SJSU SAN JOSÉ STATE UNIVERSITY

FRESNO STATE
Discovery. Diversity. Distinction.

Taming the Data in the Internet of Vehicles

Shahab Tayeb

Project 2014
January 2022

To stay ahead of novel attacks, cybersecurity professionals are developing new software programs and systems using machine learning techniques. Neural network architectures improve such systems, including Intrusion Detection System (IDSs), by implementing anomaly detection, which differentiates benign packets from malicious packets. For an IDS to best predict anomalies, the dataset the model is trained on is typically pre-processed through normalization and feature selection/reduction. Pre-processing techniques play an important role in training a neural network to optimize its performance.

In this study, we extend the current research on the importance of data normalization through developing, training, and testing a DNN on the CIDDS network data. To this end, we evaluate the effect of Z-Score and Min-Max normalization on the model's accuracy, loss, F-Score, and AUC-ROC. Additionally, an analysis and comparison of the performance of the model on the NSL-KDD and CIDDS datasets are carried out. Note that the studied normalization techniques are considered lightweight, due to their minimal overhead, which makes them potential solutions for in-vehicle applications. This is particularly true for pre-trained implementations of the model.

Study Methods

This study utilizes state-of-the-art pruning and normalization techniques in training neural network architecture. Various learning metrics, including accuracy, Area Under Curve (AUC), Receiver Operator Characteristic (ROC), F-1 Score, and loss, are used to evaluate the performance.

Findings

Normalization and other pre-processing techniques applied to the data used for training an IDS are important for optimizing the performance metrics. We propose a DNN using 27 input features for binary classification trained using the NSL-KDD dataset. As expected, the experimentation on the pruned dataset outperforms experimentation on the complete dataset across most metrics. Our proposed model determines for the complete dataset that Z-Score normalization, and Min-Max normalization to a lesser degree, improves the performance of the proposed IDS compared to no normalization. Our results demonstrate that Z-Score improves on no normalization by 4.46% for accuracy, 2.04% for loss, 4.70% for F-Score, and 23.64% for AUC-ROC. Min-Max normalization presents similar improvements with a 1.99% increase in accuracy, 1.00% decrease in loss, 0.32% increase in F-Score, and 23.65% increase in AUC-ROC. Implementing Z-Score normalization as a pre-processing step can improve the performance of DNN-based IDSs. Such pre-processing of the training model adds little to no overhead on the in-vehicle implementation of such a model, justifying the use of these pre-processing techniques. The study also highlights the optimal number of epochs for the training, after which little gain in accuracy is observed.

Security of the Internet of Vehicles (IoV) is an emerging field, and IoV has a myriad of security vulnerabilities.

Policy/Practice Recommendations

Security of the Internet of Vehicles (IoV) is an emerging field and IoV has a myriad of security vulnerabilities. Such security has usually been an afterthought, which has led to the adoption of decade-old standards in the underlying infrastructure of vehicular communication systems. One approach to mitigating such vulnerabilities would be to use lightweight IDS solutions at the edge of the network. The pre-processing of the training model ensures the optimal usage of in-vehicle resources while providing security via a learning architecture.

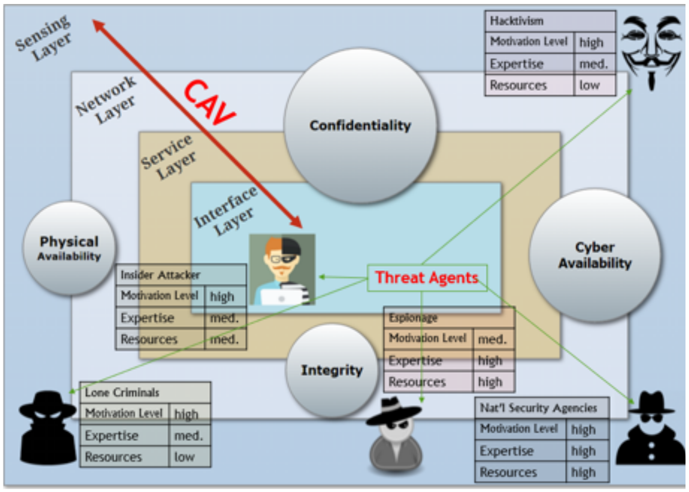

Figure 1. Vulnerability Surfaces of the Emerging Internet of Vehicles

About the Author

Dr. Shahab Tayeb is a faculty member with the Department of Electrical and Computer Engineering in the Lyles College of Engineering at California State University, Fresno. Dr. Tayeb's research expertise and interests include network security and privacy, particularly in the context of the Internet of Vehicles. His research incorporates machine learning techniques and data analytics approaches to tackle the detection of zero-day attacks. Through funding from the Fresno State Transportation Institute, his research team has been working on the security of the network backbone for Connected and Autonomous Vehicles over the past two years. He has also been the recipient of several scholarships and national awards including a US Congressional Commendation for STEM mentorship.

To Learn More

For more details about the study, download the full report at **transweb.sjsu.edu/research/2014**

MTI MINETA TRANSPORTATION INSTITUTE

MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.