https://doi.org/10.7840/kics.2025.50.12.1842

# 차량용 침입 탐지 에이전트를 위한 순환 신경망 성능 연구

박 시 연◆, 최 다 영*, 박 형 곤°

# A Study on Recurrent Neural Network Performance for In-Vehicle Intrusion Detection Agents

Siyoun Park◆, Dayoung Choi*, Hyunggon Park°

요 약

Controller Area Network (CAN) bus 프로토콜은 네트워크 보안 취약성으로 인해 메시지 주입 공격에 취약하다는 구조적 한계가 있어 이를 위한 침입 탐지 에이전트가 요구된다. 본 논문에서는 메시지 주입 공격에 대한 효과적인 침입 탐지를 위해서, 시계열적 패턴을 파악할 수 있는 4가지 종류의 순환 신경망으로 침입 탐지를 수행하여 성능을 비교하고, 신경망 구조가 성능 변화에 미치는 영향을 분석하였다.

**Key Words :** CAN, Intrusion detection agent, Recurrent neural network, Multi-class classification

ABSTRACT

The lack of security mechanism of Controller Area Network (CAN) bus protocol introduces structural vulnerabilities, requiring intrusion detection agents for message injection attacks. This study presents an intrusion detection system based on Recurrent Neural Networks (RNNs), capturing time-series patterns. We compare the performance of various RNNs and analyze how each architecture influences its detection capability.

## Ⅰ. 서 론

최근 차량 내부 네트워크의 연결성과 자동화 수준이 높아지면서, 차량 시스템 보호를 위한 지능형 침입 탐지 에이전트(agent)의 필요성이 대두되고 있다. 특히, 차량 내 통신을 담당하는 Controller Area Network (CAN) bus 프로토콜은 메시지 인증이나 암호화 같은 보안 기능이 부족하여 기능 마비 등 심각한 사고를 야기할 수 있어 통신 패턴을 분석하고 이상 징후를 탐지하는 지능형 보안 시스템의 도입이 필수적이다. CAN 메시지 패킷들은 차량 내 노드들의 식별자 우선순위에 따라 처리 순서가 결정되기 때문에, 교환되는 메시지의 순서나 주기를 변경시켜 혼선을 유발하는 공격이 주로 수행된다[1]. 따라서, 순서와 시간 정보 반영을 통한 신속한 탐지가 필수적이기에 순환 신경망(Recurrent Neural Network, RNN)을 비롯한 딥러닝 모델을 기반으로 침입 탐지 시스템을 구축하려는 연구가 이루어지고 있다[1-9].

본 논문에서는 Long Short-Term Memory (LSTM)[10], Gated Recurrent Unit (GRU)[11], Bi- directional LSTM (BiLSTM), Bidirectional GRU (BiGRU)[12]를 이용하여 침입 탐지 성능을 시퀀스 길이에 따라 비교하고 분석한다.

## Ⅱ. CAN 메시지 데이터

본 논문에서는 그림 1과 같은 다차원 시계열 형태의 CAN 메시지 데이터를 고려하며, Timestamp, CAN ID, DLC (Data Length Code), Payload 등 $M$개의 특징으로 이루어져 있다. 총 $T$개의 시점에서 수집된 전체 메시지 데이터 $\boldsymbol{D} \in \mathbb{R}^{T \times M}$ 중에서, 신경망에 입력되는 데이터는 연속된 시퀀스 길이 $L(L \leq T)$개로 구성된

※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2021-0-00739, 분산/협력 AI 기반 5G+ 네트워크 데이터 분석 기능 및 제어 기술 개발)과 2024년도 (No.RS-2024-00344830, 6G 네트워크 구조/산업융합 표준기술 개발 및 표준화)을 받아 수행된 연구임.

◆ First Author : (ORCID:0009-0003-9707-9041) Ewha Womans University, Department of Electronic and Electrical Engineering, siyoun0116@ewha.ac.kr, 학생(학사과정), 학생회원

° Corresponding Author : (ORCID:0000-0002-5079-1504) Ewha Womans University, Department of Electronic and Electrical Engineering, hyunggon.park@ewha.ac.kr, 교수, 종신회원

* (ORCID:0009-0008-9270-0528) Ewha Womans University, Department of Electronic and Electrical Engineering, dayoung.choi@ewha.ac.kr, 학생(석/박사통합과정), 학생회원

논문번호 : 202505-124-A-LU, Received May 28, 2025; Revised June 30, 2025; Accepted June 30, 2025

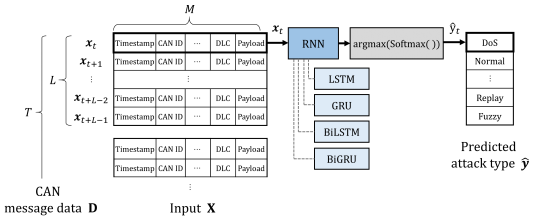


그림 1. 순환 신경망 기반 차량 CAN 침입 탐지 시스템
Fig. 1. Intrusion Detection System for In-Vehicle CAN based on RNN

$\boldsymbol{X} \in \mathbb{R}^{L \times M}$이다. 따라서 순환신경망은 고정된 길이의 연속적인 메시지들을 순차적으로 분석할 수 있으며, $L$이 짧을수록 탐지에 요구되는 시간 길이와 처리 데이터의 양이 감소하므로 더욱 빠른 침입 탐지에 유리하다.

실제 공격 유형 $\boldsymbol{y} \in \mathbb{R}^{L}$은 Normal, DoS, Fuzzy, Replay로 구성되며, 예측 $\hat{\boldsymbol{y}} \in \mathbb{R}^{L}$는 신경망 출력에 Softmax 함수를 적용한 후 가장 높은 확률을 갖는 클래스로 결정된다. Normal은 어떠한 공격도 수행되지 않은 정상 상태를, DoS는 높은 우선순위의 식별자를 갖는 다량의 메시지가 빠른 주기로 주입되어 정상 메시지의 전송을 방해하는 공격을, Fuzzy는 랜덤한 식별자를 가진 임의의 payload가 포함된 메시지를 주입하는 공격을, Replay는 특정 시간 동안의 유효한 정상 메시지를 재주입하는 공격을 의미한다[1].

## Ⅲ. 차량 CAN 침입 탐지를 위한 순환 신경망

### 3.1 LSTM

LSTM은 3개의 게이트(gate)와 2개의 상태(state)로 구성되며, 특정 시점 $t$에서의 입력 벡터 $\boldsymbol{x}_t \in \mathbb{R}^{M}$에 대해 수식 (1)과 같이 나타낼 수 있다.

$$\begin{aligned} \boldsymbol{i}_t &= \sigma(\boldsymbol{W}_i \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \\ \boldsymbol{f}_t &= \sigma(\boldsymbol{W}_f \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f) \\ \boldsymbol{o}_t &= \sigma(\boldsymbol{W}_o \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \\ \tilde{\boldsymbol{C}}_t &= \tanh(\boldsymbol{W}_c \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_c) \\ \boldsymbol{C}_t &= \boldsymbol{f}_t * \boldsymbol{C}_{t-1} + \boldsymbol{i}_t * \tilde{\boldsymbol{C}}_t \\ \boldsymbol{h}_t &= \boldsymbol{o}_t * \tanh(\boldsymbol{C}_t) \end{aligned} \tag{1}$$

$\boldsymbol{W}_i$, $\boldsymbol{W}_f$, $\boldsymbol{W}_o$, $\boldsymbol{W}_c$는 각 게이트와 상태에 대응하는 가중치 행렬, $\boldsymbol{b}_i$, $\boldsymbol{b}_f$, $\boldsymbol{b}_o$, $\boldsymbol{b}_c$는 편향 벡터이다. 입력 게이트 $\boldsymbol{i}_t$는 새로운 정보를, 망각 게이트 $\boldsymbol{f}_t$는 과거 정보를 반영하는 것을 조절하며, 출력 게이트 $\boldsymbol{o}_t$는 입력 게이트와 망각 게이트가 모두 적용된 셀 상태 $\boldsymbol{C}_t$를 통해 출력되는 정보를 결정한다. 셀 입력 $\tilde{\boldsymbol{C}}_t$는 $\boldsymbol{C}_t$에 새로 추가되는 입력 후보를, 은닉 상태 $\boldsymbol{h}_t$는 $t$에서의 출력을 의미한다.

### 3.2 GRU

GRU는 2개의 게이트와 1개의 상태로 LSTM을 간략화한 구조로, 수식 (2)와 같이 표현할 수 있다.

$$\begin{aligned} \boldsymbol{r}_t &= \sigma(\boldsymbol{W}_r \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_r) \\ \boldsymbol{z}_t &= \sigma(\boldsymbol{W}_z \cdot [\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_z) \\ \tilde{\boldsymbol{h}}_t &= \tanh(\boldsymbol{W}_h \cdot [\boldsymbol{r}_t * \boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_h) \\ \boldsymbol{h}_t &= \boldsymbol{z}_t * \boldsymbol{h}_{t-1} + (1 - \boldsymbol{z}_t) * \tilde{\boldsymbol{h}}_t \end{aligned} \tag{2}$$

$\boldsymbol{W}_r$, $\boldsymbol{W}_z$, $\boldsymbol{W}_h$는 각 게이트와 상태에 대응하는 가중치 행렬, $\boldsymbol{b}_r$, $\boldsymbol{b}_z$, $\boldsymbol{b}_h$은 편향 벡터이다. 초기화 게이트 $\boldsymbol{r}_t$는 과거의 은닉 상태 $\boldsymbol{h}_{t-1}$의 반영 정도를, 업데이트 게이트 $\boldsymbol{z}_t$는 $\boldsymbol{h}_{t-1}$와 새로운 은닉 상태 $\tilde{\boldsymbol{h}}_t$의 비율을 조정하여 $\boldsymbol{h}_t$ 갱신 정도를 결정한다. 따라서, GRU의 $\boldsymbol{h}_t$는 LSTM의 $\boldsymbol{C}_t$와 $\boldsymbol{h}_t$가 통합되어 표현된 것으로써, GRU가 LSTM의 단순화된 형태라는 것을 알 수 있다.

### 3.3 BiLSTM, BiGRU

BiLSTM과 BiGRU는 각각 LSTM과 GRU의 입력을 순방향과 역방향에서 모두 처리하는 것으로, 순방향 계층(layer)에서는 입력 데이터가 그대로 처리되며, 역방향 계층에서는 입력 데이터가 반전되어 역으로 처리된다. 따라서 과거와 미래의 정보를 모두 활용하여 보다 정교한 학습이 가능하다.

## Ⅳ. 성능 비교 실험

### 4.1 실험 설정

본 연구에서 사용한 데이터셋은 KU-CISC2017-OTIDS-2nd[13,14]로, Normal, DoS, Fuzzy, Replay 메시지를 포함한다. $M = 11$로, 원본 데이터셋에서 단일값으로 구성된 FLAG는 제외하고, CAN ID, DLC, Payload[1-8], logIAT (log Inter Arrival Time)를 특징(feature)으로 사용하였다. logIAT는 각 CAN ID에 대해 현재와 이전 메시지 사이의 시간 간격의 로그를 취한 값으로, Timestamp를 대신하여 사용하였다. 여러 순환 신경망 종류에서 시퀀스 길이에 따른 탐지 성능 비교를 위해 $L \in \{100, 300, 500\}$으로 설정하였다. 성능 평가 지표로는 Matthews Correlation Coefficient (MCC), Precision, Recall, F1-score를 사용하였다. MCC는 불균형한 데이터셋에서도 전반적인 성능을 균형 있게 평

표 1. $L$에 따른 순환 신경망 종류별 차량 CAN 침입 탐지 성능
Table 1. In-Vehicle CAN Intrusion Detection Performance of RNNs for $L$

| | $L=100$ | | | | $L=300$ | | | | $L=500$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MCC | Precision | Recall | F1-score | MCC | Precision | Recall | F1-score | MCC | Precision | Recall | F1-score |
| LSTM | 0.7988 | 0.8565 | 0.8910 | 0.8728 | 0.8375 | 0.8795 | 0.9219 | 0.8981 | 0.8123 | 0.8574 | 0.9070 | 0.8797 |
| GRU | 0.8252 | 0.8773 | 0.9082 | 0.8914 | 0.8659 | 0.8885 | **0.9415** | 0.9128 | **0.8949** | **0.9110** | **0.9594** | **0.9325** |
| BiLSTM | 0.8282 | 0.8725 | 0.9108 | 0.8904 | 0.8194 | 0.8704 | 0.9088 | 0.8874 | 0.8262 | 0.8702 | 0.9117 | 0.8895 |
| BiGRU | **0.8487** | **0.8867** | **0.9301** | **0.9056** | **0.8894** | **0.9373** | 0.9235 | **0.9297** | 0.8877 | 0.9089 | 0.9562 | 0.9293 |

가하는 지표이다. Precision은 시스템이 특정 클래스로 예측한 것 중에서 실제로 해당 클래스에 속하는 비율을 의미하며, 예측의 정밀도를 나타낸다. Recall은 재현율로, 실제로 특정 클래스에 속하는 것 중에서 시스템이 올바르게 예측한 비율을 의미한다. F1-score는 Precision과 Recall의 조화평균으로, 정밀도와 재현율을 모두 고려한 결과이다.

### 4.2 실험 결과

표 1은 $L$에 따른 순환 신경망 종류별 차량 CAN 침입 탐지 성능을 비교한 결과이다. 실험 전반에 걸쳐 BiGRU와 GRU, 즉 GRU 계열의 순환 신경망이 LSTM 계열의 신경망보다 일관적으로 우수한 성능을 보였으며, 이는 CAN 메시지에 GRU 계열의 비교적 단순한 구조가 더 적합함을 시사한다. 특히, $L=500$에서 GRU가 가장 뛰어난 성능을 보였다. 또한, 양방향 구조인 BiGRU와 BiLSTM이 각각 GRU와 LSTM 대비 더욱 우세한 성능을 보였다. 이는 메시지의 과거와 미래 정보를 모두 활용하는 양방향 순환 구조가, 시계열의 시간 및 순서 정보가 중요한 메시지 주입 공격 탐지에 효과적임을 나타낸다. $L=100$인 경우에 BiGRU가 가장 우수한 성능을 보여, 빠른 탐지와 연산 효율이 중요한 차량 환경에 가장 유리한 구조로 판단된다.

그림 2는 공격 유형별 탐지 성능을 나타낸 것이다. Normal, DoS, Fuzzy는 조건에 상관없이 높은 탐지 성능을 유지했다. 반면, 정상 메시지 묶음을 가공 없이 재전송하는 공격 방식인 Replay는 다른 유형에 비해 낮은 성능을 보였다. 이는 Replay 메시지가 정상 메시지와 매우 유사하여 구분이 어렵기 때문이다. 그러나, 학습 시 $L$이 증가함에 따라 GRU 계열 신경망에서 성능이 향상되는 경향으로 보아, 장기적인 시계열 정보를 활용하는 것이 Replay 공격을 탐지하는 데에 효과적임을 알 수 있다. 따라서, Replay 공격 대응 측면에서도 GRU 기반 구조가 효과적임을 확인하였다.

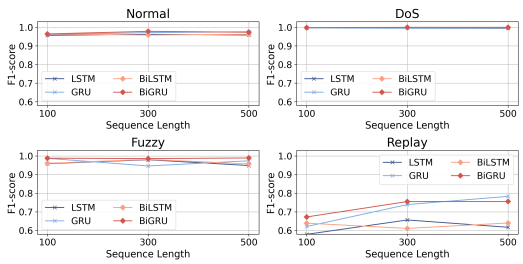


그림 2. 공격 유형별 순환 신경망의 차량 CAN 침입 탐지 성능
Fig. 2. In-Vehicle CAN Intrusion Detection Performance of RNNs for Attack Classes

## V. 결 론

본 논문에서는 차량 CAN에서 발생하는 메시지 주입 공격 탐지를 목적으로 4가지 순환 신경망 LSTM, GRU, BiLSTM, BiGRU의 탐지 성능을 비교하였다. 실제 차량 환경에서 수집된 KU-CISC 2017-OTIDS 데이터셋을 활용하여 실험한 결과, GRU 계열 신경망이 간결한 구조로 인해 뛰어난 일반화 능력을 보이며 일관적으로 우수한 탐지 성능을 보였다. 향후 특정 공격 유형에 특화된 침입 탐지 에이전트를 설계할 때, 본 연구 결과를 기반으로 순환 신경망 구조와 시퀀스 길이를 전략적으로 설정한다면, 더욱 정밀하고 효과적인 침입 탐지 시스템 구축이 가능할 것으로 기대된다.

## References

[1] M. Almehdhar, et al., "Deep learning in the fast lane: A survey on advanced intrusion detection systems for intelligent vehicle networks," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 869-906, Jul. 2024.

[2] S. Park, et al., "A study on intrusion Detection for in-vehicle CAN based on xLSTM memory structures," in *Proc. JCCI 2025*, pp. 295-296, Apr. 2025.

[3] D. Choi, et al., "Attack-specific feature

analysis framework for NetFlow IoT datasets," *Comput. & Secur.*, vol. 157, no. 104536, Oct. 2025.

[4] J. Rheey and H. Park, "Robust hierarchical anomaly detection using feature impact in IoT networks," *ICT Express*, vol. 11, no. 2, pp. 358-363, Apr. 2025.

[5] J. Rheey, et al., "Symmetrical pruning for lightweight network anomaly detector," *ACM MobiSys 2024*, pp. 634-635, Jun. 2024.

[6] J. Rheey and H. Park, "Game-theoretic lightweight autoencoder design for intrusion detection," in *Proc. IEEE WCNC 2024*, pp. 1-6, Apr. 2024.

[7] M. Kim, et al., "Improving network attack classification on imbalanced real-world intrusion incident datasets," in *Proc. ACM MobiSys 2025*, Jun. 2025.

[8] H. Kye, et al., "Hierarchical detection of network anomalies: A self-supervised learning approach," *IEEE Signal Process. Lett.*, vol. 29, pp. 1908-1912, Aug. 2022.

[9] J. Kwon, et al., "Automatic classification of network traffic data based on deep learning in ONOS platform," in *Proc. ICTC 2020*, pp. 1028-1030, Oct. 2020.

[10] S. Hochreiter, et al., "Long short-term memory," *Neural comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

[11] K. Cho, et al., "Learning Phrase Representations using RNN encoder-decoder for statistical machine translation," in *Proc. EMNLP 2014*, pp. 1724-1734, Oct. 2014.

[12] M. Schuster, et al., "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673-2681, Nov. 1997.

[13] H. Lee, et al., "OTIDS: A novel intrusion detection system for in-vehicle network by using remote frame," in *Proc. PST 2017*, pp. 57-66, Aug. 2017.

[14] H. Song, et al., "Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network," in *Proc. ICOIN 2016*, pp. 63-68, Jan. 2016.