Reinforcement Learning-based Misbehaviour Detection in V2X Scenarios

Roshan Sedar*, Charalampos Kalalas*, Francisco Vázquez-Gallego*, Jesus Alonso-Zarate†
*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Barcelona, Spain
†i2CAT Foundation, Barcelona, Spain
{roshan.sedar, ckalalas, francisco.vazquez}@cttc.es, jesus.alonso@i2cat.net

***Abstract*—Emerging vehicle-to-everything (V2X) services rely on the secure exchange of periodic messages between vehicles and between vehicles and infrastructure. However, transmission of false/incorrect data by malicious vehicles may pose important security perils. Therefore, it is essential to detect safety-threatening erroneous information and mitigate potentially detrimental effects on road users. In this paper, we assess the effectiveness of a reinforcement learning (RL) approach for misbehaviour detection in V2X scenarios using an open-source dataset. Considering the case of sudden-stop attacks, the performance of RL-based detection is evaluated over commonly used detection metrics. Our research outcomes reveal that misbehaving vehicles can be accurately detected by exploiting real-time position and speed patterns.**

***Index Terms*—V2X, Misbehaviour Detection, Reinforcement Learning.**

I. Introduction

Pervasive vehicle-to-everything (V2X) connectivity and the emergence of effective data-driven methods based on artificial intelligence and machine learning (AI/ML) drive a paradigm shift towards connected and automated mobility (CAM) services and applications [1]. A key functionality that can benefit from AI/ML is cybersecurity, which is essential for ensuring road safety in CAM environments [2]. V2X security threats and attacks can be originated from malicious outsiders and/or insiders. In contrast to an outsider user, an insider possesses valid credentials to interact with other legitimate entities in the system [3]. Sophisticated insider attacks are often difficult to detect and contain, particularly when attackers behave intelligently while conforming to normal system behaviour. Malicious/selfish behaviours from such rogue insiders are commonly referred to as misbehaviours in V2X, and they pose a serious threat when transmitting erroneous/incorrect data in safety-critical situations. Ensuring the semantic correctness of exchanged V2X information is thus of paramount importance.

In existing literature, several data-driven approaches have been proposed to detect misbehaving vehicles, some of which rely on statistical and conventional ML techniques [4]. Nevertheless, current misbehaviour detectors are not designed to dynamically improve their detection experience according to evolving attack patterns in rapidly changing V2X environments. In addition, the use of security thresholds in detectors (e.g., anomaly score-based methods) limit their applicability to very specific V2X scenarios. To this end, reinforcement learning (RL) can be identified as an effective approach to deal with misbehavior detection in V2X. RL-based detection algorithms can improve their detection experience over time from the interactions with unknown environments without relying on security threshold values.

In this paper, we assess the applicability and performance of the RL-based anomaly detection method proposed in [5] for identification of sudden-stop attacks in V2X time-series data. Such attacks may lead to unnecessary traffic congestion and potential road accidents due to hard braking. They can be difficult to detect due to the attacker's erratic behaviour over time; attacker may behave normally for a specific time-period, and transmit repeatedly falsified information, i.e., fixed-position coordinates and zero-speed values, in subsequent time-steps. We adopt an ensemble approach utilizing multiple features of the sudden-stop attack (e.g., position, speed, heading angle, etc.) to train the RL model separately for each feature. We further assess the detection performance of the RL-based method for each selected feature while observing that some features yield superior performance over others. For performance evaluation, we have used the open-source VeReMi dataset [6], generated with a V2X simulator and including several V2X misbehaviour attacks, e.g., position falsification and sudden stop.

II. Reinforcement Learning Model

Time-series anomaly detection constitutes a sequential decision-making process that can be modelled as a Markov decision process [7]. The action of anomaly detection will change the environment based on the decision of either normal or anomalous behaviour at time t; subsequently, the next decision at time step $t+1$ will be influenced by the changing environment at previous time-step t. The application of an RL model is thus a natural fit for time-series anomaly detection. In the context of V2X, vehicle's mobility data is a time-series consisting of periodic beacon messages. Each beacon message includes information of the vehicle's speed, position, heading angle, etc., and this information is evolving over time along the vehicle's trajectory. Hence, misbehaving vehicles can be potentially detected by sequentially analysing their mobility patterns using an RL model. In what follows, we briefly discuss the components pertaining to the RL model introduced in [5] and applied for sudden-stop attack detection.

The **agent** is the core part of the RL model. It takes the time-series (i.e., vehicle's mobility data) and prior related

decisions as inputs (i.e., state s), and generates the new decision made (i.e., action a) as output. Each action made by the agent is rewarded (i.e., reward r) as feedback, and the agent subsequently updates its model in order to improve the accuracy in decision-making. The iterative model update is performed through Q-learning [8], i.e.,

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)), \quad (1)$$

where α and γ denote the learning rate and discount factor, respectively.

The **environment** of the RL model controls the training of the agent. It takes the action a performed by the agent as its input, and consequently generates a reward r and the next environment state s for the agent. The environment is a time-series repository of vehicles' mobility data and contains a large population of periodic beacon messages with sudden-stop attack labels.

The **state** contains two sequences: the sequence of previous actions, denoted by $s_{action} =< a_{t-1}, a_t, ..., a_{t+n-1} >$, and the current vehicle's time-series data, denoted by $s_{time} =< X_t, X_{t+1}, ..., X_{t+n} >$, where X_t is the value of a feature at time t. According to the state design, the next action taken by the agent is dependent on the previous actions and the current vehicle's information.

The **action** space is defined as $\mathcal{A}$ = {0,1} where 1 indicates the detection of an attack and 0 represents the normal behaviour. In a given state s, the agent selects the action as

$$a = \arg\max_a Q(s, a). \quad (2)$$

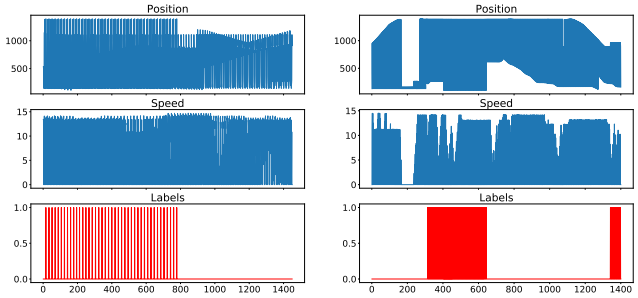
The **reward** r is offered to the agent when an action a is taken in state s. In particular, the agent is given a positive reward for correctly identifying the sudden-stop attack, i.e., true positive (TP), or a normal state, i.e., true negative (TN); otherwise, a negative reward is given to the agent for incorrect identification of a normal state as an attack, i.e., false positive (FP), or an attack as a normal state, i.e., false negative (FN). In safety-critical V2X scenarios, FNs are more hazardous than FP alarms; thus, an agent is penalised more for FN actions than for FPs. The reward function can be expressed as

$$r(s, a) = \begin{cases} A & \text{if the action is a TP,} \\ B & \text{if the action is a TN,} \\ -C & \text{if the action is an FP,} \\ -D & \text{if the action is an FN,} \end{cases} \quad (3)$$

where $A, B, C, D > 0$, with $A > B$ and $D > C$.

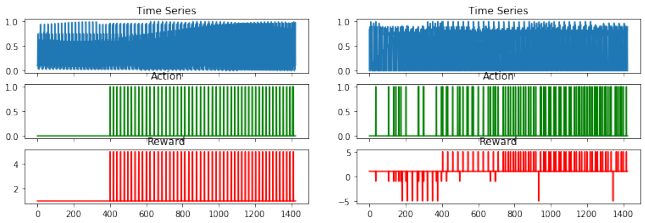
III. Experiments and Results

In this section, we demonstrate the effectiveness of the RL-based approach [5] applied for V2X misbehaviour detection by performing experiments on specific parts of VeReMi [6] dataset for detection of sudden-stop attacks.



(a) A sample of training data (b) A sample of test data

Fig. 1: Time-evolving position and speed features of sudden-stop attack dataset. Label 1 indicates an attack and label 0 indicates normal operation.



(a) Position feature (b) Speed feature

Fig. 2: Training results of the RL model where {A,B,C,D}={5,1,1,5} determine the reward function of Eq. 3.

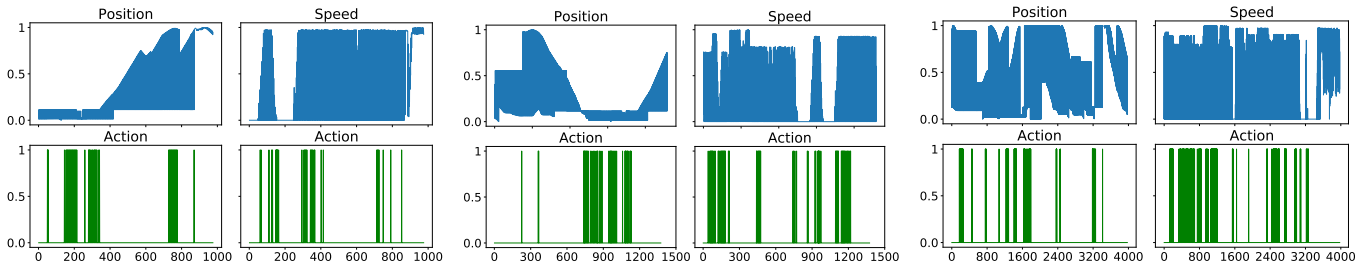
A. Datasets

Two datasets corresponding to traffic scenarios of high-density (37.03 vehicles/km^2) and low-density (16.36 vehicles/km^2) of vehicles are utilized as input in this work. The proportion between misbehaving and legitimate vehicles ranges from 30% to 70%, although such high proportion of attackers is rather unlikely in real-world scenarios. Exchanged beacon messages among vehicles include features such as position, speed, acceleration and heading angle, while each feature constitutes a three-dimensional vector[1]. Based on feature analysis and data pre-processing, position and speed were selected as the most relevant features related to the sudden-stop attack detection. In particular, the Euclidean norm of the speed vector and the x-dimension of the position vector are used. In our experiments, a subset of the high-density dataset was used to train the RL model with 2368 labelled attack messages from several misbehaving vehicles. On the other hand, a subset of the low-density dataset with 1846 labelled attack messages was used to test the capability of the RL-based approach in detecting sudden-stop attacks. Fig. 1 shows the time-evolving position and speed features of the two datasets with attack labels. The input datasets to the RL model are provided as samples, where each sample represents a chunk of the time-series dataset for the selected feature.

B. Results and Discussion

The detection performance of the RL model was evaluated based on commonly used metrics, i.e., *precision*, *recall* and

[1]It is noted that z-dimension entries are zero-valued for all features.



(a) Sample 1 (b) Sample 2 (c) Sample 3

Fig. 3: Test results of the RL model.

F1 score, which are defined as

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{5}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{6}$$

respectively. A higher F1 score indicates a better detection performance. Similarly, higher precision values indicate low FP rates and higher recall values indicate low FN rates.

In Fig. 2, training performance of the RL model is illustrated for position and speed features on a particular sample. The training performance of the model for the position feature (Fig. 2a) is 100%, which means that neither FPs nor FNs were reported. However, training performance of the model for speed feature is not as good as for position due to the occurrences of several FP (e.g., precision is 0.7692) and FN (e.g., recall is 0.8571) alarms (Fig. 2b). It can be observed that the speed values of some legitimate vehicles are close to zero due to stopped or slow-moving vehicles, which, in turn, resemble the sudden-stop attack; thus, this behaviour tends to mislead the RL model for an incorrect action in a state.

Regarding the RL model evaluation with the test dataset, we hereby present only the results for the samples with the best achieved performance due to space limitations. Fig. 3 shows the resulting performance for three sample sequences (i.e., sample 1, sample 2 and sample 3) of test data. The detection performance for each selected sample is shown in Table I. From the evaluation results, it can be noticed that the RL model performs better (except for sample 2) when using position instead of speed feature. For example, F1 score is very low in sample 3 when tested using speed feature. As mentioned before, close-to-zero speed values of some legitimate vehicles are causing the RL model to perform incorrect actions in detection; in turn, an increased number of FP (low precision) and FN (low recall) alarms are reported by the RL model. Similarly, it is noted that transmission of constant or close-to-constant positions by stopped/slow-moving legitimate vehicles can also affect the detection performance of the RL model. For example, low recall performance is reported for sample 2 when using the position feature, and this can be attributed to the increased number of FN alarms, by incorrectly identifying attackers as stopped/slow-moving legitimate vehicles.

TABLE I: Detection performance over the test dataset

Test sequence	Feature	Precision	Recall	F1
Sample 1	Position	1.0	0.9250	0.9610
	Speed	1.0	0.9091	0.9524
Sample 2	Position	0.9915	0.7347	0.8440
	Speed	0.9442	0.9979	0.9703
Sample 3	Position	0.8558	0.8017	0.8279
	Speed	0.6423	0.5360	0.5843

IV. Conclusion

In this paper, we have presented the evaluation of an RL-based approach for detecting sudden-stop attacks in V2X scenarios. Performance results confirm that misbehaving vehicles can effectively be detected with high accuracy by sequentially analysing their mobility patterns, i.e., real-time position and speed, using an RL model. In the path forward, we will investigate the integration of multi-dimensional vehicles' data into RL-based misbehaviour detectors with varying proportion of attackers.

Acknowledgment

This work is partly supported by the H2020-INSPIRE-5Gplus project (under grant agreement No. 871808), by the Spanish MINECO project SPOT5G (TEC2017-87456-P), and by the Generalitat de Catalunya under Grant 2017 SGR 891.

References

[1] Ye, H. et al., "Machine learning for vehicular networks: Recent advances and application examples," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 94–101, 2018.

[2] Maimó, L. F. et al., "Dynamic management of a deep learning-based anomaly detection system for 5G networks," *Journal of Ambient Intelligence and Humanized Computing*, 2019.

[3] Sakiz, F. et al., "A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV," *Ad Hoc Networks*, vol. 61, pp. 33 – 50, 2017.

[4] van der Heijden, R. W. et al., "Survey on misbehavior detection in cooperative intelligent transportation systems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 779–811, 2019.

[5] Huang, C. et al., "Towards experienced anomaly detector through reinforcement learning," in *AAAI*, vol. 32, no. 1, 2018.

[6] Kamel, J. et al., "Veremi extension: A dataset for comparable evaluation of misbehavior detection in vanets," in *ICC*, Jun 2020.

[7] Sutton, R. et al., *Reinforcement learning: An introduction*. MIT press, 2018.

[8] Watkins, C. et al., "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.