A Multi-Level Dempster-Shafer and Reinforcement Learning-Based Reputation System for Connected Vehicle Security

Pranay Chowdary Jasti
Department of Computer Science
Texas State University
San Marcos, Texas
xxp12@txstate.edu

Henry Griffith
Department of Engineering
San Antonio College
San Antonio, Texas
hgriffith5@alamo.edu

Heena Rathore
Department of Computer Science
Texas State University
San Marcos, Texas
heena.rathore@txstate.edu

2024 IEEE 21st Consumer Communications & Networking Conference (CCNC) | 979-8-3503-0457-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CCNC51664.2024.10454856

***Abstract*—Data falsification attack in connected vehicles (CV) refer to the manipulation or alteration of data within the vehicle's communication systems. This paper discusses the critical challenges in ensuring the security of CV networks where vehicle data integrity is paramount to prevent data falsification. Various existing solutions, such as machine learning and reputation-based approaches, have limitations in terms of scalability and robustness. To address these issues, we propose a novel multi-level Dempster-Shafer with reinforcement learning (RL)-based reputation system for CV networks. We use decentralized validation that combines self and peer reports of vehicles along with centralized feedback from road side unit, merging reputation-based trust management with Deep RL. By incorporating a multi-level Dempster-Shafer model, we elevate prediction accuracy and reward values while dynamic RL optimizes the process of reputation updates.**

I. Introduction

Connected vehicles (CV) are equipped with sensors to streamline navigation and facilitate easy transportation [1]. This is achieved by sharing basic security messages (BSM) which contains the crucial information of each vehicle such as latitude and longitudinal position, speed, acceleration and others [2]. Ensuring the security of this data is essential, serving as a safeguard for the integrity of the CV network and preventing malicious actors from tampering with it.

BSM data is manipulated by compromising the integrity of the data by injecting falsified information in it [3], [4], [5]. This type of falsification pose a considerable challenge in terms of detection and mitigation since the attackers possess a deeper knowledge of the system's internals. The state of art solutions for data falsification comprises of machine learning [6], and reputation based approaches [7], [8], [9] and others [10]. Machine learning approaches face challenges as they require large and diverse labelled datasets for the detection of data falsification. Further, reputation systems often rely on static and predefined trust scores based on historical data and may lack the ability to adapt well to the rapidly changing CV network.

Reinforcement learning (RL) based reputation systems offer the potential to improve trust and security in CV networks. In [11] the authors had discusses deep RL with the reputation system to build a dynamic reputation update policy. The model is characterized by a centralized deployment strategy, which, while commendable in its initial application, raises concerns about scalability and overall robustness. By predominantly focusing on vehicle feedbacks for reputation updates, the scheme does not incorporate self and peer report validation [2], thus limiting the comprehensiveness of its trust assessment framework. Moreover, the scope of the presented results is constrained within attacker densities ranging from 40% to 60%, omitting exploration of its effectiveness in more hostile scenarios that could significantly impact real-world deployment. Additionally, the reliance on the Simulation of Urban MObility (SUMO) traffic simulator [12] for simulations poses a potential limitation, as this platform may not fully encapsulate the intricacies and complexities inherent in RL framework. These identified gaps collectively underscore the need for an enhanced and comprehensive approach to trust and security within CV.

Our research addresses issues by introducing a hybrid deployment model that optimizes scalability and adaptability in CV. By integrating self and peer reports alongside traditional vehicle feedbacks, our framework enhances reputation updates and data falsification detection accuracy. Unlike previous studies, we comprehensively evaluate our scheme at an 80% attacker density, shedding light on its robustness in highly challenging scenarios. To address a critical gap, we define a sophisticated internal attacker model, enriching our threat analysis with a realistic representation of potential adversarial behavior. Furthermore, we enhance simulation realism by transitioning to a multi-agent RL controlled environment, ensuring practical relevance in complex CV networks.

The paper outline is organized as follows: Section II summarizes related work for security in CV towards data falsification attacks [13]. Section III presents proposed multi level system architecture, and attacker model. Section IV shows the simulation results. Finally, Section V concludes the paper with proposed future work.

II. Background and Related Work

A. Related Work

Feng et al. [14] explores diverse threat scenarios, attack strategies, and protective measures for BSM related traffic safety and control systems. The authors propose a security analysis framework, encompassing risk assessment, defense solutions, security testing, and utilizing data from the hybrid Mcity model with cross-validation, proof of validation, and trajectory-based hierarchical defense. The threat model outlined in their paper centers around a central management system, which, as a single point of failure, poses significant reliability and scalability challenges.

In [15], the authors present artificial intelligence and statistical data classification framework to analyze messages in CV. The model is trained on the US Department of Transportation Safety Pilot Deployment Model, which integrates a ML algorithm and a local trust manager. Experimental results show that the trained model can accurately predict false alerts, it attains an 98% accuracy rate while maintaining a relatively low 0.55% standard deviation when subjected to 25% malicious data. The model's performance heavily rely on the specific characteristics of the data it was trained on, and its ability to handle novel, unseen data remains unaddressed.

Chen et al. [10] focuses on a trust-based service management mechanism to secure information dissemination, emphasizing the need for a Decentralized Trust Management System (DTMS). The authors propose a Blockchain-based DTMS and evaluate its performance using data from a test bed and comparing it to a blockchain-based non trust evaluation scheme. The findings show that the DTMS exhibits an efficient consensus design, with high throughput and low latency, making it suitable for large-scale transportation environments.

Suo [24] draws inspiration from Zacharia's work on reputation systems [25]. The authors assess the effectiveness of both centralized and distributed architectures, as described in reference [26]. In the centralized approach the trust authority underwent a plausibility assessment vs in the decentralized approach, the plausibility assessment was done by the peer vehicle. Though the model had learning rate and forgetting rate as the parameter, rule-based reputation systems are characterized by their static nature, and lack the ability to adapt the rules over time.

The authors in [11] develop a dynamic reputation update policy using deep RL and DS theory for feedback combination. The simulation involves a scenario with normal and malicious vehicles, and the results demonstrate the effectiveness of the proposed scheme in predicting true messages accurately. Overall, the paper highlights the use of dynamic reputation policy as a collaborative misbehavior detection system in 5G-based CV networks [11]. However, it also gives rise to concerns regarding scalability and overall robustness. The primary focus on vehicle feedback for reputation updates neglects the inclusion of self and peer report validation, as highlighted in reference [2]. The study's reported results are confined to scenarios with attacker densities ranging from 40% to 60%, excluding an exploration of its effectiveness in more hostile environments that could substantially impact real-world deployment. Another limitation is the absence of a detailed examination of attacker models, leading to a deficiency in providing a comprehensive threat analysis. This gap arises due to the lack of a well-defined attacker profile. Additionally, the reliance on the SUMO traffic simulator [12] for simulations where the platform did not fully capture the RL framework. In this paper, we address these issues by providing a comprehensive approach to trust and security in CV networks. Following are the contributions:

- Present a multi level hierarchical structure for evidence and hypothesis.
- Implementation of multi level DS technique to handle uncertainties at Road Side Unit (RSU). In many real-world situations, sources of information may provide contradictory data. Multi level DST provides a systematic way to combine and manage these conflicting pieces of evidence, allowing RSU to weigh the importance of different sources.
- Introduce well defined attacker profile with the simulations performed in RL framework dedicated for vehicle navigation.

III. Proposed Work

A. Dataset for Vehicle Networks Simulation

We used an open source vehicle navigation model controlled by multi agent RL agent framework [22], [23] where the navigation of the vehicles is governed in a on ramp merging scenario simulating general traffic conditions. Each vehicle shares BSM which contains the details about its position, speed, headed direction, acceleration. At vehicle level we implemented multi level system architecture, where each vehicle calculates its peer reputation by comparing the BSM reported by the vehicle with the self reports generated through its own sensor. We assume that each vehicle is equipped with necessary sensors which help them to make peer reports.

B. System Model

The system architecture (Figure 1) combines decentralized validation with centralized feedback, integrating reputation-based trust management with a Deep Q-learning agent at the Centralized Authority (CA). A multi-level implementation of DS at the RSU predicts false reports from vehicles, influencing the Deep RL agent's rewards. It starts with inter-platoon communication, where vehicles transmit BSM, sensed by peer vehicles. Decentralised reputation calculation at the vehicle level identifies malicious vehicles, generating reputation scores sent to the RSU, with dynamic reputation updates (smoothing factor) assigned by the CA. The RSU employs DS theory to combine vehicle reputation reports, considering both peer-reported behavior and self-reported veracity. An average reputation score, reputation update policy, and previous-time rewards feed into the RL agent, which determines an optimal smoothing factor disseminated across the network for decentralized vehicle-level reputation calculation. This factor

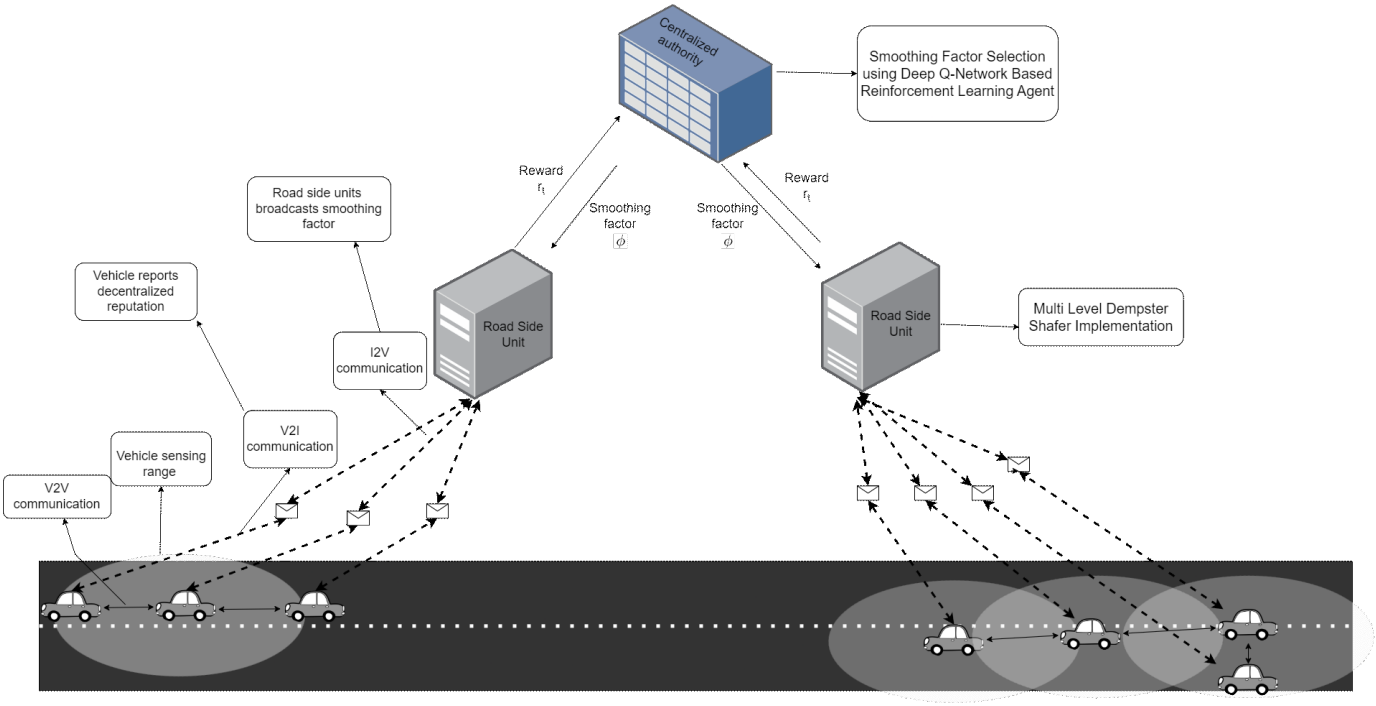


Fig. 1: System Model

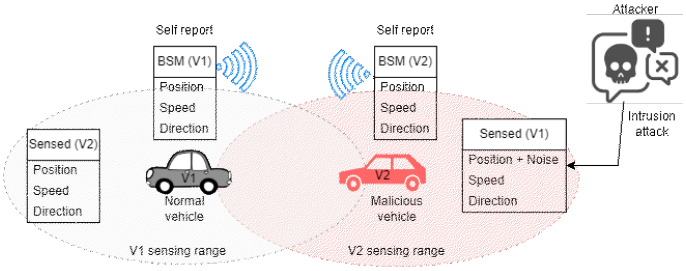


Fig. 2: Attacker Model

encourages malicious vehicles to report truthfully, while a robust reputation update policy continuously evaluates each vehicle's overall behavior.

C. Attacker model

In our model we implement data falsification attack by adding a white Gaussian noise to the sensed kinematics of the peer vehicles as shown in Figure 2. The white Gaussian is generated with $\mu = 2$ and $\sigma = 0.9$. The noise is introduced by intercepting the smoothing function $(\phi_1) \geqq 0.4$.

D. Dempster-Shafer Technique

The DS theory of evidence is a mathematical framework for quantifying belief in statements by combining independent evidence from multiple sources using belief functions. Unlike traditional probability theory, it handles uncertainty by assigning degrees of belief to subsets of possible events. It assumes inherent ignorance leading to uncertainty and uses the DS rule to aggregate belief functions.

a) Frame of discernment Θ*:* Frame of discernment is defined as $\Theta = \{\Theta_1, \Theta_2 \ldots\ldots \Theta_n\}$ that covers individual, mutually exclusive, discretized values of all viable outcomes of Θ. In our approach Θ contains two elements $\Theta = \{M, N\}$; where M= malicious vehicle, N= normal vehicle.

b) Power Set P(Θ*):* The power set P(Θ) of aforesaid random variable Θ is a set of all subsets of Θ including the individual elements, represents the DS frame of Θ. For our model it contains P(Θ)=$\{\phi, M, N, MN\}$.

c) Evidence: Evidences are events/symptoms and one evidence maps to single hypothesis or set of hypotheses. We consider the reputation score from the vehicles as the evidences in the level 1. In level 2, the difference between the plausibility values from level 1 and the reputation reports from the vehicles are considered as evidences. For level 3, plausibility scores from level 1, level 2 and reputation calculated by RSU is considered as a evidence.

d) Mass Function (m-value): Our trust regarding the truth value of a proposition 'A' is dependent on the evidence that supports the proposition which is denoted as mass function(m-value). It relates to the weights of the elements in the P(Θ), $m : 2^{\Theta} \rightarrow [0, 1]$, where weight of the null set is 0, $m(\phi) = 0$ and $m(A) \geq 0$. The overall sum of the mass function's of all elements in the P(Θ)=1 or $\sum\{m(A) \; \forall \; A \in 2^{\Theta}\} = 1$. Thus, $m(A)$ is a measure of belief assigned by a given evidence to A, where A is any element of 2^{Θ}, $\forall A \in 2^{\Theta}$, and non belief is forced by the lack of knowledge. We can get the lower and upper bound of an interval from the mass function. The lower bound is used as the belief function and the upper bound is used to calculate

the plausibility function.

e) Plausibility function (Pl): The upper bound of the interval is called as plausibility, and it is determined by taking the sum of all the mass function of the subsets (B) that intersects (A) where $(B \cap A \neq \phi)$, $Pl(A) : 2^{\Theta} \rightarrow [0, 1]$ [13].

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) \tag{1}$$

f) DS Rule of Combination: The data collected from the different sources are combined rationally, to focus on the consensus opinion and use normalization to ignore all the conflicting evidence. A cartesian product of two mass functions is employed for the combination of evidence. The DS combination rule determines the joint m_{1-2} from the combination of two mass function using equation:

$$m_{1-2}(A) = \frac{\sum_{B \cap C = A}\{m_1(B)m_2(C)\}}{1 - K} \tag{2}$$

when $A \neq \phi$, $m(\phi) = 0$ and $K = \sum_{B \cap C = \phi} m_1(B)m_2(C)$

E. Multi Level System Architecture

1) Level-1 Plausibility Calculation: Level 1 corresponds to decentralized reputation calculation at the vehicle level. Each vehicle is equipped with the capability to sense the kinematics of its peers. Each vehicle calculates the difference between the sensed kinematics and the broadcasted BSMs of peer vehicle (j). If this disparity exceeds a predefined threshold noise, the vehicle (i) assigns a trust value of 0.1 to j; otherwise, it assigns a trust value of 0.9.

$$T^i_{j,t} = \begin{cases} 0.1 & If \quad \Delta > 0.2 \\ 0.9 & Else \end{cases} \tag{3}$$

where $\Delta = x^j_j - x^i_j$ and x is the position reported. These trust values are then integrated with the current reputation score through the dynamic reputation update policy with dynamic smoothing factor (denoted as ϕ_1):

$$R^i_{j,d_t} = \phi_1 R^i_{j,d_{t-1}} + (1 - \phi_1) T^i_{j,t} \tag{4}$$

Each vehicle then share the calculated reputation scores to RSU. The reputation scores shared by peer vehicles serve as the basis for building mass functions. The RSU treats the reputation reports provided by different vehicles as individual pieces of evidence. Specifically, one mass function characterizes the likelihood of a vehicle being normal (N), while the other, its complement (1 - reputation score), represents the likelihood of a vehicle being malicious (M). For each vehicle's reputation report, the RSU generates a mass function tuple consisting of $m(N)$ (the mass function for normal behavior), $m(M)$ (the mass function for malicious behavior):

$$m_{1i}(N)^{RSU} = R^j_{i,d_t} \tag{5}$$

$$m_{1i}(M)^{RSU} = 1 - R^j_{i,d_t} \tag{6}$$

The generated mass functions from the reputation reports shared by the peer vehicles are combined using Eq. 2 which is later used to calculate the plausibility values of vehicles being malicious and normal. The plausibility values generated at this stage are referred to as level-1 plausibility values where $Pl_{1,j}(M)$ and $Pl_{1,j}(N)$ represents the level-1 plausibility of the vehicle being malicious and normal respectively.

2) Level-2 Plausibility Calculation: The plausibility calculation described in the previous step solely relies on reputation reports from peer vehicles. However, in real world scenarios attackers can exploit a model by deliberately sending false, low reputation reports of peers to the RSU. To address this challenge, in addition to aggregating peer reputation reports, it is essential to validate the accuracy of the reputation reports provided by each vehicle about their peer vehicles. In this step, RSUs generates mass functions by calculating the difference between the reputation report submitted by the vehicle i for peer vehicle j and the level-1 plausibility value of the peer vehicle j being normal indicating the vehicle's malicious behavior.

$$m_{2i}(M)^{RSU} = \begin{cases} |R^i_{j,d_t} - Pl_j(N)^{RSU}| & If \quad |R^i_{j,d_t} - Pl_j(N)^{RSU}| > 0.2 \\ 0.1 & Else \end{cases} \tag{7}$$

$$m_{2i}(N)^{RSU} = 1 - m_{2i}(M) \tag{8}$$

This validation process is iteratively applied to each report submitted by a vehicle for its peer vehicles. The plausibility values from this step are considered as the level-2 plausibility values where $Pl_{2,i}(M)$ and $Pl_{2,i}(N)$ represents the level-2 plausibility of the vehicle being malicious and normal respectively.

3) Random Validation by RSU:

When the proportion of malicious vehicles surpasses 50%, attackers succeed in their objective of diminishing the reputation of peer vehicles at the CA. To counteract this vulnerability, RSUs intermittently intercept vehicle communications in every 20 discrete intervals and maintains this policy for next 20 iterations. RSUs validate vehicle self-reports against its own sensed kinematic data to establish vehicle reputations.

$$R^{RSU}_{i,c_t} = \phi_2 R^{RSU}_{i,c_{t-1}} + (1 - \phi_2) T^{RSU}_{i,c_t} \tag{9}$$

$\phi_2 = 0.2$. RSUs accumulates reputation reports that vehicles share regarding their peers. The acceptance of reputation reports despite variance is grounded in the disparate nature of trust update policies employed by vehicles and RSUs. When the variance follows a descending order, a vehicle is attributed a trust level of 0.9; otherwise, a trust level of 0.1 is assigned:

$$T^{RSU}_{i,c_t} = \begin{cases} 0.1 & If \quad \Delta > 0.2 \\ 0.9 & Else \end{cases} \tag{10}$$

$\Delta = x^{RSU}_i - x^i_i$. Later using Eq. (5) and (6) the centralized mass functions $(m_{3,i}(M))$ and $(m_{3,i}(N))$ are calculated.

F. Level-3 Combination of Level-1, Level-2 and Centralized Validated Scores

RSU merges information from level 1, level 2 and RSU calculated reputation to predict vehicle behavior. Employing

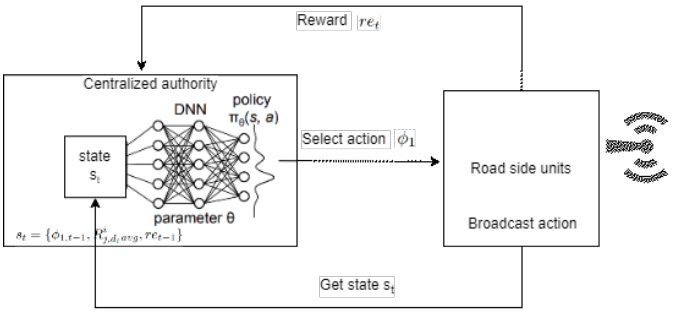


Fig. 3: Centralized Reinforcement Learning Model

the DS theory of combination, RSU fuses these mass functions to determine the final plausibility of a vehicle being malicious and normal using Eq. 1. When the final plausibility of a vehicle's malicious behavior exceeds 0.3, the vehicle is identified as malicious, rendering all its transmitted reputation reports as malicious. This validation procedure extends to all vehicles within the network. These predictions, serving as the foundation, contribute to the formulation of rewards for the RL agent located at CA. This multi-stage process not only bolsters the integrity of the vehicular network against malicious behavior but also showcases the integration of RL to incentivize reliable reporting. The final plausibility of vehicle's malicious behavior is tracked using a robust trust update policy (Eq. 11) as mentioned in [27], for the over all behavior history and this is served as the malicious score of the vehicle which is used as a indicator for vehicles behavior at the RSU level. $M_{i,t}^{RSU}$ is the cumulative malicious score, $Pl_{i,t}^{RSU}$ is the final plausibility score of the vehicle being malicious and the values of D, θ and σ are set as 200, 2, 20 respectively.

$$M_{i,t}^{RSU} = M_{i,t-1}^{RSU} + \frac{1}{\theta}\Phi(M_{i,t-1}^{RSU})D(Pl_{i,t}^{RSU} - E_{i,t-1}^{RSU}) \quad (11)$$

$$\Phi(M_{i,t-1}^{RSU}) = 1 - \frac{1}{1 + \exp\left(\frac{-(M_{i,t-1}^{RSU} - D)}{\sigma}\right)} \quad (12)$$

$$E_{i,t-1}^{RSU} = (M_{i,t-1}^{RSU})/D \quad (13)$$

G. Deep-Q RL model at CA

In our model we have tailored the Q-learning RL agent to select the optimal smoothing factor. The state s_t contains previous smoothing factor ϕ_1, average reputation of the vehicles and reward (See Figure 3).

$$s_t = \phi_{1,t-1}, R_{avg}^{RSU}, re_{t-1} \quad (14)$$

The RSU calculates the average reputation as described in [11]. The reward is computed as the ratio of the number of true reputation reports to the total number of reputation reports. The agent after observing the current state will select the optimal smoothing factor.

$$a_t = \phi_1 = \{\phi_{1_1}, \phi_{1_2}, \phi_{1_3}, ..., \phi_{1_n}\} \quad (15)$$

Subsequently, this reward is furnished to the RL agent located in CA, which leverages it to make informed selections of actions, represented as smoothing factors. The overarching aim of these actions is to maximize the cumulative reward. In a comprehensive loop, the actions determined by the RL agent are disseminated to the vehicles, effectively compelling them to transmit accurate reports. For our work, the deep Q-network is a full connected network with two hidden layers.

IV. Simulations and Performance Results

A. Simulation Settings

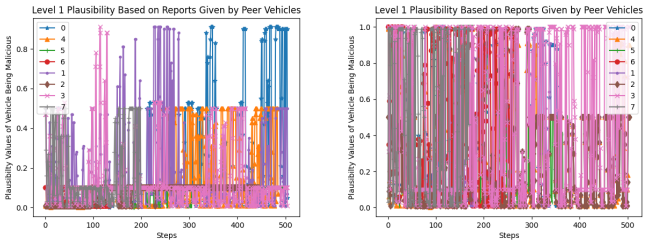
The simulations are carried out on a Lambda GPU workstation AMD(R) Ryzen threadripper pro 3955wx 16 cores x32 with 128 GB RAM on a Ubuntu 20.04.5 LTS. The dynamics of the vehicles is controlled by a Multi agent RL based navigation model which is designed for the navigation of the vehicles in a on ramp merging scenario, Every vehicles will broadcast the BSM's periodically and vehicles can also sense over a predefined space called as sensing range with the help of the onboard sensors, using which the vehicles will validate the BSM of other vehicles and build a reputation score using a dynamic reputation update policy and sends reputation report to the roadside unit, multi level DS model along with a centralized reputation calculation is carried out in the road side unit to predict the false reports from the vehicles, later the number of true reports out of total number of reports is sent to the deep Q RL agent model as a part of the state to find the optimal smoothing factor residing in the CA. The Deep Q Network is a fully connected neural network which contains one input layer, one hidden layer and one output layer. The RL agent select the random value for smoothing factor to explore the network for 300 episodes and starts selecting optimal smoothing factor as a part of exploitation.

B. Performance analysis

The performance of the model is showcased in two scenarios one with 20% attacker density and other with 80% attacker density. The rest of the section contains plots of plausibility scores of vehicle being malicious from level 1 reputation report combination, level 2 reputation reports calculation, reputation scores from the RSU and final plausibility scores of the vehicles. A subset of the vehicles are selected for the better representation. In this subset there are 8 vehicles names range in [0,7] in the network, In a 20% attacker scenario vehicles 2, 5 are malicious. In a 80% attacker density vehicle 3,6 are normal vehicles.

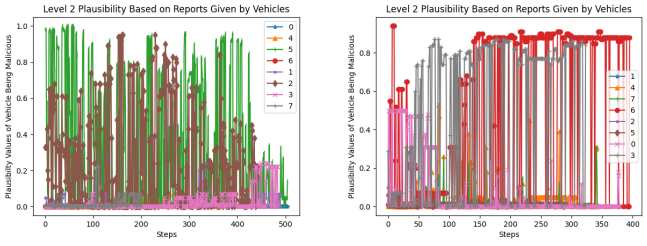
1) Level 1 malicious plausibility: Figures 4 (a) and (b) represents the level 1 plausibility scores for 20% and 80% attacker density respectively. It can be seen that the malicious vehicles were successful in framing the normal vehicles as malicious vehicles at the RSU.

2) Level 2 malicious plausibility: Figures 5 (a) and (b) are generated using the plausibility of vehicle being malicious which are calculated in the level 2 peer report verification process for 20% and 80% attacker density respectively.



(a) (b)

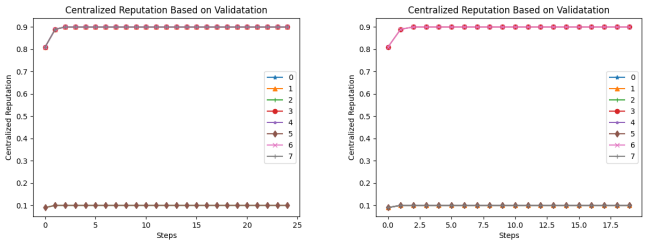
Fig. 4: (a) 20% attacker density where vehicles 2,5 are malicious, (b) 80% attacker density where vehicles 3,6 are normal



(a) (b)

Fig. 5: (a) 20% attacker density where vehicles 2,5 are malicious, (b) 80% attacker density where vehicles 3,6 are normal

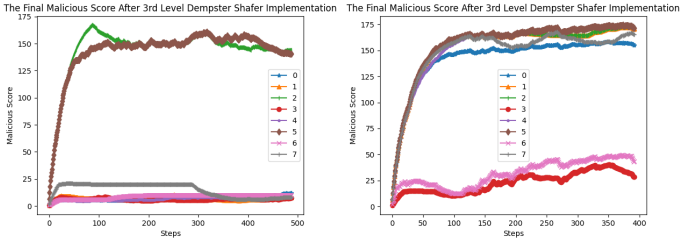
3) Reputation values calculated by RSU: Figures 6 (a) and (b) presents the values of vehicles centralized reputation in both 20% and 80% attacker density scenario. We can see that the RSU was able to identify the malicious vehicles behavior in every cycle.



(a) (b)

Fig. 6: (a) 20% attacker density where vehicles 2,5 are malicious, (b) 80% attacker density where vehicles 3,6 are normal

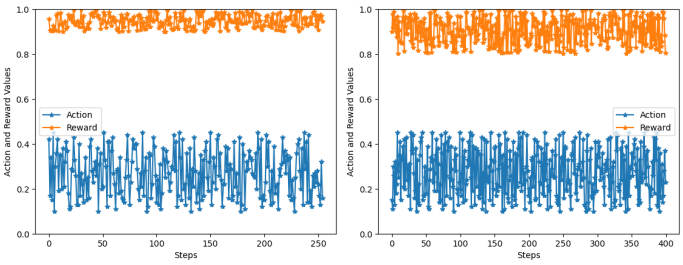
4) Final malicious score of vehicles: It can be seen from the final malicious score (See Figure 7) generated using Eq. 11, that our hybrid malicious detection model is able to identify the malicious vehicles in both 20% and 80% attacker density scenarios starting from 20 iterations.



(a) (b)

Fig. 7: (a) 20% attacker density where vehicles 2,5 are malicious, (b) 80% attacker density where vehicles 3,6 are normal

5) Action and reward values: Figures 8 (a) and (b) show the action and reward values of RL agent in 20% attacker density and 80% attacker density respectively. It is evident from Figure 8 that the RL agent performs well in both scenarios. The reward values in the 20% attacker density ranges between 0.87 - 0.99, where as in the 80% attacker density the reward ranges between 0.80 - 0.99. In both the cases, the RL agent, learns the smoothing factor to be in the range of 0.1-0.4 suggesting higher weights to trust values rather than reputation accumulated.

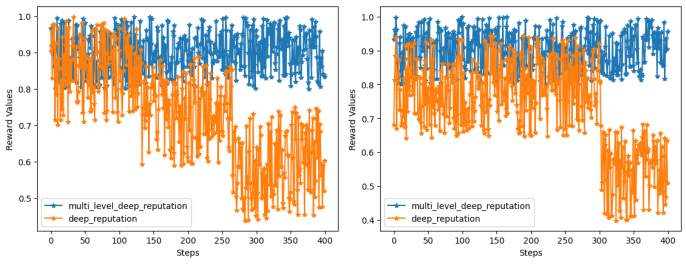


(a) (b)

Fig. 8: Action and reward values for: (a) 20% attacker density where vehicles 2, 5 are malicious, (b) 80% attacker density where vehicles 3, 6 are normal

C. Comparative Analysis

We have compared our model with deep RL algorithm proposed in [11]. The superiority of the our multi level DS along with deep RL over standard deep RL reputation method [11] in the 20% and 80% attacker density can be seen from Figure 9. Our proposed work reputation reward is high and ranges between 0.8-0.95 where as in the case of deep RL algorithm, reputation varies between 0.5 - 0.95. It is also observed that the reward value decreases after 100 steps in case of 20% attacker density and after 300 steps in 80% attacker density.



(a) (b)

Fig. 9: (a) 20% attacker density where vehicles 2, 5 are malicious, (b) 80% attacker density where vehicles 3, 6 are normal

V. Conclusion

We propose a three level DS model complemented by a RL-based dynamic reputation system. By amalgamating plausibility values from levels 1 and 2, as well as mass functions generated through centralized reputation calculations, the RSU consistently and precisely identifies malicious vehicles at each step. This precise detection enables the calculation of accurate reward values. These rewards are crucial for the RL agent to determine the optimal smoothing factor based on the current state. In both 20% and 80% attacker density scenarios, the RL agent has learned to select ϕ_1 values below 0.4 to maximize rewards at each step. In the 20% attacker density scenario, the RL agent consistently achieves high rewards ranging from 0.87 to 0.99. In the 80% attacker density scenario, reward values range between 0.80 and 0.99. In the future work, we plan to introduce other attacker models and implement different RL algorithms such as actor-critic models, and policy gradient RL model.

VI. Acknowledgment

This publication has been supported by NSF CISE Research Initiation Initiative (CRII) grant #2153510 and #2313351.

References

[1] S.A. Bagloee et al., "Autonomous vehicles: challenges, opportunities, and future implications for transportation policies," *J. Mod. Transport*. vol. 24, pp. 284–303, 2016.

[2] H. Griffith, M. Farooq and H. Rathore, "A Data Generation Workflow for Consensus-Based Connected Vehicle Security," *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, NV, USA, 2023, pp. 1-2, 2023.

[3] H. Rathore, S. Sai and A. Gundewar, "Social Psychology Inspired Distributed Ledger Technique for Anomaly Detection in Connected Vehicles," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7092-7107, 2023.

[4] M. Castillo et al., "Poster: Decentralized Simulation Workflow for Enhancing Connected Vehicle Security". *In Proceedings of the Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23). Association for Computing Machinery, New York, NY, USA*, pp. 574–576. 2023.

[5] G. Voce et al., "Poster: Opinion Dynamics for Enhancing Trust and Security in Connected Vehicle Networks". *In Proceedings of the Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '23). Association for Computing Machinery, New York, NY, USA*, pp. 562–564. 2023.

[6] Anyanwu et al, "Novel hyper-tuned ensemble random forest algorithm for the detection of false basic safety messages in internet of vehicles," in *ICT Express*, 9(1), pp.122-129, 2023.

[7] H. Rathore, A. Samant, and M. Jadliwala. 2021. "TangleCV: A Distributed Ledger Technique for Secure Message Sharing in Connected Vehicles," in *ACM Trans. Cyber-Phys. Syst.* vol 5, Article 6, pp 1–25, 2021.

[8] H. Rathore and H. Griffith, "Leveraging Neuro-Inspired Reinforcement Learning for Secure Reputation-based Communication in Connected Vehicles," *2023 IEEE Conference on Communications and Network Security (CNS)*, Orlando, FL, USA, pp. 1-6, 2023.

[9] H. Rathore and H. Griffith, "GNN-RL: Dynamic Reward Mechanism for Connected Vehicle Security using Graph Neural Networks and Reinforcement Learning," *2023 IEEE International Conference on Smart Computing (SMARTCOMP)*, Nashville, TN, USA, pp. 201-203, 2023.

[10] X. Chen, J. Ding and Z. Lu, "A Decentralized Trust Management System for Intelligent Transportation Environments," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 558-571, Jan. 2022.

[11] S. Gyawali, Y. Qian and R. Q. Hu, "Deep Reinforcement Learning Based Dynamic Reputation Policy in 5G Based Vehicular Communication Networks," in *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 6136-6146, 2021.

[12] Lopez et al, "Microscopic traffic simulation using sumo," in *21st international conference on intelligent transportation systems (ITSC)* pp. 2575-2582, 2018.

[13] Sentz, K. and Ferson, "Combination of evidence in Dempster-Shafer theory". 2002

[14] Y. Feng et al, "On the Cybersecurity of Traffic Signal Control System With Connected Vehicles," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16267-16279, 2022.

[15] E.O. Eze et al.,"A Context-Based Decision-Making Trust Scheme for Malicious Detection in Connected and Autonomous Vehicles". in *Proc. IEEE2022 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, pp. 31-36, 2022.

[16] G. -P. Antonio and C. Maria-Dolores, "Multi-Agent Deep Reinforcement Learning to Manage Connected Autonomous Vehicles at Tomorrow's Intersections," in *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7033-7043, 2022.

[17] Chen et al, "Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles," in *Comput Aided Civ Inf.* vol 36, pp 838– 857, 2021.

[18] A. Qayyum et al, "Securing Connected and Autonomous Vehicles: Challenges Posed by Adversarial Machine Learning and the Way Forward," in *IEEE Communications Surveys and Tutorials*, vol. 22, no. 2, pp. 998-1026, 2020.

[19] N. Papernot et al, "The Limitations of Deep Learning in Adversarial Settings," in *IEEE European Symposium on Security and Privacy (EuroS and P)*, pp. 372-387, 2016.

[20] Chen, T., Liu, J., Xiang, Y. et al. "Adversarial attack and defense in reinforcement learning-from AI security view," in *Cybersecurity*, 2019.

[21] A. BOUBAKRI and S. METTALI GAMMAR, "Intra-Platoon Communication in Autonomous Vehicle: A survey," in *9th IFIP International Conference on Performance Evaluation and Modeling in Wireless Networks (PEMWN)*, pp. 1-6, 2020.

[22] Dong Chen et al, "Deep Multi-agent Reinforcement Learning for Highway On-Ramp Merging in Mixed Traffic," 2022.

[23] https://github.com/DongChen06/MARL_CAVs [accessed on august 31, 2023].

[24] D. Suo, "Towards security by design of connected and automated vehicles: cyber and physical threats, mitigations, and architectures" in *(Doctoral dissertation, Massachusetts Institute of Technology)*, 2021.

[25] G. Zacharia and P. Maes. "Trust management through reputation mechanisms". *Applied Artificial Intelligence*, 14(9):881–907, 2000.

[26] A. Angelogianni, I. Krontiris and T. Giannetsos, "Comparative Evaluation of PKI and DAA-based Architectures for V2X Communication Security," *2023 IEEE Vehicular Networking Conference (VNC)*, pp. 199-206, 2023.

[27] D. Suo and S. E. Sarma, "Proof-of-Travel: A Protocol for Trustworthy V2I Communication and Incentive Designs," in *IEEE Vehicular Networking Conference (VNC)*, pp. 1-4, 2020.