# Distinguishing Sensor Faults and Attacks in Connected Vehicle Networks Using Two-Step Verification

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

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

2023 Eighth International Conference On Mobile And Secure Services (MobiSecServ) | 979-8-3503-1649-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/MOBISECSERV58080.2023.10329092

***Abstract*—As the reliance on vehicle-to-vehicle communication within transportation networks increases, the risk of cyber attacks is an escalating concern. In addition to attacks, faulty sensors within vehicles can also lead to transportation network failures. To address these challenges, this research proposes a two-step process for identifying and classifying malicious and faulty vehicles within a network using sequential reputation estimation and validation. The reputation estimation step calculates the cumulative trustworthiness of vehicles by validating their broadcasted Basic Safety Messages (BSMs) using the sensors of neighboring vehicles, while the reputation verification step classifies low reputation vehicles as faulty or attacked by assigning them additional tasks and analyzing their responses. Simulation results demonstrate the effectiveness of the proposed approach in accurately detecting and distinguishing faulty and malicious vehicles.**

***Index Terms*—connected vehicles, reputation, two step verification, trust**

## I. Introduction

As vehicles become more connected and automated, they rely on accurate and trustworthy information exchange to ensure safe and efficient transportation [1]. This reliance on shared information creates natural vulnerabilities to cyber attacks. Cyber attacks in connected vehicles (CV) can lead to unauthorized access to critical vehicle systems, manipulation of sensor data, and interference with navigation and control systems. While compromising CV is possible through various attack vectors, including conventional communication attacks like denial of service, the specific targeting of information integrity reported by vehicles presents significant challenges [2].

Information integrity attacks involve the falsification of data exchanged within the network. These attacks can rapidly undermine the integrity of CV networks, potentially resulting in severe consequences. The presence of faulty sensors in vehicles can also introduce erroneous data, leading to similar incorrect decision-making and system failures to data falsification. Ensuring robust security measures in CV to distinguish between faults and cyber attacks is vital to safeguarding the integrity, reliability, and safety of the transportation system.

Trust and reputation systems play a crucial role in maintaining reliable network operation [3]. By evaluating a vehicle's reputation, other vehicles in the network can make informed decisions regarding the trustworthiness of the information and messages transmitted by that particular vehicle. These systems are enabled by using the sensors of neighboring vehicles to estimate the reliability of information broadcasted by vehicles. The reputation metric serves as a measure of the vehicle's reliability, integrity, and past behavior, enabling others to assess the credibility of its data and actions.

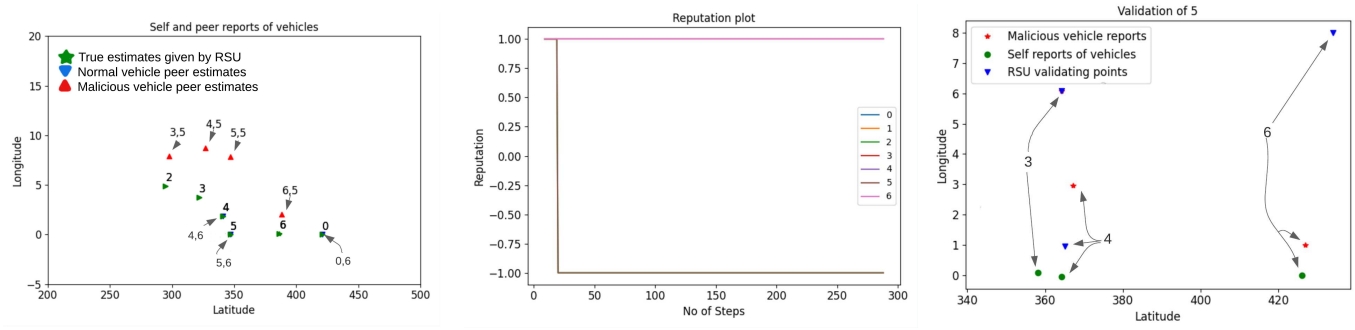
While reputation systems can provide an indication of trustworthiness, they may confuse data falsifications resulting from faults and cyber attacks. This paper addresses this issue by introducing an additional step for reputation verification. Here, the road side unit (RSU) can actively challenge suspected malicious vehicles and validate the accuracy of their messages. This is done by assigning specific tasks to suspected malicious vehicles and analyzing their responses. The RSU can thus differentiate between sensor faults (resulting in inaccurate data) and deliberate attacks (resulting in intentionally falsified data). This differentiation can enable tailored mitigation strategies and appropriate responses based on the specific nature of the issue.

## II. Methods

To begin the reputation estimation process, the trustworthiness of vehicles is estimated at each timestamp by comparing the BSM data broadcasted by each vehicle versus the peer estimates of this data produced by neighboring vehicles. These trust estimates are defined as a matrix $T_{n\times n}$, where $n$ is the number of vehicles in the network. Each element $T_{i,j}$ corresponds to the estimation of trust of vehicle $i$ by vehicle $j$. If vehicle pairs are not within range of one another, elements of the trust matrix are not defined (e.g.: $nan$). Trust estimates are then utilized to develop a cumulative reputation estimate.

### A. Reputation Estimation

Reputation estimation is performed by the BSM based upon the evaluation of the peer estimation accuracy of each vehicle.




(a) (b) (c)

Fig. 1: (a) Self and peer reports of malicious (#5) and benevolent vehicle (#6). Here peer reports are denoted by $(i, j)$ stating report of $i$ by vehicle $j$, (b) Reputation scores of malicious and benevolent vehicles, (c) RSU BSM data points given to (#5) for (#3, #4, and #6)

For the initial simulations described herein, we assume that all vehicles broadcast correct BSM data in an attempt to maintain high reputation scores and stay within the network. The RSU then forms an aggregated measure of the trustworthiness of each vehicle by assessing the accuracy of their peer-estimates. Namely, if any peer-estimate produced by the vehicle varies from the self-reported values (assumed to be ground truth based upon the aforementioned assumptions) outside of a threshold (initially set at -0.1), the RSU assigns a trust value $T(t)$ of -1 (malicious vehicle); otherwise, a trust value of 1 (benevolent vehicle) is assigned. In every iteration, we update the reputation of the vehicle given by:

$$R_t = wT_t + (1 - w)R_{t-1} \tag{1}$$

Where $R(t)$ is the reputation (cumulative estimate of trust) and $w \in (0, 1)$ is a weighting factor (set at 0.5 herein) which serves to modulate the rate at which a vehicle's reputation is updated.

### B. Reputation Verification

In this step, the RSU assesses the reputation of vehicles to formulate a list of potentially malicious vehicles by thresholding reputation scores. For each vehicle in the potentially malicious list, the RSU selects benevolent vehicles (e.g.: those not on the list) in close proximity to the malicious vehicle and alters a subset of their BSM data randomly. The BSM data of the benevolent vehicles is then sent to the potentially malicious vehicles for verification. These vehicles then verify the BSM data sent from the RSU and returns whether the data is true or false. We assume that malicious vehicles will correctly report each position in an attempt to stay in the network, while faulty vehicles will maintain incorrect reports. The RSU classifies whether vehicles are attacked or faulty using the malevolence factor $(M_f)$:

$$M_f = True\_verifications/A_i \tag{2}$$

If $M_f$ is less than 1, the vehicle is classified as faulty; otherwise, it is classified as attacked.

### C. Results

Simulations were carried out in a Python environment using OpenAI Gym and Highway-ENV for vehicle generation and simulation. Vehicle navigation was governed by the multi-agent reinforcement learning (RL) algorithm designed for CV [4]. The code has been developed on the existing code base available at [5]. We utilized an 8-vehicle grid. After 20 simulation steps, vehicles 2 and 5 begin reporting inaccurate measurements of other vehicles. Figure 1(a) shows the self and peer reports of vehicles. Figure 1(b) shows the reputation scores of benevolent vehicles and malicious vehicles (closer to +1 and -1, respectively). Vehicles are added to the malicious list and are given additional task as discussed in Section II (B). Here, vehicle number 5 is flagged as faulty since $M_f$ is computed as 1/3 (Figure 1(c)).

## III. Future Works

Future work will improve the sophistication of the reputation model utilizing RL. RL algorithms for reputation will be evaluated using additional relevant metrics like mean episode reward, sample efficiency, and learning curves.

## IV. Acknowledgment

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

## References

[1] 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.

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

[3] H. Rathore and H. Griffith, "GNN-RL: Dynamic Reward Mechanism for Connected Vehicle Security using Graph Neural Networks and Reinforcement Learning", *IEEE SmartComp*, 2023.

[4] 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.

[5] https://github.com/DongChen06/MARL_CAVs [accessed on June 3, 2023].