# Using a Synthetic Data Generation Workflow to Introduce Faults and Attacks in Connected Vehicles

**Marbella Castillo**
Mentors: Dr. Henry Griffith and Dr. Heena Rathore
Department of Computer Science, Texas State University, USA

## Introduction

- Connected vehicles (CV) are equipped with advanced sensors that facilitate seamless communicate with both their internal systems and the surrounding environments.
- These vehicles can exchange their kinematic data (i.e., speed, direction, location etc.) through basic safety messages (BSM).
- While this real- time communication between these interconnected vehicles is essential for road safety, it also exposes them to potential cyber-attacks.
- To address these cybersecurity concerns, this research focuses on simulating various faults and attacks using a reliable source dataset[1].

## Background

**Goal:** Addressing limitations in existing BSM datasets for CV networks by incorporating diverse simulated noise models and establishing a baseline for future security algorithm development and evaluation.

**Decentralized Security**
The exchanging of real-time BSM data between vehicles and infrastructure can expose vulnerabilities, making CVs potential targets for cyber-attacks. Traditional centralized security approaches may prove inadequate in protecting these interconnected systems, especially as cyber threats continue to evolve. Decentralized approaches to CV security can significantly enhance resilience against cyber threats by distributing security responsibilities across the network. The two vulnerabilities focused on in this research are sensor faults and malicious cyber-attacks.

- **Fault –** A sensor reading that deviates from the normal pattern exhibited by faulty sensor.
- **Attack –** A malicious attempt to falsify or deceive in order to report incorrect readings on peer-reported data.

**Dataset:** The dataset is an open-source sample released by the U.S. Department of Transportation, comprising BSM records captured by vehicles operating within the pilot study area situated in downtown Tampa, FL, USA. The data was truncated to only include 5,400 samples over a 4-hour time period and modified to simulate peer-report estimations [2].

## Motivation

- **Introducing Realism in Open-Source Dataset**

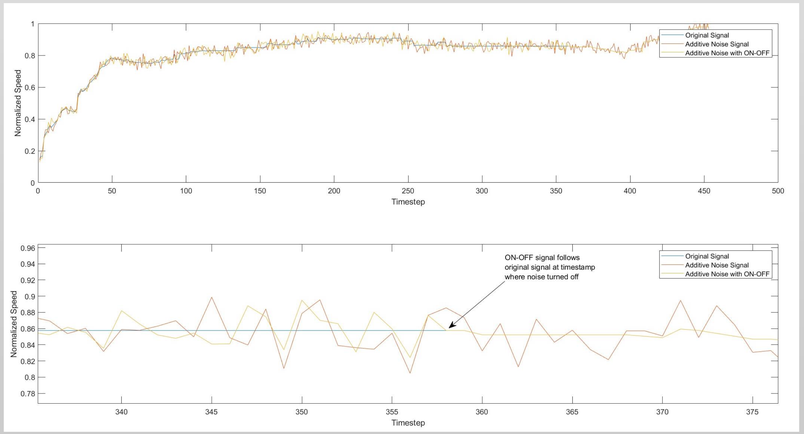
One of the reasons we used an open-source dataset is because it reflects real-world situations accurately. Many previous research papers relied on unrealistic closed-source or custom datasets, making replication difficult. Our research stands out by using an open-source real world dataset, ensuring others can reproduce and build upon our findings. This is vital for improving the security of CVs, as realistic datasets contribute to more effective security measures.

- **Emulating Realistic Attacks in Noise Model**

Another motivation for our research was to simulate attacks more realistically. In other research, fault/attack simulations were often binary - either fully on or completely off. Our approach introduced dynamic simulations, like a white noise attack with an on/off feature, mimicking real-world scenarios better. This adds credibility to our findings and emphasizes the need for proactive security strategies to protect CVs from sophisticated threats.

## Results

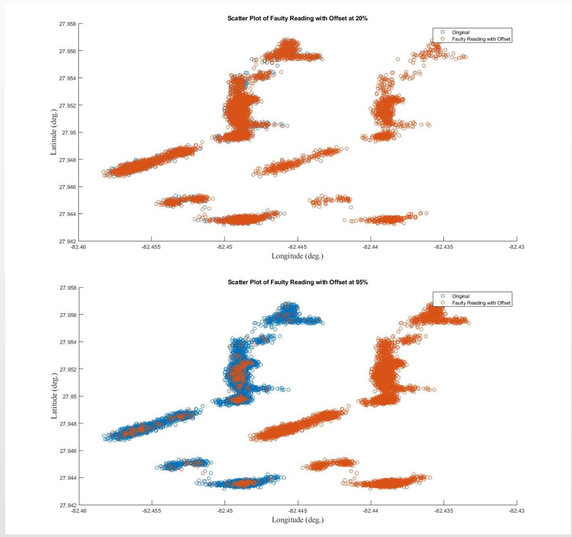
### (1) White Noise Attack with ON-OFF Feature




$s_{A\_OF}(t) = s(t) + \eta(t).w(t)$

Where, $s(t)$ is the uncorrupted signal and $\eta(t)$ is an additive noise process and $w(t)$ is a windowing function.

### (2) Random Offset Fault




$s_{A\_RO}(t) = s(t) + s(t)*r$

Where, $s(t)$ is the uncorrupted signal and $s(t)*r$ represents the random offset amount to be added to the original signal.

## Future Works

- **Incorporate/Simulate More Types of Faults**

Building upon the current research, future investigations can focus on simulating a broader range of connected vehicle faults, including outlier, spike, drift, and stuck-at scenarios.

- **Integrating a Trust Estimation Into The CV Workflow:**

Future research could explore techniques for assessing the trustworthiness of neighboring CVs within the network to improve the overall security of the system.

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. CNS-2149950. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## Contacts

Marbella Castillo: marbellacastillo50@gmail.com
Henna Rathore: heena.rathore@txstate.edu
Henry Griffith: hgriffith5@alamo.edu

## References

[1] https://data.transportation.gov/Automobiles/Tampa-CV-Pilot-Basic-Safety-Messages-BSM-Sample/nm7w-nvbm
[2] H. Griffith, M. Farooq, and H. Rathore, "A Data Generation Workflow for Consensus-Based Connected Vehicle Security". In *IEEE International Conference on Consumer Electronics (ICCE),* pp. 1-2, 2023.