Mark Demore, 2d Lt

CSCE623 – Dr. Borghetti

Project Final Draft

ABSTRACT – GPS Spoofing has become more common, as the ease to implement such an attack increases. All USAF assets rely on GPS data for navigation and timing, and are vulnerable to spoofing attacks. A data focused approach would provide a flexible, lightweight solution that could be easily implemented on a variety of systems. Machine learning techniques provide a promising way to develop such a system. Logistic regression, naïve bayes, and random forests classifiers were trained on simulated spoofing data to detect a spoofing attack. All of these methods showed promise when given appropriate training data, and random forests proved to be the best solution with high recall, even outperforming a deep learning method implemented in previous research. Future work is needed to test the approach with more subtle spoofing attacks.

1. Introduction

The proliferation of Software-Defined Radio (SDR) and other Commercial Off-the-Shelf (COTS) systems has led to increased vulnerability of United States Air Force (USAF) systems through the spoofing of Global Positioning System (GPS) signals. USAF assets (drones, airplanes, ground systems) rely on GPS signals to coordinate timing and for navigation. While military GPS signals are encrypted and more difficult to spoof, many systems also use civilian GPS messages. The MIL-STD-1553 avionics bus coordinates traffic between systems on USAF assets. It times messages between a bus controller (BC) and the different components (GPS, INS, Flight Computer), or remote terminals (RTs). The most effective lightweight solution to counter GPS spoofing would be in a bus monitor for MIL-STD-1553 that can parse the messages between the avionics and detect GPS spoofing, using a machine learning approach.

In this work, I will address the following key research question: Can a supervised machine learning algorithm classify MIL-STD-1553 messages as spoofed based on observations from the bus? I hypothesize that a machine learning algorithm can correctly classify MIL-STD-1553 messages with statistical significance using bus traffic from other avionics components as features, 13 in total. I will also hypothesize that a more drastic spoofing attack, one with more deviation in position from true position, will be easier to detect and have higher accuracy. The features used in this project are: GPS Latitude, GPS Longitude, GPS Altitude, ADC Altitude, ADC Air Speed, EFIS Angle of Attack, INS X-Velocity, INS Y-Velocity, INS Z-Velocity, INS X-Acceleration, INS Y-Acceleration, and INS Z-Acceleration. The Inertial Navigation System (INS) data will likely have the most significance in classifying when compared to the GPS data .My research objective is to develop a classification model that can correctly classify spoofed GPS messages on the MIL-STD-1553 bus with accuracy of 90% or greater.

The data used in this project has been procured from a previous student’s thesis work. It was generated using the Air Force Research Lab’s (AFRL) Avionics Vulnerability Assessment System (AVAS). The data was collected from 5 flights of lengths of about 4 hours, for a total of approximately 700,000 observations per flight, each collecting all of the aforementioned features. Spoofing occurred for periods of 3 minutes and was recorded as truth data, with equal portions of the flight being spoofed and not spoofed.

1. Related Work

A number of papers have been published on the application of machine learning to anomaly detection in time-based systems. Marvin attempted a deep learning approach to detect spoofing with mixed results, likely due to issues with the data and the additional complexities of a neural network application [5]. Genereux, et al. has used the timing of messages over the 1553 bus to detect anomalies and intrusion using a simple histogram approach [7]. Other work includes the use of Support Vector Machines to detect GPS spoofing using the signals themselves and a supervised approach [8]. More generalized research includes the use of Markov models and relative probabilities [1], and applications of deep learning to detect ADS-B spoofing [4] and network intrusion [2]. The aforementioned research is representative of the applicability of machine learning methods to this problem. This project differs mostly in the data that is being used, messages from the MIL-STD-1553 bus. The advantage of using 1553 data is that developing a simple bus monitor with the detection algorithm would provide a lightweight solution that can be implemented on any USAF asset.

1. Methodology

The data has been collected using the AVAS system in tandem with AFRL’s Vampire, an FPGA-based bus monitor, that sent the 1553 messages as UDP packets to a Windows 10 machine running a Python script to save the data to a CSV file. The flights were flown by hand in level and banked flight orientation, as well as with varying altitudes with climb and descent maneuvers. The spoofing attacks implemented drift and hold position attacks of varying displacement rates (5, 10, 50, and 100 m/s). The data is then parsed using an XML file encoded from the 1553 bus Interface Control Document (ICD), which details how the data in the message is wrapped. This is necessary because each platform has a different ICD, and the intermediate step will allow the solution to be flexible and fit any system once an XML is made for its specific bus architecture. Messages that do not contain information about the 13 features of interest is dropped upon parsing and left out of the data.

To solve this binary classification problem, I will compare the effectiveness of logistic regression, Naïve Bayes, and Support Vector Machines (SVM). These methods will be compared to random guessing as a baseline alternative. I will use feature scaling and regularization of the data before feeding it to the algorithms. I will forgo feature selection methods, as I believe all features to be significant, although this step may be included in future work. Two different methods will be used in determining test data, in one I will randomly select one of the flights data as test data to withhold, and the other I will randomly select 20% from all flight data to withhold as test data. I will use k-fold cross validation of varying k-values (3, 5, and 10) to tune my models before giving them the test data. The primary performance measure will be the accuracy of predictions on unseen flight data. Other measures will include true and false, positive and negative rates, as well as precision. These measures will be gathered using the truth data and analyzed in tabular form, as well as using a receiver operating characteristic curve.

1. Results

I expect the results to indicate that an SVM approach would be most accurate, followed by Naïve Bayes and then logistic regression, due to the large number of observations and features. I also expect these methods to be more than 90% accurate, with low false positive and false negative rates. These results will prove with statistical significance that a supervised machine learning approach can be implemented as a MIL-STD-1553 bus monitor to effectively detect GPS spoofing attacks.

1. References

[1] Du, Y., Wang, H., & Pang, Y. (2004). A Hidden Markov Models-based anomaly intrusion detection method. Proceedings of the World Congress on Intelligent Control and Automation (WCICA), 5, 4348–4351. https://doi.org/10.1109/wcica.2004.1342334

[2] Van, N. T., Thinh, T. N., & Sach, L. T. (2017). An anomaly-based network intrusion detection system using Deep learning. Proceedings - 2017 International Conference on System Science and Engineering, ICSSE 2017, 210–214. https://doi.org/10.1109/ICSSE.2017.8030867

[3] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). An Introduction to Statistical Learning - with Applications in R | Gareth James | Springer. Retrieved from https://www.springer.com/gp/book/9781461471370%0Ahttp://www.springer.com/us/book/9781461471370

[4] Ying, X., Mazer, J., Bernieri, G., Conti, M., Bushnell, L., & Poovendran, R. (2019). Detecting ADS-B Spoofing Attacks Using Deep Neural Networks. 2019 IEEE Conference on Communications and Network Security, CNS 2019, 187–195. https://doi.org/10.1109/CNS.2019.8802732

[5] Marvin, J. M. (2019). Detecting GPS Spoofing with Deep Learning. https://doi.org/10.1109/TE.1962.4322266

[6] Wang, J., Tu, W., Hui, L. C. K., Yiu, S. M., & Wang, E. K. (2017). Detecting Time Synchronization Attacks in Cyber-Physical Systems with Machine Learning Techniques. Proceedings - International Conference on Distributed Computing Systems, 2246–2251. https://doi.org/10.1109/ICDCS.2017.25

[7] Genereux, S. J. J., Lai, A. K. H., Fowles, C. O., Roberge, V. R., Vigeant, G. P. M., & Paquet, J. R. (2020). MAIDENS: MIL-STD-1553 Anomaly-Based Intrusion Detection System Using Time-Based Histogram Comparison. IEEE Transactions on Aerospace and Electronic Systems, 56(1), 276–284. https://doi.org/10.1109/TAES.2019.2914519

[8] Semanjski, S., Muls, A., Semanjski, I., & De Wilde, W. (2019). Use and validation of supervised machine learning approach for detection of GNSS signal spoofing. 2019 International Conference on Localization and GNSS, ICL-GNSS 2019 - Proceedings, 1–6. https://doi.org/10.1109/ICL-GNSS.2019.8752775