Article title: Addressing Polymorphic Attack Strategies with Misbehavior Detection for ITS

Authors: Sultan Almalki[1], Frederick Sheldon [2]

Affiliations: Computer Science Department, University of Idaho, Moscow, ID, USA[1]

Orcid ids: 0000-0003-4534-2461[1]

Contact e-mail: dr.sultanalmalki.sa@gmail.com

License information: This work has been published open access under Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Conditions, terms of use and publishing policy can be found at <https://www.scienceopen.com/>.

Preprint statement: This article is a preprint and has not been peer-reviewed, under consideration and submitted to ScienceOpen Preprints for open peer review.

Links to data: Links to data:

||<https://data.transportation.gov/Automobiles/Next-Generation-Simulation-NGSIM-Vehicle-Trajectory/8ect-6jqj>

DOI: 10.14293/S2199-1006.1.SOR-.PPVNX1G.v1

Preprint first posted online: 10 February 2022

Keywords: MDS, cITS, VANET, Deep Learning.

Addressing Polymorphic Attack Strategies with Misbehavior Detection for ITS

Sultan Ahmed Almalki ¹, Frederick Sheldon ²

¹ University of Idaho, USA, alma6989@vandals.uidaho.edu

² University of Idaho, USA, sheldon@uidaho.edu

ABSTRACT

Cooperative Intelligent Transportation Systems (cITSs) is one of the Internet of Things (IoT) applications whose purpose is to improve road safety and traffic efficiency. Within this system, vehicles can communicate with one another by establishing a Vehicular Ad-Hoc Network (VANET) along the road section. Although such connectivity facilitates the exchange of information related to road safety and traffic efficiency, it puts the vehicles at risk in that an attacker could compromise one or more vehicles and use them to share false information causing congestions and/or life-threatening accidents. Although several studies tried to address this issue, they assume that the network topology and/or attack behavior is stationary, which is not realistic as the cITS is dynamic in nature and the attackers may have the ability and resources to change their behavior continuously. Therefore, these assumptions are not suitable and lead to low detection accuracy and high false alarms. To this end, this paper proposes a misbehavior detection model that can cope with the dynamicity of both cITS topology and attack behavior. The model starts by addressing the issue of missing data that happen at the early stages of the model formation after a topology change. Then, the deep learning approach is used to select the discriminative features used to train. We expect that the proposed model will help to overcome the limitations of related solutions by detecting attacks that change their behavior continuously.

Key words : MDS, cITS, VANET, Deep Learning.

1. INTRODUCTION

The purpose of Cooperative Intelligent Transportation Systems (cITSs) is to improve road safety and traffic efficiency. Within a road section, vehicles create a network to communicate and exchange different types of data regarding the traffic situation and safety information. However, such connectivity poses many threats to the cITSs vehicles (called nodes) outside as well as inside the network in the form of hijacked, rogue, and/or faulty nodes.

2. Research Motivation

- Vehicles in cITS are vulnerable to many forms of cyberattacks that compromise the data exchanged between the vehicles causing many operational disruptions like road

congestions and accidents. Existing MDS solutions try to protect the cITS systems by introspecting the patterns of the behavioral signatures of the cyberattacks. However, these solutions are built based on non-realistic assumptions on the reliability of the data exchange between the nodes in the cITS systems. The attackers could hijack one or more vehicles and manipulate the data they generate and share with the neighboring nodes. Therefore, relying on such compromised information to build detection models adversely affects the accuracy of those solutions.

- Existing MDS solutions assume that the attacks target only a limited number of vehicles. However, this assumption does not necessarily hold as some advanced attacks like botnets can compromise most of the nodes in the network, which invalidates the majority honest. These solutions adopted the anomaly detection approach to define the boundaries of normal behavior and considers the instances falling outside this boundary as attack.

- Existing solution incorporated an adaptation mechanism to the detection model to adjust the security thresholds dynamically in real-time and cope with the non-stationary nature of the cITS topology. However, they assume that the data related to the new driving situation are sufficient. This does not hold as the amount of data collected at early stages after cITS's topology changes might not be sufficient to accurately determine and build the new thresholds of the security profiles. Therefore, these solutions lack the ability to detect the attacks accurately.

- Existing solutions assume that the attackers always follow identical or similar attack strategies. However, many attackers and malicious software use obfuscation techniques and change attack strategies to deceive the detection. Therefore, they will become outdated very quickly and unable to cope with the new attack strategies.

3. Problem Statement

Although several studies have proposed MDS solutions for the cITS that assess the trustworthiness of the data exchanged between the nodes and cope with the dynamic nature of the topology, these solutions have several limitations affect the detection accuracy.

Limitation #1

Amount of data collected at early stages after cITS's topology change is insufficient to accurately determine/build the new thresholds of the security profiles because they contain lots of missing (null) as well as immature values which have a

negative impact on the ability of detection models to estimate features significance accurately

Limitation #2

Several existing solutions are unable to select significant features accurately

Limitation #3

Several attacks have dynamic nature (polymorphic attacks) that use sophisticated approach to change the behavior at each round to deceive the existing measures

4. Expected Innovative Contributions

- I will develop A correlation-based averaged fuzzy C-Means data imputation technique that estimates the missing values in the data collected at the early stages of online model's formation based on the local(univariate) and global(multivariate) average.

Based on the literature review, the existing solutions estimate missing values in the same attributes while mine will also consider the other correlated attributes in the calculation as well.

- I will develop A Deep Mutual Information Feature Selection (DMIFS) which incorporates the Deep Autoencoder (DAE) to calculate the values of redundancy and relevancy coefficients accurately when selecting the informative features from the dataset collected at the early stages of online model's formation.

Based on the literature review, the existing solutions estimate redundancy and relevance coefficients by using a statistical approach so my contribution here I will use the autoencoder to estimate these coefficients.

- Semantical features will be extracted and fed into the detection model in order to cope with the dynamic nature of the attacks' behavior to improve detection polymorphic attacks and re-adjust the security parameters accordingly.

The semantics features meaning the feature related to the behavioral. For examples, *How many time the vehicles are changing lane? How many data packets this vehicle sent per second? How many accidents this vehicle reports per second?* it will give me more insights about the behavior itself so the

model will become able to distinguish between the misbehaving vehicle and the honest one more accurately.

5. Research Methodology

6. Dataset

The dataset in this research will be the Next Generation Simulation (NGSIM) Vehicle Trajectories. It is an open source and publicly available data set with a collection of real-world vehicles' trajectories collected by smart vehicles and consist of different patterns representing different drive situations and driver behavior. Particularly, NGSIM was built by collecting data from vehicles moving on a road section with 500 m long and seven lane highway. For each vehicle, the data is collected (recorded) for 45 minutes using 16 sensors.

7. CONCLUSION

A misbehavior detection model is proposed that address the limitation of existing solutions that assume that the nature of cITS topology as well as the attack strategy are stationary. The model consists of three main components, pre-processing, feature selection, and training/testing. The pre-processing estimates the missing data that occur at the early stages of model formulation after the change occurs. The discriminative features are chosen at the feature selection stage by employing a deep auto encoder to calculate the redundancy coefficients more accurately. Deep learning is then used to build the model by training the DBN using the data and features prepared and selected during the 1st and 2nd phases. In our future work, the three components of the model will be developed and integrated into an MDS model that can

prevent evasive attacks that target the cITS ecosystem and protect the data exchanged between vehicles from being manipulated, jammed and/or spoofed.

REFERENCES

1. Almalki, S. A., & Sheldon, F. (2020, December). Prospectus: An Online Polymorphic Attack Detection Model for Intelligent Transportation Systems. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1115-1119). IEEE.
2. Almalki, S. A., & Song, J. (2020). A review on data falsification-based attacks in cooperative intelligent transportation systems. *International Journal of Computer Science and Security (IJCSS)*, 14(2), 22.
3. Almalki, S. A., & Sheldon, F. T. (2021, October). Deep Learning to Improve False Data Injection Attack Detection in Cooperative Intelligent Transportation Systems. In *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)* (pp. 1016-1021). IEEE.