를 업데이트할 수 있음

- 위 알고리즘은 비선형, 비볼록 및 다중 모드 최적화 문제에 직면하여 뛰어난 성능을 보여줌
- 하지만, 이룹 상황에서는 로컬 최적에 갇히거나 조기 수렴에 직면할 수 있음

MBA 제안

수정된 박쥐 알고리즘 MBA

■ MBA(Modified Bat Algorithm)

1. 박쥐 개체군 다양성을 증가시키기 위한 돌연변이 및 교차 연산자에 기반한 새로운 수정 방법인 두 가지 접근 방식을 기반으로 구성
   - 가능한 조기 수렴을 피할 뿐만 아니라 BA의 전역 검색 능력을 절대적으로 향상시킴

$$X_{mut} = X_{z1,k} + \theta_1 \times (X_{z2,k} - X_{z3,k})$$
$$X_{mut} = [x_{mut,1}, x_{mut,2}, \ldots, x_{mut,N}] \qquad (13)$$

Z : 박쥐
N : 제어 변수의 수

   - 각 반복에서 각 박쥐에 대해 3개의 다른 박쥐가 모집단에서 선택되어 $z_1 \neq z_2 \neq z_3 \neq j$.가 됨
   - 위와 같은 식을 사용하여 새로운 돌연변이 박쥐가 생성됨
   - 두 개의 새로운 테스트 박쥐가 아래와 같이 생성됨

이 중 최고의 솔루션이 Xj로 대체

$$x_{j1,v}^{test1} = \begin{cases} x_{mut,v} & \text{if } \theta_1 \leq \theta_2 \\ x_{g,v} & otherwise \end{cases} \qquad (14)$$
$$X_{j2}^{test2} = \theta_3 \times X_{mut} + \theta_4 \times (X_g - X_{mut}) \qquad (15)$$

Xmut와 V : 박쥐 벡터 Xmut에서 V번째의 요소
$\theta 1, L, \theta 4$: (0,1] 범위의 임의의 값

2. BA의 수렴율을 높이기 위한 새로운 수학 기반 수정 방법
   - 각 반복에서 최고의 박쥐는 다른 박쥐의 위치를 개선하려고 시도
   - 먼저 박쥐 개체수의 평균을 평가한 후 최고의 박쥐로부터의 거리에 따라 각 박쥐의 위치를 개선하려고 시도함으로써 가능함

$$X^{test3} = X_j + \theta_5(X_g - \phi_F A_D) \qquad (16)$$

[0,1] 의 임의 값

이동 가속도를 나타내는 1 또는 2와 같은 임의의 정수

결과 및 논의

HALLYM
UNIVERSITY

수정된 박쥐 알고리즘 MBA

□ 인가된 수정되지 않은 차량과 두 개의 다른 공용 CAN 버스 트래픽 데이터 세트에서 수집한 실제 데이터를 기반으로 시뮬레이션 결과를 제공하여 제안된 모델의 정확도를 검사함
- CAN 트래픽이 차량에 존재하는 OBD-II 포트를 통해 VN1630A 장치에 의해 기록된 전기 자동차

□ 원본 형식 데이터셋
- 타임스탬프, 메시지 ID 및 데이터 필드가 있는 쉼표로 구분된 값(csv)을 포함하는 텍스트 파일 모음

□ 데이터 분할(무작위)
- Train : 70%
- Validation : 10%
- Test : 20%

□ 우수성 입증을 위해 세 알고리즘을 비교
- MBA-OCSVM
- 기존 OCSVM
- Isolation Forest

수정된 박쥐 알고리즘 MBA

□ 제안된 이상 탐지 방법의 효율성 평가
- 공격자가 보내지 않은 것으로 예상되는 메시지를 보내려고 하는 조건을 시뮬레이션

① CAN 트래픽에서 메시지ID의 빈도를 증가시켜 공격 시나리오 복제
② 차량이 정상작동 > ID의 주기 시간은 100ms
③ 의도적으로 메시지 빈도를 2배로 증가시키고, 100ms마다 다른 메시지 전송
④ 모든 노드가 단일 버스를 공유
- 버스 점유가 증가하면, 다른 메시지의 대기 시간이 발생
- 운전자 명령에 응답하지 않는다면, 가용성에 대한 위협 발생할 수 있음

□ 정상 주행 중 캡처된 CAN 트래픽에서 각 메시지 발생에는 트래픽에서 고유된 주파수 패턴이 있음을 확인

수정된 박쥐 알고리즘 MBA

- DoS 공격과 같은 동일한 공격 시뮬레이션 시나리오가 본 연구에서 사용되는 다른 데이터 세트에 대해 개발되었음

CAN Identifier	Frequency
6FF	101.010101
308	85.74311927
340	50
2A0	48.7804878
670	99.00990099
3F0	100.1666667
D21	38.7804878
210	51.02040816
238	108.6956522
410	93.45794393
200	61.02040816
A7F	49.01960784
B61	10
212	68.54368932
240	78.01010101
4EB	113.6363636
2C1	110.3595506
312	50
5AE	80.01960784

표1 CAN 식별자(ID)와 빈도

- 각 CAN 식별자는 고유한 발생 빈도를 가지고 있다
- 제안된 방법의 train 단계를 시작하기 전 CAN 식별자가 동일한 빈도를 유지하면서 10진수 표시기로 변환되도록 데이터를 분석하여 올바른 구조 형식으로 저장하기 위해 데이터 전처리가 필요함
- 데이터셋의 최대값과 최소값 사이의 매우 큰 범위를 피하기 위해 속성 재조정을 통해 속성을 정규화

수정된 박쥐 알고리즘 MBA

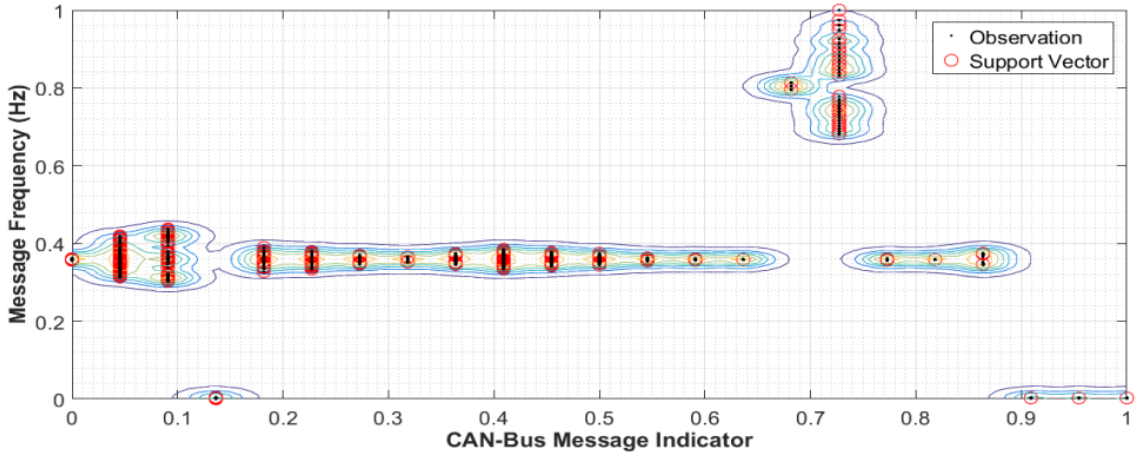


그림5 표 1에 표시된 선택된 CAN 식별자에 대해 관찰된 주파수, 지원 벡터 및 결정 경계를 나타냄

□ 데이터 전처리 단계를 완료하면 각 CAN 식별자와 해당 주파수를 제안된 방법에 입력 매개변수를 가져온 다음 MOCSVM 의 train 단계가 시작된다

수정된 BAT 알고리즘의 성능 평가

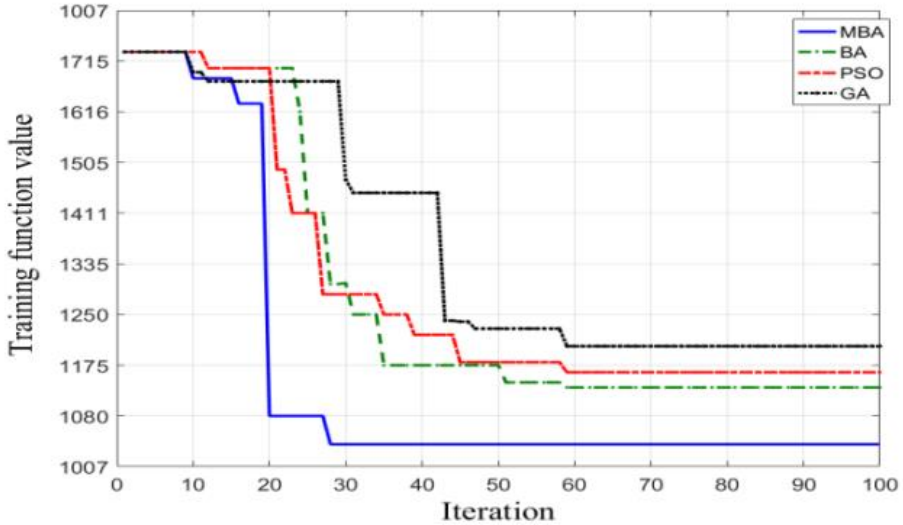


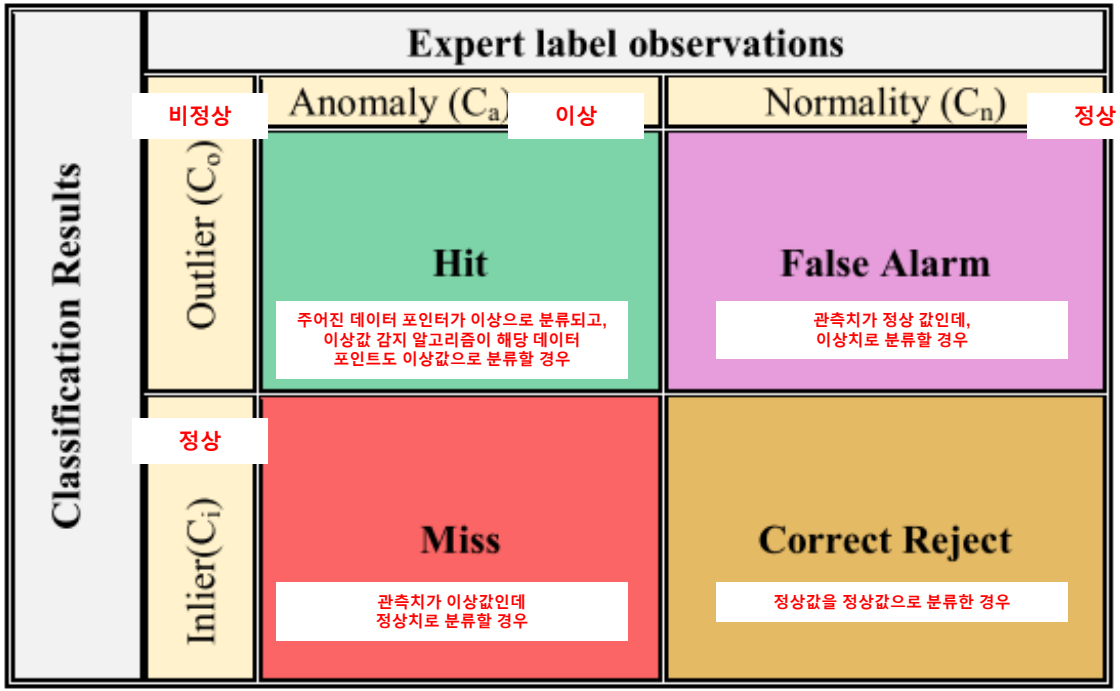
그림 6 훈련과정에서 MBA, BA, PSO, 및 GA의 수렴 특성

- PSO(Particle Swarm Optimization)
- GA(Genetic Algorithm)
- Original BA
- MBA(Modified Bat Algorithm)

- 제안된 MBA는 먼저 수렴했을 뿐만 아니라 다른 알고리즘에서는 찾을 수 없는 최적의 솔루션에 도달할 수 있었음을 보여줌
- MBA의 높은 검색 능력과 수렴 특성을 보여줌

제안된 MBA-OCSVM의 성능 평가

■ 4가지 혼동행렬을 통하여 성능 평가



제안된 MBA-OCSVM의 성능 평가

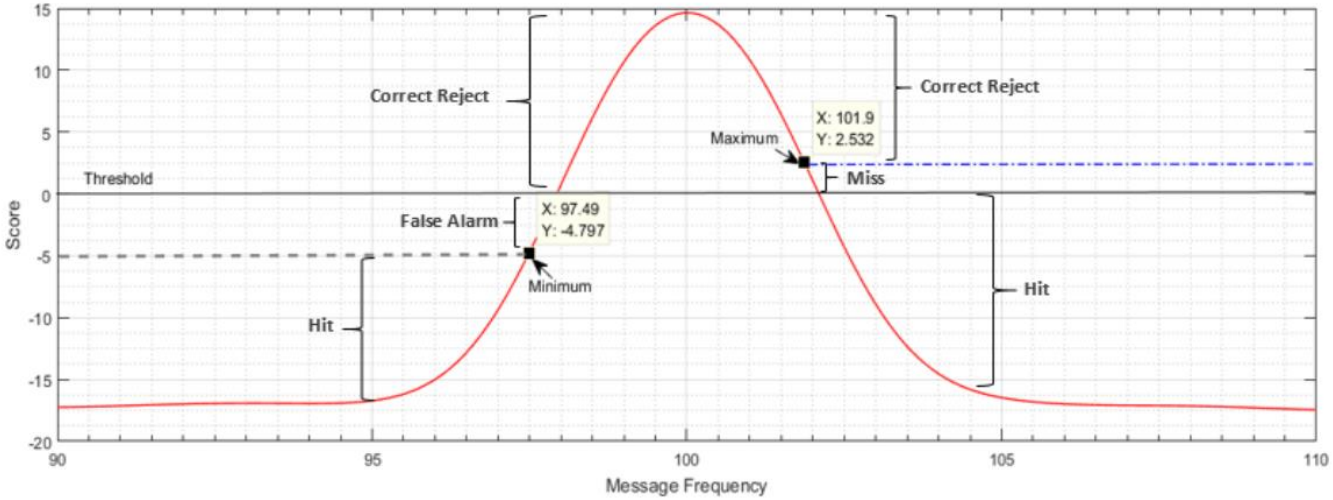
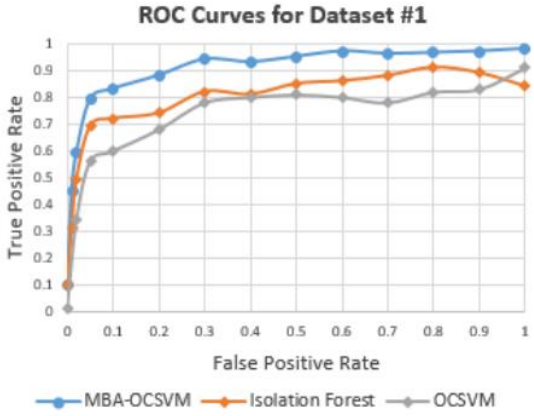


그림 8 주어진 CAN 버스 메시지에 대한 HIT, MISS, FALSE ALARM, CORRECT REJECT

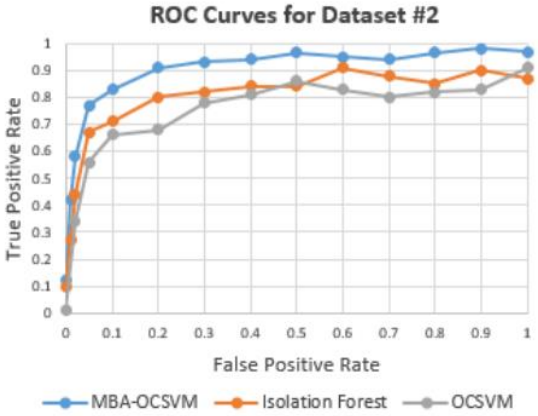
- 공격자가 주어진 CAN 메시지의 발생 빈도를 조작하려고 하면 제안된 알고리즘은 해당 동작을 비정상적인 상황으로 감지하여 폐기할 수 있다
- 일반적으로 공격자는 중재 시나리오에서 승리하여 버스에서 데이터를 게시할 수 있도록 메시지 빈도를 두 배로 늘리려고 한다
- 이 동작은 손상된 메시지 주파수가 일반 CAN 트래픽에서 주어진 메시지에 대한 최대 유효 주파수보다 높은 적중 영역에 해당됨

제안된 MBA-OCSVM의 성능 평가

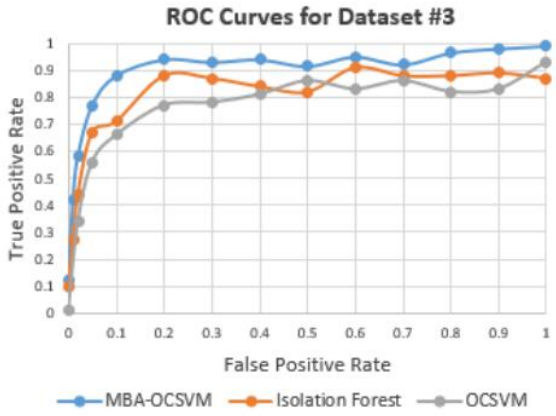
■ 제안된 MOCSVM, Isolation Forest 및 Original Classic SVM에 대한 ROC 성능 곡선



DATASET 1
정상 작동 상태에서 수정되지 않은
면허 차량의 CAN BUS 트래픽



DATASET 2
Dodge RAM Pickup track(Tulsa 대학
충돌 재건 연구 컨소시엄)의
CAN BUS 트래픽 데이터 세트



DATASET 3
사용 가능한
온라인 CAN BUS 트래픽 데이터 세트

□ ROC 곡선

- 높은 적중률과 낮은 오탐율, 낮은 오경보율을 나타내므로 가파른 경사로 가능한 한 왼쪽 상단 모서리에 가까워야 함
- 제안된 MBA는 높은 적중률과 낮은 오경보율을 보임

제안된 MBA-OCSVM의 성능 평가

■ OCSVM, Isolation Forest, MBA-OCSVM 에 대한 혼동행렬

Dataset #1	Hit Rate	Miss Rate	False Alarm Rate	Correct Reject Rate
One-Class SVM	83.12%	16.88%	18.07%	81.93%
Isolation Forest	89.01%	10.99%	13.57%	86.43%
Proposed MBA-OCSVM	95.5%	4.5%	8.54%	91.45%
Dataset #2	Hit Rate	Miss Rate	False Alarm Rate	Correct Reject Rate
One-Class SVM	85.62%	14.38%	16.04%	83.96%
Isolation Forest	89.99%	10.01%	13.34%	86.66%
Proposed MBA-OCSVM	96.99%	3.01%	7.11%	92.89%
Dataset #3	Hit Rate	Miss Rate	False Alarm Rate	Correct Reject Rate
One-Class SVM	86.01%	13.99%	15.8%	84.2%
Isolation Forest	88.85%	11.15%	13.34%	86.66%
Proposed MBA-OCSVM	97.01%	2.99%	6.45%	93.55%

결론

HALLYM
UNIVERSITY

결론

■ CAN 버스 트래픽에 대한 효과적인 이상 탐지 모델 제안

□ MOCSVM은 MBA라는 수정된 박쥐 알고리즘을 기반으로 구성
- 알고리즘이 로컬 최적에 갇히는 것과 조기 수렴을 방지하는 데 도움이 됨

□ MOCSVM 방법은 CAN 트래픽에서 악성 사이버 공격 행위를 탐지하는 데 사용
- 주어진 일반 트래픽에서 전송되는 메시지 ID에 반복되는 링 패턴을 포함하는 일반 CAN 버스 트래픽을 기반으로 모델 설정

□ MBA-OCSVM 알고리즘
- 정상적인 트래픽과의 편차를 이상값으로 탐지할 수 있음

□ 입증
- 기존의 OCSVM, Isolation Forest 및 MBA-OCSVM 세 가지 방법 비교
- 제안된 방법이 다른 이상 탐지 방법에 비해 가장 높은 적중률과 가장 낮은 실패율을 보임

□ 결론
- 제안된 모델은 공격자에 의해 피해 입지 않도록 매우 안전하고 정확한 모델을 제공
- 높은 검색 능력과 수렴성을 보여줌

참고문헌

■ **Bat algorithm**

- http://richardlecture.blogspot.com/2018/09/bat-algorithm-2010-xin-she-yang-1.html
- https://www.mdpi.com/2227-7390/7/2/135/htm

■ **ROC**

- https://www.ibm.com/docs/ko/spss-statistics/27.0.0?topic=features-roc-analysis
- https://huidea.tistory.com/178

■ **IVN프로토콜**

- https://www.itbiznews.com

■ **OCSVM**

- https://limitsinx.tistory.com/147

Q&A

HALLYM
UNIVERSITY

Thank You

HALLYM
UNIVERSITY