ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
j.inst.Korean.electr.electron.eng.Vol.28,No.3,412~418,September 2024
논문번호 24-03-24
https://doi.org/10.7471/ikeee.2024.28.3.412
170

# 랜덤 샘플 합의를 사용한 초경량 차량용 침입 탐지 시스템

# Ultra-Light-Weight Automotive Intrusion Detection System Using Random Sample Consensus

김 종 권*, 임 형 철*, 이 주 석*, 이 성 수*★

Jonggwon Kim*, Hyungchul Im*, Joosock Lee*, and Seongsoo Lee*★

## Abstract

This paper proposes an effective method for detecting hacking attacks in automotive CAN bus using the RANSAC (Random Sample Consensus) algorithm. Conventional deep learning-based detection techniques are difficult to be applied to resource-constrained environments such as vehicles. In this paper, the attack detection performance in vehicular CAN communication has been improved by utilizing the lightweight nature and efficiency of the RANSAC algorithm. The RANSAC algorithm can perform effective detection with minimal computational resources, providing a practical hacking detection solution for vehicles.

## 요 약

본 논문은 RANSAC(Random Sample Consensus) 알고리즘을 활용하여 차량용 CAN 통신에서 발생하는 해킹 공격을 효과적으로 탐지하는 방법을 제안한다. 기존에 제안된 딥러닝 기반 탐지 기법은 차량과 같이 리소스가 제한된 환경에는 적용하기 어렵다는 한계가 있다. 본 논문에서는 RANSAC 알고리즘의 경량성과 효율성을 활용하여 차량용 CAN 통신에서의 공격 탐지 성능을 향상시켰다. RANSAC 알고리즘은 적은 연산 자원으로도 효과적인 탐지를 수행할 수 있어서 차량에 탑재 가능한 실용적인 해킹 탐지 솔루션을 제공할 수 있다.

*Key words : Controller Area Network, Intrusion Detection System, Random Sample Consensus, Support Vector Machine, K-Nearest Neighbor*

## Ⅰ. 서론

자동차 내부의 전자장치 제어에 사용되는 차량 내 네트워크는 차량 기능의 다양화와 고도화에 따라 점점 더 복잡해지고 있다. 이러한 발전은 운전의 편리성과 안전성을 크게 향상시키고, 자율 주행과 V2X(Vehicle-to-Everything) 통신과 같은 첨단 기술의 구현을 가능하게 한다. 그러나 네트워크 복잡성의 증가로 인해 외부 공격에

* School of Electronic Engineering and Department of Intelligent Semiconductor, Soongsil University (Student, Student, Professor, Professor)
★ Corresponding author
E-mail : sslee@ssu.ac.kr, Tel : +82-2-820-0692
※ Acknowledgment
This work was supported by the R&D Program of the Ministry of Trade, Industry, and Energy (MOTIE) and Korea Evaluation Institute of Industrial Technology (KEIT). (RS-2022-00154973, RS-2023-00232192, RS-2024-00403397). It was also supported by MOTIE and Korea Institute for Advancement of Technology (KIAT) (P0012451). The authors wish to thank IC Design Education Center (IDEC) for CAD support.
Manuscript received Sep. 23, 2024; revised Sep. 24, 2024; accepted Sep. 25, 2024.

대한 노출 위험도 높아지고 있다[1]. 특히 차량 내부 통신의 핵심 역할을 하는 CAN(Controller Area Network)은 암호화와 인증 메커니즘의 부족으로 해커들의 공격에 매우 취약하다. 이 문제를 해결하기 위해 CanNet, DCNN (Deep Convolutional Neural Network), GIDS (Generalized Intrusion Detection System)와 같은 다양한 딥러닝 기반 침입 탐지 시스템들이 제안되었다 [2]-[4]. 하지만 이러한 시스템들은 높은 연산 능력과 많은 메모리를 필요로 하며, 이는 리소스가 제한된 차량 환경에서 실시간 데이터 처리를 수행하기에 적합하지 않다. 차량의 제어 시스템은 낮은 지연 시간과 실시간 처리가 필수적이므로, 고도의 연산 자원을 요구하는 시스템은 실용적이지 않다.

본 논문에서는 RANSAC(Random Sample Consensus) 알고리즘을 적용한 경량화된 침입 탐지 시스템을 제안한다. RANSAC 알고리즘은 적은 연산 자원으로도 효과적인 성능을 발휘할 수 있어, 차량 환경에 적합한 침입 탐지 시스템 구현에 적합하다. 본 논문에서는 RANSAC 기반 시스템의 성능을 평가하고, 기존의 딥러닝 기반 방법들과 비교하여 실제 차량 네트워크 환경에서의 적용 가능성을 검토하였다.

## II. RANSAC 기반 침입 탐지 시스템

본 논문에서 제안하는 침입 탐지 시스템은 그림 1에 나타낸 과정을 거치며, 각 단계별로 다음과 같이 수행한다.

### 1. 데이터 전처리

RANSAC(Random Sample Consensus) 알고리즘은 아웃라이어가 포함된 데이터에서 모델을 추정하기 위한 알고리즘이다. RANSAC 알고리즘은 먼저 주어진 데이터에서 일부 샘플을 무작위로 선택하여 모델을 추정한다. 이후, 추정된 모델을 바탕으로 전체 데이터에서 모델에 잘 맞는 데이터 포인트(인라이어)를 선택하고, 나머지 데이터 포인트는 아웃라이어로 간주한다. 이 과정을 여러 번 반복하여 각 반복에서 무작위로 샘플을 선택하고 새로운 모델을 추정한다. 각 반복에서 가장 많은 인라이어를 포함하는 모델이 선택되며, 최종적으로 가장 적합한 모델을 선정한다. RANSAC은 사전에 정의된 반복 횟수나 충분히 많은 인라이어가 발견될 때까지 반복한다. 인라이어는 모델과 잘 일치하는 데이터로, 이들을 통해 정확한 모델을 형성하며, 아웃라이어는 모델과 부합하지 않아 제거된다. RANSAC 알고리즘을 적용하기 위해서는 데이터의 분포를 파악하고, 분포에 적합한 회귀 모델이

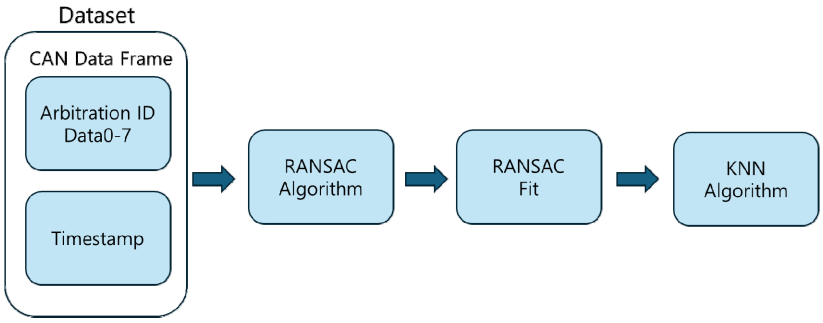


Fig. 1. Flowchart of the poposed intrusion detection system.
그림 1. 제안하는 침입 탐지 시스템의 흐름도

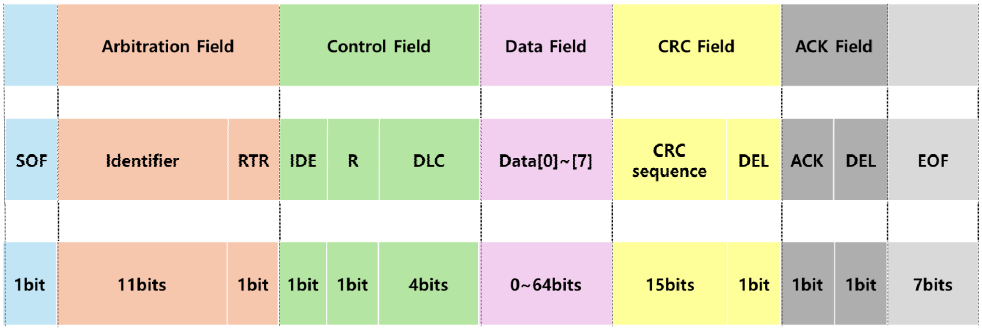


Fig. 2. CAN data frame.
그림 2. CAN 데이터 프레임

필요하다.

본 논문에서는 데이터의 비선형적 관계를 처리하기 위해 SVR(Support Vector Regression) 모델을 사용하고 RANSAC 알고리즘을 적용하여 CAN(Controller Area Network) 통신에서 발생할 수 있는 정상 메시지와 비정상 메시지를 구별하는 방법을 제안한다. 먼저, 공격이 포함되어 있지 않은 정상 데이터 세트와, RPM, Gear, Fuzzy 공격을 포함하고 있는 공격 데이터 세트[5]를 각각 준비한다. 그 후, 정상 데이터에서 차량 내 중요한 장치와 관련된 CAN ID를 추출하고, 이들의 데이터 흐름을 분석하여 RANSAC 알고리즘을 적용한다. 모든 데이터 세트는 Arbitration ID와 Data Field의 데이터 8개를 포함하고 있고, 메시지가 차량 네트워크 상에서 발생한 시간을 의미하는 Timestamp를 포함한다. Arbitration ID와 Data Field를 보이기 위해 CAN 데이터 프레임을 그림 2에 나타내었다. 이렇게 추출된 데이터들에 식(1)과 식(2)를 적용하여 전처리를 시행한다. 그 후 RANSAC 알고리즘을 적용한다. 8개의 데이터를 갖는 각 Arbitration ID를 하나의 데이터 포인트로 나타내기 위해 식 (1)을 사용하였다. 그 후, 식 (2)을 사용하여 y값의 범위를 통일하는 정규화를 실시하였다.

$$y_i = \frac{\sum_{j=0}^{7} Data_j[i]}{8} \tag{1}$$

$$y_{scaled\ i} = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \tag{2}$$

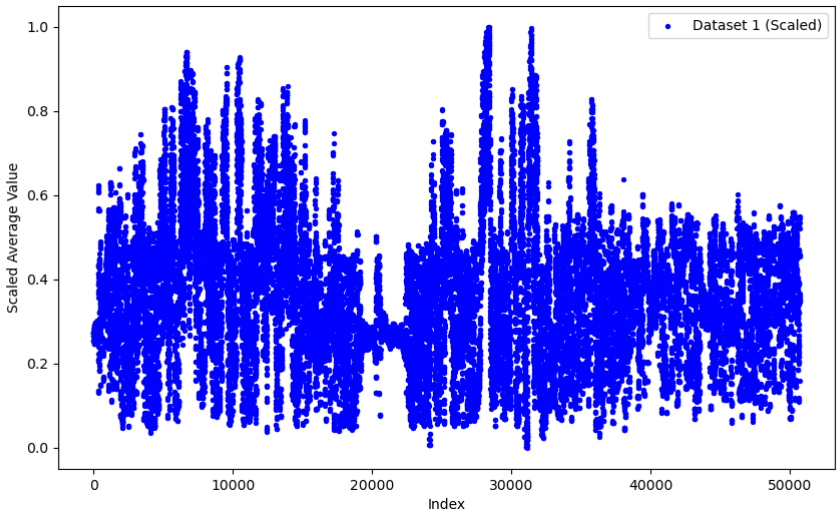


Fig. 3. Data distribution of the normal dataset.
그림 3. 정상 데이터 세트의 데이터 분포

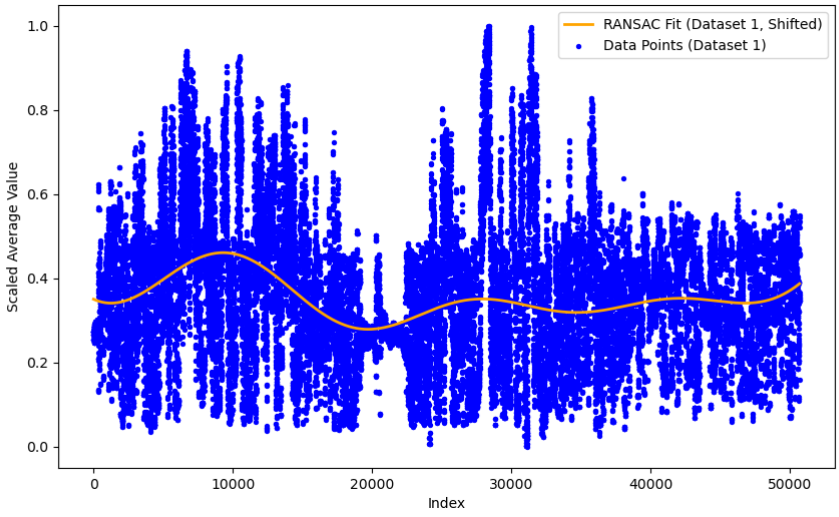


Fig. 4. RANSAC fit estimated from the normal dataset.
그림 4. 정상 데이터 세트로 추정된 RANSAC Fit

각 Arbitration ID 마다 8개의 데이터를 포함하고 있기 때문에 이 데이터들의 평균을 y값으로 사용한다. 식 (2)는 식 (1)에서 구한 y값에 적용하는 min-max 정규화다. 이를 통해 넓은 범위의 y값을 0과 1 사이로 축소한다.

## 2. SVR(Support Vector Regressor) 회귀 모델

RANSAC 알고리즘을 적용하기 위해 준비된 데이터 세트에 맞는 모델을 설정해야 한다. 데이터 세트는 비선형적인 특징을 갖기 때문에 그에 맞는 모델이 필요하다. 그림 3은 정상 데이터 세트에서 특정 ID(0×316)의 데이터 분포를 나타낸다. 본 논문에서는 SVR 회귀 모델을 적용하였다. SVR(Support Vector Regression) 회귀 모델은 Support Vector Machine 알고리즘을 기반으로 한 회귀 분석 기법으로, 회귀 문제를 해결하는 데 효과적인 방법이다. SVR은 데이터의 분포와 관계없이 고차원 공간에서 데이터를 선형으로 구분할 수 있도록 해주며, 예측 모델을 만드는 데 사용된다. SVR의 핵심 아이디어는 마진으로, 이는 데이터를 잘 설명하는 회귀선을 찾는 것이다. SVR은 마진 내에 최대한 많은 데이터 포인트가 포함되도록 회귀선을 설정하고, 이 마진을 벗어나는 데이터 포인트에 대해서만 패널티를 부과한다. 이 과정에서 중요한 요소는 '입실론($\varepsilon$) 튜브'라는 개념인데, 이는 마진 내의 오차를 허용하는 영역을 의미한다. SVR은 이 입실론 튜브 내의 오차는 무시하고, 튜브 밖에 있는 오차에 대해서만 패널티를 부여함으로써 전체적인 모델의 복잡도를 줄인다.

SVR은 SVM 알고리즘에 커널 함수(kernel function)를 사용하여 비선형적인 데이터를 다룰 수 있는 능력을 갖추고 있다. 커널 함수를 통해 입력 데이터를 고차원 공간으로 변환하여 비선형적인 관계를 선형적으로 처리할 수 있게 된다. 이를 통해 SVR은 선형 모델로는 해결할 수 없는 복잡한 패턴의 데이터도 효과적으로 분석할 수 있다. 식 (3)는 가우시안 커널을 사용하기 전의 선형 SVM 식으로, 데이터 포인트 x의 예측을 나타낸다. 식 (4)는 가우시안 커널을 사용한 후의 SVR 함수이다. 가우시안 커널을 사용하여 비선형 데이터를 효과적으로 분류할 수 있도록 하였다. 정상 데이터 세트는 비선형 데이터이기 때문에 RANSAC 알고리즘을 적용하기 위해 SVR 모델을 사용한다. SVR 모델을 사용하여 정상 데이터들의 분포에 대해, RANSAC 알고리즘을 적용할 수 있도록 RANSAC Fit을 만든다.

$$f_{SVM}(x) = sign(w \circ x + b) \quad (3)$$

$$f_{SVR}(x) = sign(\sum_{i=1}^{n} \alpha_i y_i \exp(-\frac{||x - x_i||^2}{2\sigma^2}) + b) \quad (4)$$

## 3. RANSAC 알고리즘 적용

SVR 회귀 모델로 RANSAC 알고리즘을 적용하여, RANSAC Fit을 만든다. 정상 데이터 세트에서 특정 CAN ID(0×316)의 RANSAC Fit을 그림 4에 나타내었다. 추정된 RANSAC Fit을 공격이 포함된 데이터 세트에 같은 길이만큼 반복해서 적용한다. 그 후, 잔류 임계값(Residual Threshold)보다 낮은 데이터 포인터는 인라이어로, 높은 데이터 포인트는 아웃라이어로 지정한다. 최소 데이터 포인트 수를 0.5로 설정하였고, 잔류 임계값을 0.35로 설정하였다. 즉, 모델 예측값과 실제값 사이의 최대 거리를 0.35로 지정하여, 이 임계값보다 낮은 데이터 포인트는 인라이어로 분류되고, 임계값보다 높은 데이터 포인트는 아웃라이어로 간주되도록 설정하였다.

RANSAC Fit을 사용하여 구분된 인라이어와 아웃라이어를 그림 5에 나타내었다. 그림 5에서 정상 데이터의 일부가 아웃라이어로 오분류 되어있고, 이를 해결하기 위해 KNN 알고리즘을 적용하는 후처리 과정을 실시한다.

## 4. KNN(K-Nearest Neighbor) 후처리

K-Nearest Neighbors(KNN) 알고리즘은 새로운 데이터 포인트를 분류하거나 회귀 분석을 수행할 때, 해당 포인트와 가장 가까운 K개의 이웃 데이터를 기반으로 결정을 내리는 비지도 학습 방법이다. 이 알고리즘은 거리 측정법을 사용하여 데이터 포인트 간의 유사성을 평가한다. 본 논문에서는 KNN 알고리즘을 후처리 단계에서 활용하여 초기 예측 결과의 정확도를 개선하고자 한다.

RANSAC 알고리즘이 생성한 예측 결과에 대해 KNN을 적용하여, 주변 데이터 포인트의 정보를 바탕으로 예측값을 보정한다. KNN 알고리즘에서는 유클리드 거리를 사용하여 주변 이웃을 찾는다. 이 과정을 통해 정상 데이터 세트 중 아웃라이어로 판단된 데이터들을 인라이어로 보정할 수 있다. 두 데이터 포인트의 좌표 p, q 사이의 맨해튼 거리를 측정하여, 이 기준으로 가장 가까운 K개의 이웃을 찾아서, 아웃라이어의 여부를 판단한다. 이 알고리즘의 동작 수식을 식 (5)에 나타내었다. 본 논문에서는 알고리즘에 사용되는 최소 이웃의 수를 2로 지정하고, 인라이어와 아웃라이어 사이에 최소한으로 발생해야 하는 변경 수인 min_changes_threshold를 1로 지정하였다. 즉, 더 이상 변경 사항이 없을 때까지 반복해서 KNN 알고리즘을 적용한다.

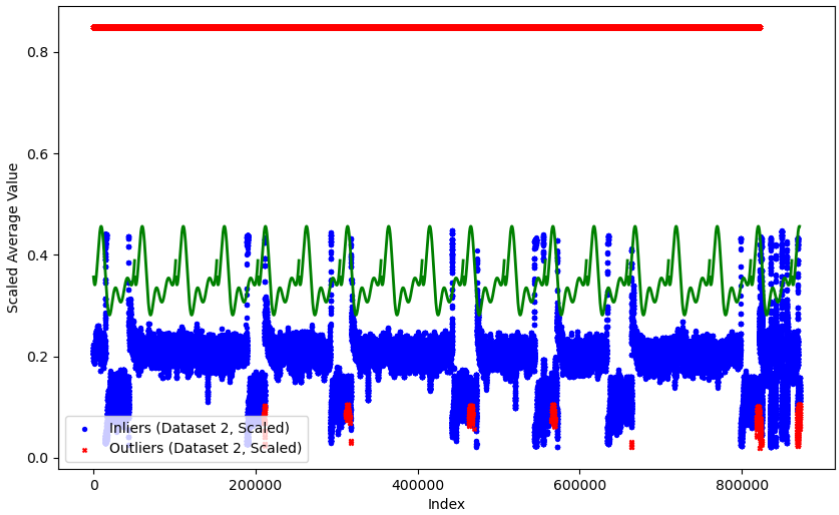


Fig. 5. RANSAC fit applied to the attack data.
그림 5. 공격 데이터에 적용시킨 RANSAC Fit

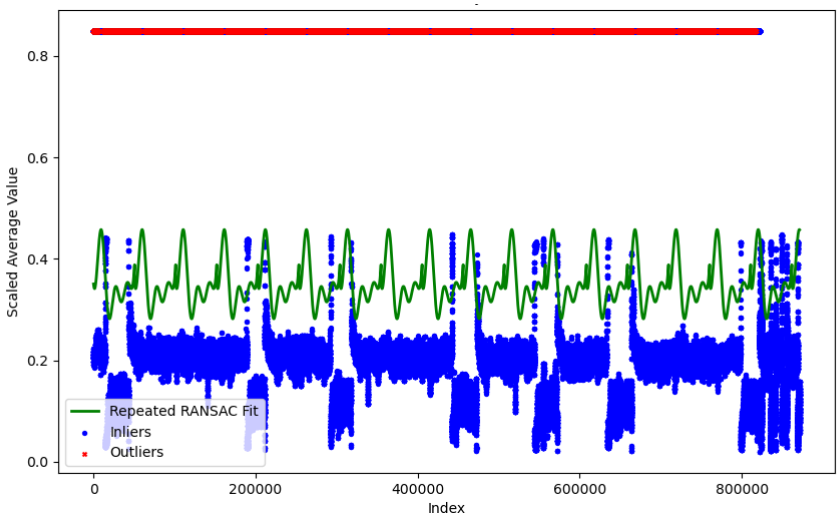


Fig. 6. Post-processing applied using the KNN algorithm.
그림 6.KNN 알고리즘을 적용한 후처리

$$d(p,q) = \sum_{i=1}^{n} | p_i - q_i | \quad (5)$$

그림 6은 KNN 알고리즘을 적용하여 후처리 과정을 거친 후의 인라이어와 아웃라이어를 나타낸다. 그림 5와 비교하였을 때, 정상 데이터에 오분류된 아웃라이어가 인라이어로 바뀐 것을 확인할 수 있다.

## Ⅲ. 실험 결과

본 논문에서 제안한 RANSAC 기반 침입 탐지 시스템은 정상 데이터를 기준으로 특정 노드 ID를 필터링하여, RANSAC 알고리즘을 적용해 RANSAC Fit을 만든다. RANSAC Fit을 공격 데이터 세트에 적용하여 인라이어와 아웃라이어를 구분한다. 그 후 KNN 알고리즘을 적용하여 정상 데이터 세트의 오분류된 아웃라이어를 인라이어로 수정하는 과정을 거친다. 본 논문에서는 침입 탐지 시스템의 성능을 평가하기 위해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall)을 사용하였다. 이들 지표는 모델의 탐지 성능과 오탐지 여부를 정량적으로 평가하는 데 유용하다.

정확도(Accuracy)는 전체 데이터에서 모델이 정확하게 분류한 비율을 나타내며, 식 (6)과 같이 정의된다. 여기서, TP(True Positive)는 실제 침입을 올바르게 탐지한 경우를, TN(True Negative)은 정상적인 상황을 정확하게 탐지한 경우를 의미한다. FP(False Positive)는

정상 상황을 침입으로 잘못 탐지한 경우를, FN(False Negative)은 침입을 탐지하지 못한 경우를 나타낸다. 정밀도(Precision)는 모델이 탐지한 침입 중 실제 침입의 비율을 평가하는 지표로, 식 (7)과 같이 계산된다. 정밀도는 FP를 최소화하는 데 중점을 둔 지표로, 탐지된 침입이 얼마나 신뢰할 수 있는지를 평가하는 데 중요하다. 높은 정밀도는 오탐지의 발생률을 줄이는 데 기여한다. 재현율(Recall)은 실제 침입 중에서 모델이 올바르게 탐지한 비율을 나타내며, 식 (8)과 같이 정의된다. 재현율은 FN을 최소화하는 데 중점을 두며, 탐지되지 않은 침입의 비율을 줄이는 것이 중요하다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$\mathrm{P}recision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

실험 결과, Fuzzy 공격과 관련된 특정 노드 ID(0xA1)에서 99.72%의 공격 탐지 정확도를 보였고, 100%의 정밀도, 100%의 재현율을 나타내었다. 또한 제한된 시스템은 다양한 공격 시나리오에서도 비슷하게 높은 성능을 유지했다. 특히 차량의 RPM과 관련된 CAN ID(0x316)의 RANSAC 알고리즘 추정에서는 정확도, 정밀도, 재현율에서 모두 100%의 성능을 보였다.

## Ⅳ. 결론

본 논문에서는 RANSAC 알고리즘을 적용하여 경량화된 차량 내 침입 탐지 시스템을 구현하였다. 이 시스템은 차량 네트워크의 핵심 요소인 CAN ID의 데이터 흐름을 분석하고, 시간 간격 기반 데이터 패턴을 일반화하여 침입을 탐지하는 데 주력했다. 실험 결과, 제안된 시스템은 높은 정확도와 낮은 자원 소모를 바탕으로 기존 비지도 학습 모델보다 우수한 성능을 보였다. 이는 제안된 시스템이 차량의 실시간 데이터 처리 환경에서도 효과적으로 작동할 수 있음을 입증하며, 연산량과 메모리 사용이 많은 기존 딥러닝 기반 침입 탐지 시스템에 비해 큰 이점을 제공한다는 것을 보여준다. 그러나 RANSAC 알고리즘 기반의 침입 탐지 시스템은 성능은 뛰어나지만, 특정 ID를 알아야 한다는 단점이 존재한다. 향후 연구에서는 RANSAC 알고리즘의 성능을 더욱 향상시키기 위해 하이퍼파라미터 튜닝이나 적응형 학습 기법을 적용하고, 메모리 사용을 최소화하면서 탐지 효율을 높이는 전략을 개발할 것이다. 또한, 실시간 처리 성능을 극대화하기 위해 병렬 처리 기술이나 GPU 가속을 활용하는 방안을 고려할 수 있다. 이러한 추가 연구를 통해 차량 내 네트워크의 안전성을 더욱 강화하고, 자율 주행 및 V2X 통신 등의 첨단 기능을 안전하게 구현할 수 있을 것으로 기대된다.

## References

[1] E. Aliwa, C. Perera, and O. Rana, “Cyberattacks and Countermeasures for In-Vehicle Networks,” https://doi.org/10.48550/arXiv.2004.10781

[2] S. Gao, “Attack Detection for Intelligent Vehicles via CAN Bus: A Lightweight Image Network Approach,” *IEEE Transactions on Vehicular Technology*, pp.16624, 2023. DOI: 10.1109/TVT.2023.3296705

[3] H. Song, “In-vehicle Network Intrusion Detection using Deep Convolutional Neural Network,” *Vehicular Communications,* vol.21. pp.100198, 2020. DOI: 10.1016/j.vehcom.2019.100198

[4] E. Seo, “GIDS: GAN based Intrusion Detection System for In-Vehicle Network,” *Proceedings of the IEEE Annual Conference on Privacy, Security and Trust*, 2018. DOI: 10.1109/PST.2018.8514157

[5] Hacking and Countermeasure Research Lab, “Car-Hacking Dataset,” https://ocslab.hksecurity.net/Datasets/car-hacking-dataset

## BIOGRAPHY

**Jonggwon Kim** (Member)

2019~ : Candidate for BS degree in Electronic Engineering, Soongsil University.

〈Main interest〉 Automotive SoC, AI SoC, Processor SoC

**Hyungchul Im** (Member)

2021 : BS degree in Mechanical Engineering, Soongsil University.
2021~: Candidate for Ph.D degree in Electronic Engineering, Soongsil University.
〈Main Interest〉 Vehicle Security, Artificial Intelligence, Automotive SoC

**Joosock Lee** (Member)

1983 : BS degree in Electronic Engineering, Sogang University.
1985 : MS degree in Electronic Engineering, Korea University.
1999 : PhD degree in Electrical Engineering, Korea University.
1985~1995 : Senior Engineer, LG Central Laboratory
2004~2005 : CTO, MtekVision Ltd
2006~2010 : Chief of SoC Center, Chungbuk Technopark
2022~Now : Professor in School of Electronic Engineering, Soongsil University
〈Main Interest〉 AI SoC, Power Management SoC, Battery Management SoC

**Seongsoo Lee** (Life Member)

1991 : BS degree in Electronic Engineering, Seoul National University.
1993 : MS degree in Electronic Engineering, Seoul National University.
1998 : PhD degree in Electrical Engineering, Seoul National University.
1998~2000 : Research Associate, University of Tokyo
2000~2002 : Research Professor, Ewha Womans University.
2002~Now : Professor in School of Electronic Engineering, Soongsil University.
〈Main Interest〉 AI SoC, Automotive SoC, Security SoC, Processor SoC, Power Management SoC, Battery Management SoC, Reliability and Safety.