White paper [PRELIMINARY- April 09, 2021]

Intrusion Detection System for On-Board Vehicle Communication

Artificial Intelligence based, supervised trained networks achieve state of the art performances in classification of cyber attacks

Overview

This White Paper is based on the talk presented at ITASEC 2021, the Italian Conference on Cybersecurity on April, 9th by Tobia Fiorese (Bluewind) and Pietro Montino.
Since the introduction of many external interfaces in modern vehicles exposes users to the risk of cyber-attacks, the need of focus on security is concrete. This paper presents the development of an Intrusion Detection System (IDS) to be deployed on a CAN bus. An IDS is a component that can detect anomalies in the behaviour of the system where it is deployed.

This machine learning based solution is composed of two parts. The first includes a supervised trained neural network that is able to distinguish among different known attacks. The second includes a discriminator that has been trained exploiting the Generative Adversarial Network (GAN) paradigm, to distinguish among the attack-free situation and an anomalous situation.

Bluewind presents here a system where supervised trained networks achieve state of the art performances in classification of known attacks and unsupervised trained networks guarantee a certain level of security, helping to strengthen the system in any case where models are weak and data are poor or missing.

Background

A connected car is a car capable of communicating bidirectionally with systems located outside of itself or with internal devices. Among them are: the collection of real-time traffic information, accident emergency services, access to status and operating conditions of the car. All these communication interfaces expose the vehicle internal networks to remote attacks, thus increasing the vehicles' cyber vulnerability.

In the future, each car will constantly communicate with the surrounding environment made of other cars, pedestrians and road signs giving rise to what is called the Vehicle to Everything (V2X) paradigm.
In a modern vehicle there are many kinds of networks: the CAN bus is the de facto standard for safety critical applications. The CAN bus connects hundreds of Electronic Control Units (ECUs) and controls most, if not all, of the electro-mechanical actuated systems: from brakes to lights, from airbags to transmission.

In today's connected world, finding ways to prevent cyber-attacks that could damage on-board systems and put the vehicle and the passengers at risk, has become of pivotal importance.

This paper presents the architecture and the safety performances of a two-step CAN bus IDS trained with machine learning techniques. This approach, avoiding the use of authentication protocols and encryption methods, allows to add a measure of safety, even for on-board electronics architectures already consolidated, without too much interference with pre-existent hardware and software.

How attacks on CAN Bus works

Most of the known cyber-attacks on vehicles' CAN bus network are based on insertion or deletion of packets in transit, or the manipulation of their content.

To be carried out, they leverage some weaknesses of the CAN bus as the multicast nature of the network, the lack of authentication, addressing and encryption and the common point of entry.

Three of the most diffused have been accounted on the development of this IDS, and are listed below in ascending order of implementation difficulties.

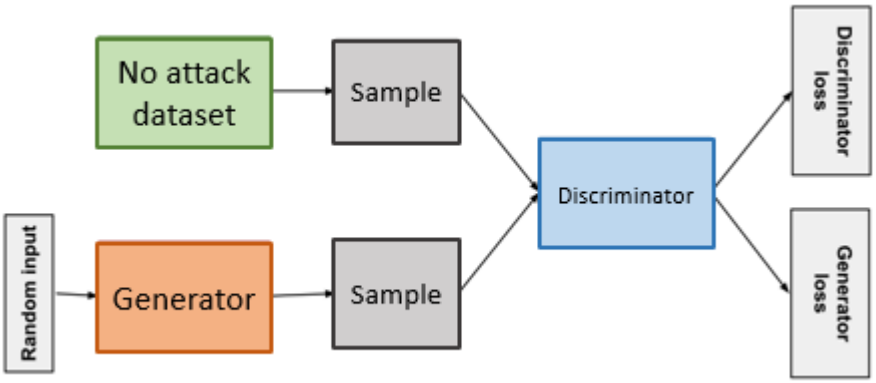
- **Denial of Service (DoS)**: An attacker can inject high priority messages in a short cycle on the bus. DoS attack messages aim at occupying the bus using the theoretically highest priority identifier, namely 0x000. Since all nodes share a single bus, increasing occupancy of the bus can produce latencies of other messages and cause threats regarding availability with no response to driver's commands.
- **Fuzzy:** An attacker can inject messages with randomly spoofed CAN ID, either with arbitrary data or with spoofed data values. All these messages are functional and structurally correct, but they can cause unintended vehicle behaviors. An attacker can passively observe in-vehicle traffic and select target identifiers to produce unexpected behaviors. Unlike the DoS attack, Fuzzy is more specific and aims at paralyzing a particular function of the vehicle.
- **Impersonation:** An attacker can manage to stop message transmission from a target node and can plant/manipulate an impersonating node that will take its role. If a victim node stops transmitting, all messages sent by the targeted node will be removed from the bus.

How GAN training works

A drawback of the supervised training approach to Machine Learning algorithms is that it relies on known data. In our case this means a dataset containing information about known attacks must be collected.
However, despite the difficulties in providing such a dataset, even with a lot of data this approach will not guarantee new attacks will be identified. Slight variations to attacks forming the dataset will increase the possibility to be confused with normal situations.

It is possible though to train a network to generate data similar to the ones of the given dataset, and at the same time train a second network to distinguish among generated data and real data. This mechanism is called Generative Adversarial Network training.

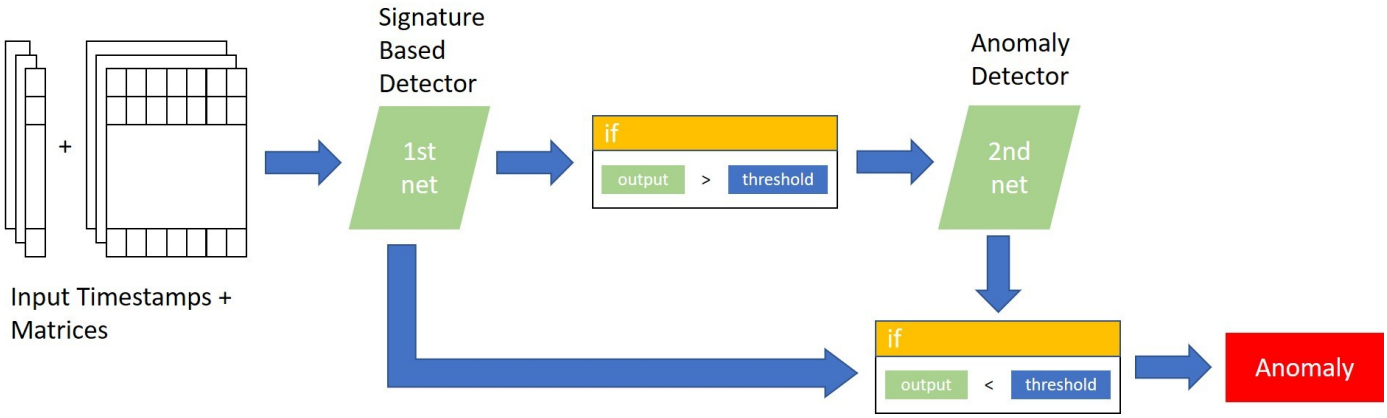
This approach, in theory may result in a reliable discriminator that enables protection against unknown attacks. Used in combination with supervised models, it may help to reduce the false positive samples yielded by such networks.



Proposed solution

The input data of the proposed IDS are windows of traffic logged directly from the bus, encoded as matrices. The only features useful to identify the three attacks mentioned above turned out to be the ID of the messages in transit and the corresponding timestamps. This results in a solution that does not need to know the particular semantics of the messages transmitted by each device.

The system is composed of two subsequent networks. The first one acts as a signature based detector, and thus recognizes attacks based on previously seen examples. The second acts as an anomaly detector: in the training phase only attack-free data has been fed to the network, but still, exploiting the GAN paradigm, the result is a fairly good detection of deviations from the normal operation.



The detection is simply based on the comparison between network outputs and two thresholds defined during training and test phases. The algorithm works as follows:

1. A buffer of messages of fixed length is encoded as a matrix of CAN IDs and a vector of associated timestamps
2. It is fed to the first network, that gives as output a score value (scores are higher for attack-free inputs)
3. The output is compared to a first threshold, that have been fixed during tests on training phase
4. If the output is lower than the threshold, an attack is detected
5. If the output is higher than the threshold then the input is forwarded to the second network to perform a double check
6. Steps 3 to 5 are repeated on the second network with a different threshold

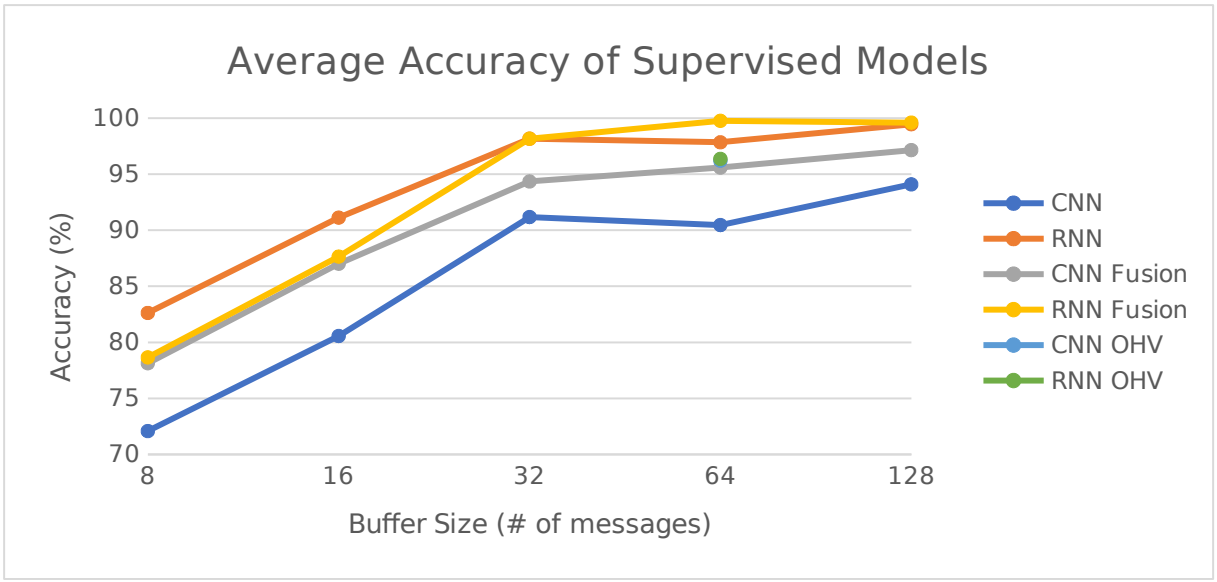
Many different setups have been tested, comprising either the encoding of raw data and the architectures of neural networks. The best choices resulted in the use of One-Hot Vector encoding, the use of Recurrent Neural Networks based on LSTM as first detector and a two-dimensional Feedforward Neural Network as second detector.

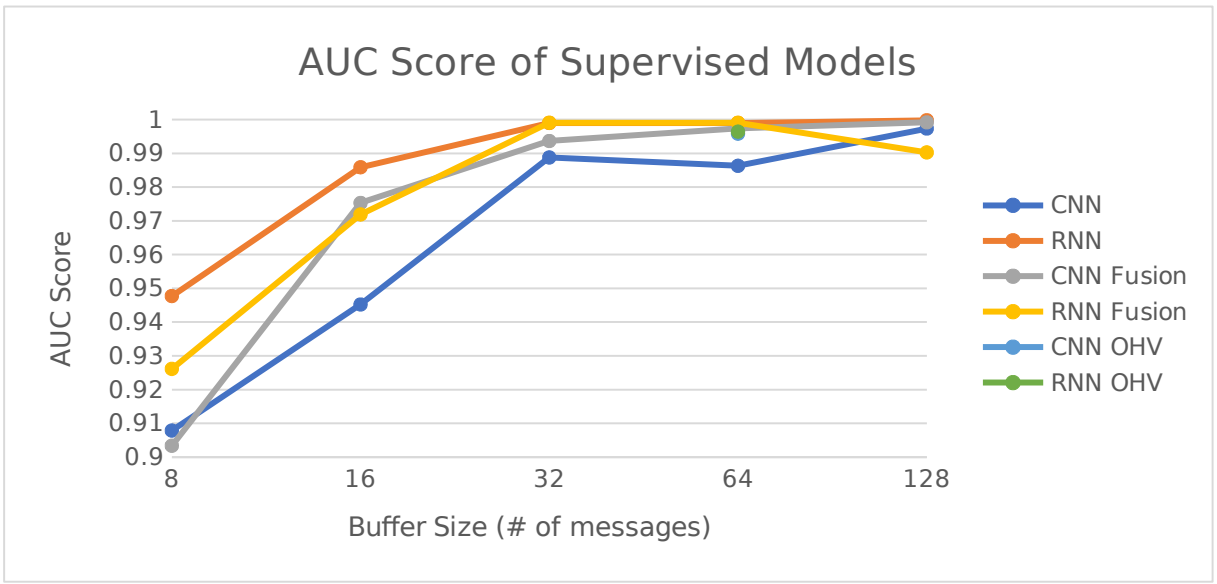
Results

Experiments on two datasets collected from cars where attacks have been carried out during drivings points out three key aspects.

Low-complexity supervised networks, such as those implemented in the first step of the system, can reach outstanding performances in classifying known attacks based on insertion or deletion of messages.

Since the proposed method does not account on content of packets, attacks based on manipulating the payload only will be more difficult to identify. However, an extension to the use of more features could improve detection also on that side.



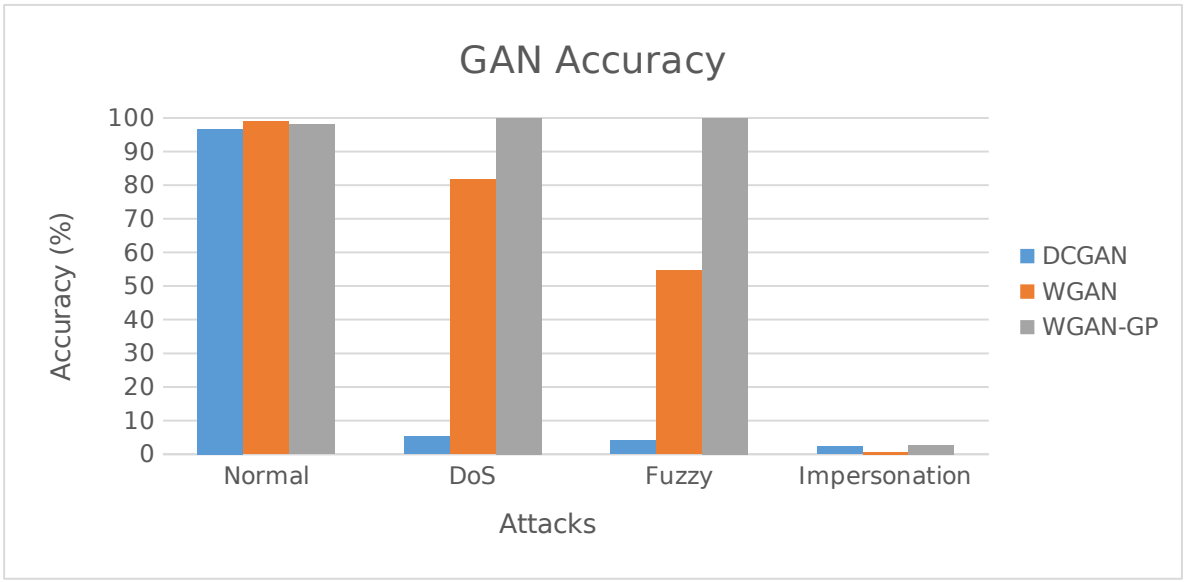


On the two images it is clearly depicted a trend where accuracy in attacks detection grows with longer temporal windows of CAN bus traffic and the addition of timestamps to IDs sequences only (Data Fusion).

The encoding plays an important role also, since One-Hot Vectors (OHV) improved performances when relying on Convolutional Neural Networks. However, the best solution plans to use Recurrent Neural Networks being the relationship of IDs sequences temporal rather than spatial.

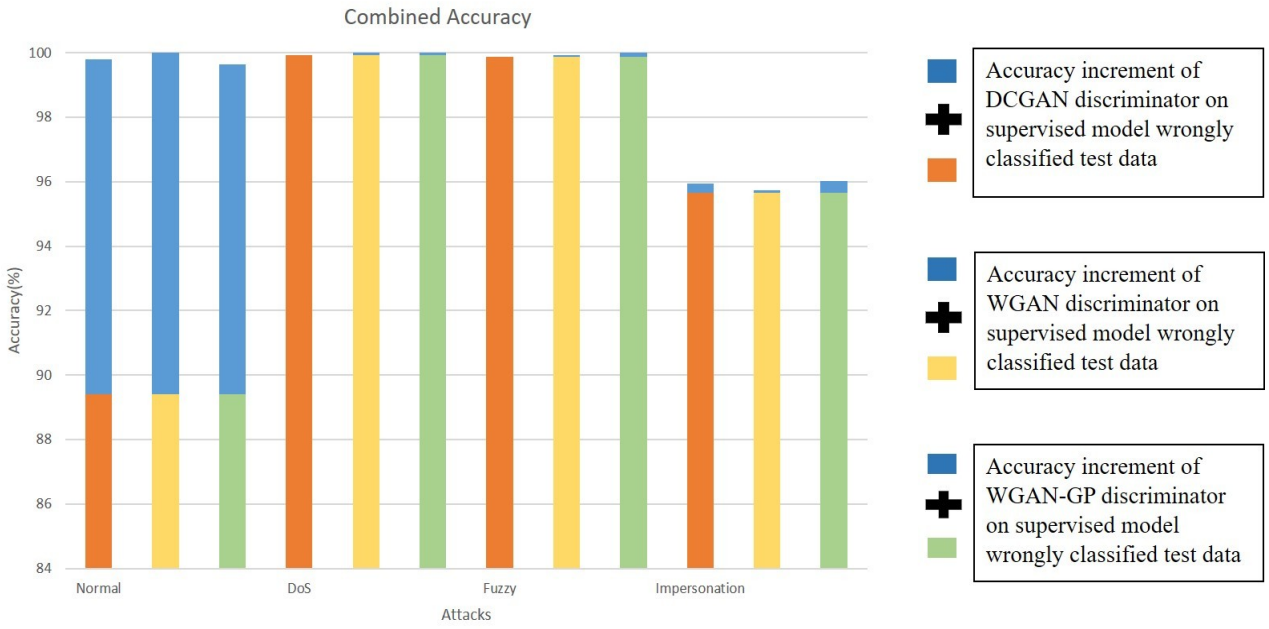
In this case, the use of timestamps become negligible, making the system even more flexible and easily adaptable in already consolidated contexts.

Datasets with various scenarios of cyber-attacks on CAN bus is not strictly necessary, as using the recent unsupervised training method based on generative adversarial networks results in a certain degree of security.



The image above clearly shows how, in two out of three cases, it is possible to identify an attack in progress on a system having no data available regarding that specific attack beforehand.

The combination of supervised and unsupervised approaches to machine learning algorithms leads to an improvement in terms of accuracy even on datasets that are small, incomplete or weak.



In this last case, the image shows the improvement in accuracy gained by the subsequent use of unsupervised models on data that were wrongly classified by supervised models. In this test, a weak discriminator was used as first step in detection in order to assess the benefits of unsupervised training.

Conclusions

The proposed system uses advanced machine learning techniques to analyze raw data exchanged on CAN bus in order to detect cyber-attacks, resulting in a fairly accurate anomaly detection that is not necessarily based on knowledge of cyber threatens against the system where it may be deployed.

About Bluewind

Bluewind, an independent engineering company, provides world-class products, engineering and software solutions in the domains of electronics, safety critical applications, and connected devices.
As a qualified researcher in the Safety and Cybersecurity domains, Bluewind is actively involved in designing next generation products in the Automotive, Industrial and Medical sectors

Bluewind Srl

Via della Borsa, 16/A - 31033
Castelfranco Veneto (TV) - Italy
+39 0423 723431 - info@bluewind.it

www.bluewind.it