Supervised ADS-B Anomaly Detection Using A False Data Generator

Ralph Karam, Michel Salomon, Raphael Couturier

Supervised ADS-B Anomaly Detection Using A False Data Generator

Ralph Karam, Michel Salomon, Raphaël Couturier
FEMTO-ST Institute, CNRS
Univ. Bourgogne Franche-Comté (UBFC)
Belfort, France
{ralph.karam,michel.salomon,raphael.couturier}@univ-fcomte.fr

***Abstract*—The ADS-B is an air traffic monitoring technology based on broadcasting messages to transfer information in between aircraft as well as between aircraft and ground stations. It was created to increase the surveillance coverage as well as reduce the cost of operation relative to traditional radars. However the messages used to communicate under the ADS-B protocol are not encrypted and thus are prone to false data injection attacks which can, for example, modify the values of the messages' components. In this paper, a supervised deep learning strategy is designed to detect attacks that modify components of ADS-B messages such as altitude, ground speed, trajectory, latitude and longitude. A false data generator based on a domain specific language was used to attack ADS-B data and obtain a dataset containing both normal and anomalous data for supervised learning. The detection performance of two types of attacks were evaluated: gradual attacks and waypoints attacks which diverge aircraft trajectories to pass through specific waypoints. The experimental results show that the proposed supervised deep learning strategy is able to recall on average 99% of anomalies in ADS-B messages, mainly property modification attacks.**

***Index Terms*—anomaly detection, deep learning, false data injection, supervised learning, ADS-B protocol**

I. Introduction

Air transport is one of the driving forces behind international economic growth. Its rising prevalence over the decades is one of many indicators of increasing global living standards. However the increase in air traffic can be a source of concern notably because of multiple safety risks. These problems are caused by the rapid growth of airspace crowding which can start to outrun infrastructure and coordination evolution. To prevent these kinds of risks the International Civil Aviation Organization (ICAO) devised a strategy called the Global Air Navigation Plan (GANP) which establishes the needed modernization for the upcoming years. This can ensure the airspace's ability to withstand the growing air traffic. In addition to that, the plan aims at conceiving more efficient routes for aircraft resulting in a decrease in carbon emissions. More concretely, GANP refers to the various steps needed to continuously improve the global air navigation system in the following areas [1]: airport operations, interoperable systems and data, globally collaborative Air Traffic Management (ATM), efficient flight paths. One of the most important tools that can help improve the mentioned areas is the Automatic Dependent Surveillance Broadcast (ADS-B) protocol.

The ADS-B is a surveillance technology able to give a much more accurate 3D position of aircraft compared to traditional radars. Based on this protocol, aircraft equipped with special transponders are able to broadcast their information such as the identification (or call sign), position, altitude, and velocity in the form of ADS-B messages. Aircraft equipped, as well as ground stations, are able to receive these types of messages to obtain a precise image of the airspace [2]. This protocol has many benefits. It can reduce total expenses relative to traditional radar systems due to its lower operational cost. In addition to that, ground stations can be located in hardly accessible and remote areas due to their non mechanical nature. The protocol can also be used in the process of finding optimal flight levels to increase fuel savings [1], and thus helps aircrafts to be more environmentally-friendly by reducing CO2 emissions. However, the ADS-B technology presents some disadvantages such as the requirement for airplanes to have special transponders and the inability to validate the position with ground stations since the position is solely obtained on board [2]. Finally the ADS-B messages are prone to false data injections since the ADS-B protocol lacks encryption and authentication for practical reasons.

In this paper is introduced a new supervised deep learning anomaly detection strategy to detect ADS-B gradual attacks as well as a another type of attack called waypoints attack (forcing a flight to pass by specific predefined points). Gradual attacks concern the following features: altitude, ground speed, track, latitude and longitude. These gradual attacks are detected whether they are separately inflicted (feature by feature) or simultaneously (all the specified features are attacked). The altitude and ground speed gradual attacks that we were able to detect (75 feet altitude gradual attack, 1.8 knots gradual ground speed attack) are harder to spot then those detected in the literature (400 feet altitude gradual attack in [3], [4] and 20 knots gradual attack in [4]). This research as well as our previous one published in [5] are the only studies which, to the best of our knowledge, use supervised learning for ADS-B anomaly detection. Our research and [6] use the

This work was supported by the DGA (French defence procurement agency) in the context of the GeLeaD project (project number ANR-18-ASTR-0011) related to the ANR ASTRID research program (specific support scheme for research works and innovation defence). It was also partially supported by the EIPHI Graduate School (contract ANR-17-EURE-0002). Computations have been performed on the supercomputer facilities of the "Mésocentre de Franche-Comté".

same false data injection generator (with a domain specific language) specifically devised for generating attacked ADS-B datasets contrary to anything else found in the literature. Finally we were able to detect gradual attacks on any of the previously mentioned features using a meta-model made of models trained on individual gradual attacks.

The remainder of this paper is organized as follows. Section 2 presents a summary of related studies. The next section details the anomaly detection process (data generation as well as the detection process). Section 4 presents the experimental results of the proposed detection strategy. Finally some conclusions are drawn regarding detection strategies.

II. Related Works

There are many studies in the field of ADS-B anomaly detection using mainly unsupervised and semi-supervised learning techniques. Obviously, the lack of datasets providing examples of attacks explains why works in a supervised setting are rare. This work, which takes place in this latter context, is thus different from this point of view from the research works reviewed thereafter.

The authors in [3] used semi-supervised learning to detect anomalous ADS-B messages. They based their technique on a LSTM encoder-decoder architecture. Normal sequences of messages can be easily reconstructed with this architecture as opposed to anomalous messages which cannot be successfully reconstructed (i.e. a high reconstruction error indicates an attack). In [7] sequences of ADS-B messages are transformed into sequences of images in which the dimension of shapes and colors represent features such as latitude, longitude and altitude. Since these images evolve relative to time, they can be treated as videos, and for this reason a Convolutional-LSTM encoder-decoder is used for anomaly detection. The authors in [4] used a model that learns to predict the next message from a window of successive messages. If the prediction is far from the observed message, an anomaly is detected. The model that they used is composed of a LSTM layer containing 14 units followed by a fully-connected layer of 7 units, with one output for each message feature.

In [8] a hidden Markov model is used, where hidden states are associated with normal historical ADS-B data and other ones for the observed ADS-B data. To obtain the parameters of the model in a dynamic manner a sticky hierarchical Dirichlet process is used. The authors in [9] were able to detect anomalies by comparing the observed flight relative to the estimated optimal waypoint route: if it is far enough it is considered as a potential anomaly. Using this technique, a route is divided into multiple segments bounded by reference waypoints. Then the points in each segment are partitioned into two clusters using Agglomerative Hierarchical Clustering. The route passing by the centroids of clusters which are the closest to reference waypoints is considered as the optimal.

In [10] a set of flights is clustered using DBSCAN. Then 50 points are sampled from each flight. After that, in each cluster an autoencoder is used which takes 50 input data points for each flight to detect anomalies based on the magnitude of the reconstruction error. This work uses thus a combination of clustering and an auto-encoder architecture, which is the deep learning model mainly considered in works studying the detection of spoofing attacks targeting the ADS-B system.

All these mentioned papers are not based on supervised learning which was an incentive for the authors of this paper to choose these conditions. In a previous work, published in [5], a first evaluation of the interest of supervised training for the detection of attacks on the ADS-B protocol was completed. However only one type of anomalies was used for detection (random punctual altitude changes). In this paper, which does not only target the altitude feature, two more types of attacks (gradual attack and waypoints attack) are tested for anomaly detection and a different detection strategy is also introduced.

III. Detection of False Data Injection

A. Generation of Labeled Attacked ADS-B Messages

As highlighted previously, anomaly detection in the literature uses only semi-supervised and unsupervised learning. In this paper we propose a technique to detect anomalies using labeled messages (supervised anomaly detection) to train a deep learning model. These messages are obtained thanks to a software, called FDI-T (which stands for False Data Injection Testing framework), designed by colleagues in our laboratory, allowing the generation of false air traffic data [11], [12]. This software implements an alteration process making it possible to define various scenario attacks according to alteration directives. Fig. 1 gives an overview of the alteration process. It consists of 5 items which are detailed below.

First, ADS-B messages are downloaded from the OpenSky Network (opensky-network.org) [13], a community-based receiver network which continuously collects air traffic surveillance data. For the sake of having balanced datasets (50% attacked messages, 50% not attacked messages) half of the recordings are totally attacked and the other half is not touched at all. Second, in order to realize attacks on ADS-B data, directives (or instructions) in a domain specific language (DSL) are used to describe the alterations to be performed on the original recorded data. Each instruction is then processed by FDI-T (third item) to call the alteration engine in charge of the specified alteration (fourth item). Finally, the resulting altered

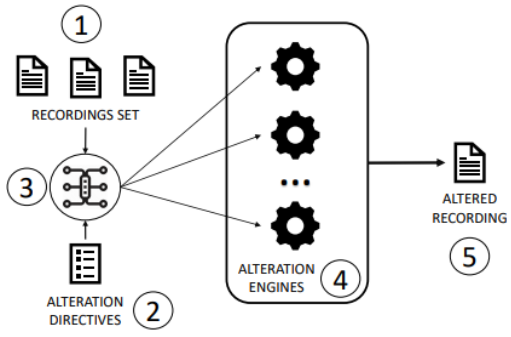


Fig. 1. Alteration process overview - image drawn from [12].

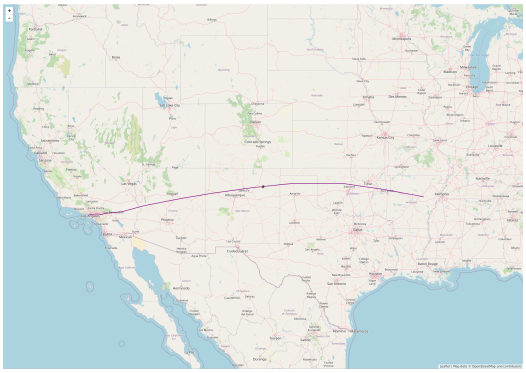
data are recorded. Note that the FDI-T software has also been used to acquire ADS-B datasets to test anomaly detection AI models in another work done in our laboratory [6]. In this work two types of attacks were considered for anomaly detection: gradual attacks and waypoints attacks.

A gradual attack of Δx to a feature x in ADS-B messages means that Δx, $2\Delta x$, $3\Delta x$, etc. are added to the feature x in the first, second, third,... message respectively. For the case of gradual attacks detection, attacks are separately inflicted on altitude, ground speed, track, latitude and longitude, i.e.: gradual attack of 75 feet, 1.8 knots, 0.9 degrees, 4.88×10^{-3} degrees, 1.28×10^{-2} degrees for the altitude, ground speed, track, latitude, longitude respectively. These values are computed by finding the mean difference of the features in non attacked messages. The following DSL instruction was used for the gradual attack of $\Delta x = 75$ feet targeting the altitude:

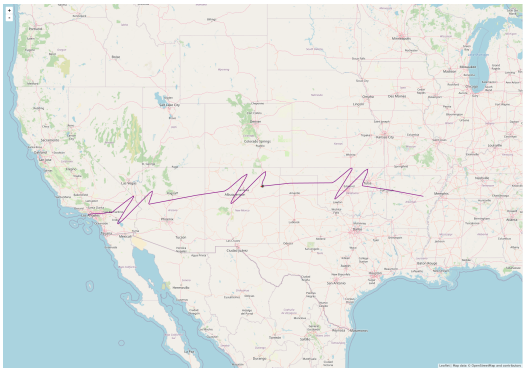
"alter all_planes at 0 seconds with_values ALTITUDE ++= 75".

The other features were gradually attacked in the same manner.

Another type of attack tested for detection is the waypoints attack. A waypoints attack is an attack in which the trajectory is deviated (using an Akima interpolation [14]) to pass by specific selected points with known latitude, longitude and altitude. An example of a waypoints attack on a flight connecting two US cities is shown in Fig. 2 and is also presented in detail in Fig. 6 which will be explained in Section IV.



(a) Before the attack



(b) After a waypoints attack

Fig. 2. Screenshots of the ADS-B false data injection software.

B. Meta-messages Generation and Detection

In order to detect attacks, the process described in Fig. 3 is followed. First ADS-B messages corresponding to 120 randomly chosen flights are downloaded from the OpenSky Network. These flights are not restricted to a specific area, they are dispersed all over the globe and contain all of the different phases of flight: take off, climb, cruise, descent and landing. Since the features that will be used for anomaly detection are distributed among multiple categories of messages (Identification, Position and Velocity messages), these messages are combined into meta-messages (as can be seen in Fig. 4).

In other terms, messages are vectors with missing data and the process of combining consecutive different types of messages into meta-messages is just filling missing features from previous messages. Then the difference for all the meta-messages between two consecutive meta-messages is computed as seen in Fig. 5. This process is applied to all the flight data gathered in the first step. Among the obtained differences of meta-messages from the 120 flights of the dataset, the messages of 100 flights are used to train a model to detect ADS-B anomalies and the remaining 20 flights to test it.

Finally for the detection of anomalies a LSTM is trained. A Long Short-Term Memory (LSTM) is a special kind of recurrent neural networks (RNN) created to tackle the problem of vanishing gradients that a regular RNN suffers from [15]. The input of our LSTM is a look back window of differences of meta-messages, where the number of time steps is defined by the lookback value. Its output is one dense neuron with a linear activation function which captures the state of the window, i.e. either a normal or an attacked window.

The choice of the LSTM architecture used in this study is based on the result of our previous work [5] which focused only on the detection of random point changes in altitude. Indeed this work concluded that an LSTM containing intermediate layers of 64 units and 32 units, trained with a "nadam" optimizer, gives the best performance for ADS-B anomaly detection compared to other architectures such as the Bidirectional LSTM, 1D Convolutional Neural Network, and so on. Note that in our case, scaling should not be applied on the data or it will deteriorate the detection performance.

IV. Experimental Results

The detection performance of our technique is computed in terms of Recall, Precision and F-score. The Recall also called Sensitivity and True Positive Rate (TPR) represents the fraction of detected attacked messages relative to the total number of attacked messages. The Precision denotes the proportion of detected attacked messages that are correctly identified as real attacks. Finally the F-score is the the harmonic mean of the precision and recall [16].

In order to train and test the LSTM model on GPUs, the supercomputer facilities of the "Mésocentre de Franche-Comté" were used. Our code is based on the Keras Python library. First the architecture is trained for 10 epochs, 50 epochs and 100 epochs in order to test the effect of the number of epochs on the detection performance as seen in Table I. The attack

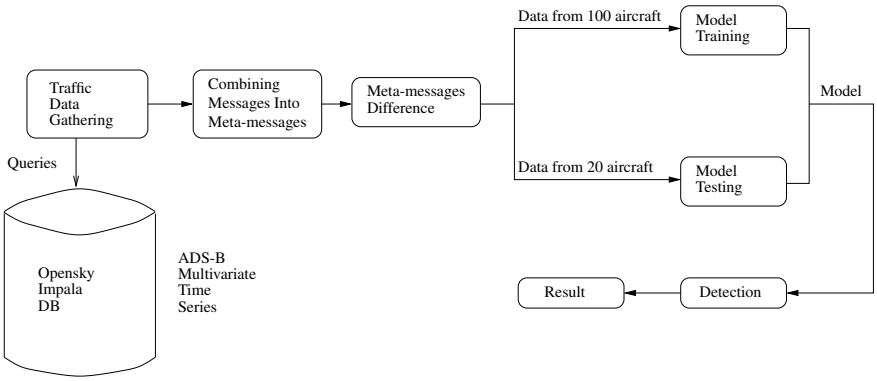


Fig. 3. The whole detection process.

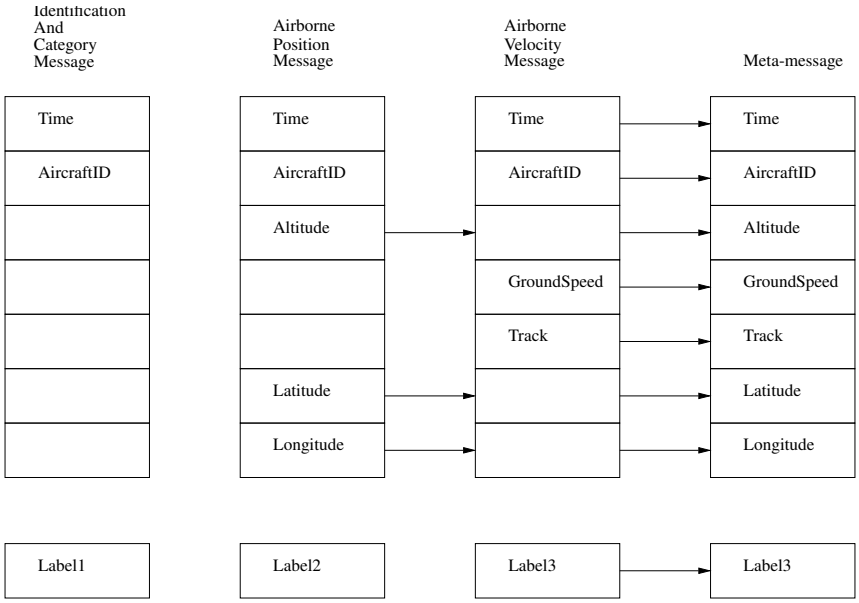


Fig. 4. Example of combining messages into a meta-message.

used for this assessment is a gradual attack of the longitude equal to 1.28×10^{-2} degrees.

Since the performance did not increase considerably, 10 epochs can be enough to have a good detection performance for the following tests while having the shorter training time. Indeed the training time is proportional to the number of epochs and an epoch takes 14 seconds to be completed (for example it takes 2.33 minutes to finish 10 epochs). The difference between training times is not negligible especially due to the following need to also train models for the gradual attacks targeting other features. Lookback values of 5, 10 and 20 were tested in the detection of average gradual attacks. Their results are summarized in Table II. It can be noticed that when the lookback value is increased, most of the time the F-score also increases. However this gain in F-score is minor. For this reason a lookback value of 10 was used for the following tests balancing in this manner the training time and the detection performance. In order to compare the gradual attack detection with other previous works [3], the altitude was attacked gradually by 400 feet. Our detection performance in terms of True Positive Rate (TPR) and False Positive Rate (FPR) compared to the best performance obtained in [3] using one detection window with a lookback equal to 15 is summarized in Table III. FPR, also called Fallout, represents the proportion of detected attacked messages that are in reality not attacked [16]. It is clear that using our technique, a considerable detection performance improvement is obtained (40.93 increase in %TPR and 3.84 decrease in %FPR).

TABLE I
EVALUATION OF THE STACKED LSTM USING DIFFERENT NUMBERS OF EPOCHS (LAYERS OF 64 AND 32 UNITS - LOOKBACK VALUE OF 10)

Epochs	%Precision	%Recall	%F-score
10	99.89	99.70	99.80
50	99.94	99.65	99.79
100	99.94	99.82	99.88

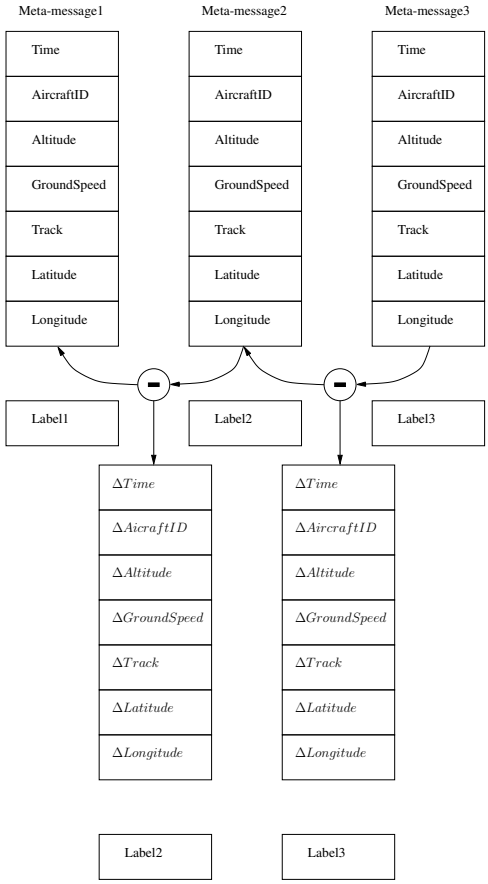


Fig. 5. Difference of meta-messages.

TABLE II
EVALUATION OF THE STACKED LSTM USING DIFFERENT LOOKBACKS FOR THE DETECTION OF GRADUAL ATTACKS

Feature attacked	Lookback	%Precision	%Recall	%F-score
altitude	5	99.35	99.43	99.39
	10	99.63	99.76	99.69
	20	99.93	99.69	99.81
ground speed	5	98.86	99.48	99.17
	10	99.23	99.60	99.42
	20	99.29	99.68	99.49
track	5	96.91	99.15	98.02
	10	97.46	99.45	98.44
	20	97.52	99.50	98.49
latitude	5	99.90	98.18	99.03
	10	99.67	98.91	99.29
	20	99.72	98.97	99.34
longitude	5	99.83	99.01	99.41
	10	99.89	99.70	99.80
	20	99.60	98.95	99.27

TABLE III
DETECTION PERFORMANCE COMPARED TO HABLER & SHABTAI (MOSCOW DATASET) FOR 400 FEET GRADUAL ALTITUDE ATTACKS

	%TPR	%FPR
Proposal	99.97	2.95×10^{-2}
E. Habler and A. Shabtai	59.04	3.87

Now a meta-model is tested. It uses the different models trained separately and detects the presence of an attack if at least one of the previously mentioned features (altitude, ground speed, etc.) is attacked. The performance of this meta-model is summarized in Table IV. The F-score obtained is 96.09% at worse with a Precision of 93.15% which is still good but a little worse than individual models as seen in the previous tables. The reason for the decrease in F-score is the slight loss of Precision which is caused by the False Positives of the different models adding up. In order to remedy the loss of Precision, we then considered an attack detected if a certain percentage of windows (and above) from a sequence of windows is attacked: the obtained performance is summarized in Table V. Note that it was considered that a given percentage of attacked windows (95% and above) from a sequence of windows need to be attacked instead of the whole sequence to prevent eventual False Negatives from deteriorating the Recall. Using this technique, the Precision becomes at worse 99.44% instead of 93.15% and the F-score 97.18% instead of 96.09%. Although a better Precision was obtained, there was a small decrease in Recall which highlights a slight trade-off between Recall and Precision using this technique.

For the waypoints attack as seen in Fig. 6, using the same technique, a good detection performance was also reached: Precision=98.77%, Recall=97.39%, F-score=98.08%. However, all the results presented so far were obtained using one training set and one testing set for each result. This way of obtaining the performance is not always enough since deep learning models can give unstable results. For this reason a stratified 6-fold cross-validation was applied whose results are shown in Table VI. This table shows the mean and standard deviation of the Precision, Recall and F-score for gradual and waypoints attacks. The mean F-score is in the order of 99% and its worst standard deviation is 0.4420 hence our technique is stable enough to detect the previously mentioned anomalies.

TABLE IV
EVALUATION OF A META-MODEL FOR THE DETECTION OF GRADUAL ATTACKS

Feature attacked	%Precision	%Recall	%F-score
altitude	93.58	99.80	96.59
ground speed	96.50	99.70	98.07
track	96.49	99.53	97.99
latitude	93.15	99.22	96.09
longitude	93.18	99.77	96.36
all features	96.97	100	98.46

TABLE V
EVALUATION OF A META-MODEL FOR THE DETECTION OF GRADUAL ATTACKS USING SEQUENCES OF 100 WINDOWS, WHERE AN ATTACK IS DETECTED IF AT LEAST 95 OF THEM ARE ATTACKED

Feature attacked	%Precision	%Recall	%F-score
altitude	99.49	98.85	99.17
ground speed	99.73	98.57	99.15
track	99.72	97.14	98.41
latitude	99.44	95.02	97.18
longitude	99.46	98.73	99.09
all features	99.76	100.0	99.88

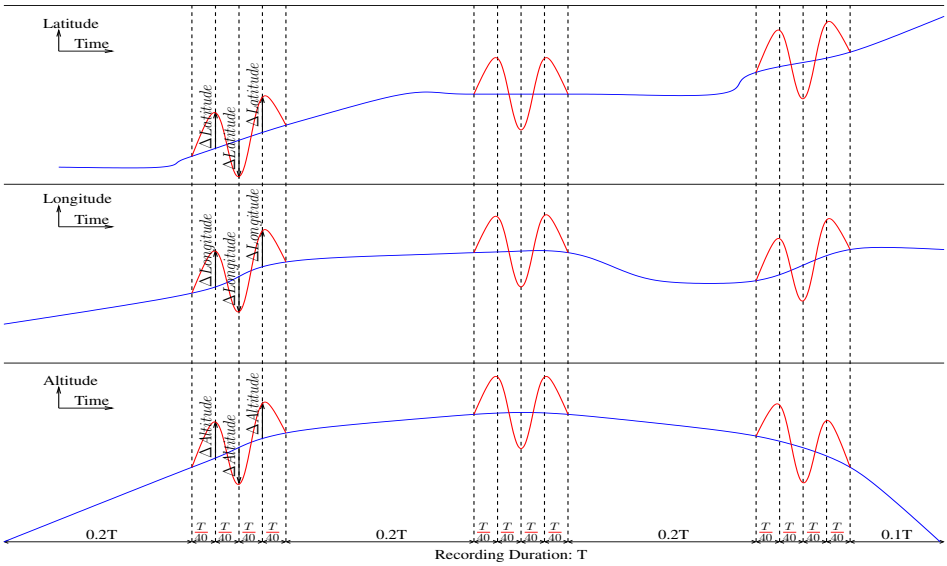


Fig. 6. Waypoints attack where $\Delta Latitude = 4.88 \times 10^{-3}$ degrees, $\Delta Longitude = 1.28 \times 10^{-2}$ degrees, $\Delta Altitude = 75$ feet.

TABLE VI
EVALUATION OF THE STACKED LSTM
USING A STRATIFIED 6-FOLD CROSS-VALIDATION

Feature attacked	%Precision	%Recall	%F-score
altitude	99.92 ± 0.0130	99.83 ± 0.0294	99.88 ± 0.0114
ground speed	99.77 ± 0.0301	99.74 ± 0.0178	99.76 ± 0.0120
track	99.45 ± 0.0941	99.41 ± 0.1680	99.43 ± 0.0648
latitude	99.41 ± 0.3280	99.26 ± 0.5940	99.34 ± 0.4420
longitude	99.89 ± 0.0482	99.66 ± 0.1320	99.78 ± 0.0728
waypoints - Fig. 6	98.42 ± 2.23	99.08 ± 0.154	98.73 ± 1.08

V. CONCLUSION

A supervised deep learning strategy to detect false data injection attacks aiming at altering properties of messages in the ADS-B protocol has been presented. This strategy was evaluated for the identification of gradual attacks and waypoints attacks created using a false data generator. The effects of the number of training epochs and the lookback value on the detection performance were tested. It was found that only a minor improvement of the detection performance results from increasing the number of training epochs and the lookback value. We were able to detect gradual attacks of the altitude, ground speed, track, latitude and longitude using individual models trained on each individual attack, as well as using one meta-model made of these individual models. In addition to that the strategy was successful in detecting waypoints attacks.

REFERENCES

[1] D. ICAO, "Draft 2016-2030 Global Air Navigation Plan," tech. rep., Doc 9750-AN/963, Fifth Edition, 2016., Montreal, Canada.

[2] "Study on the convenience and feasibility of space-based ADS-B for regional implementation." https://www.icao.int/NACC/Documents/Meetings/2018/ADSB/D05-AireonICAOPaper-EENGr.pdf.

[3] E. Habler and A. Shabtai, "Using LSTM encoder-decoder algorithm for detecting anomalous ADS-B messages," *Computers & Security*, vol. 78, pp. 155–173, 2018.

[4] J. Wang, Y. Zou, and J. Ding, "ADS-B spoofing attack detection method based on LSTM," 2020.

[5] R. Karam, M. Salomon, and R. Couturier, "A comparative study of deep learning architectures for detection of anomalous ADS-B messages," in *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, vol. 1, pp. 241–246, IEEE, 2020.

[6] A. Chevrot, A. Vernotte, P. Bernabe, A. Cretin, F. Peureux, and B. Legeard, "Improved testing of AI-based anomaly detection systems using synthetic surveillance data," in *Multidisciplinary Digital Publishing Institute Proceedings*, vol. 59, p. 9, 2020.

[7] S. Akerman, E. Habler, and A. Shabtai, "VizADS-B: Analyzing sequences of ADS-B images using explainable convolutional LSTM encoder-decoder to detect cyber attacks," *arXiv preprint arXiv:1906.07921*, 2019.

[8] T. Li, B. Wang, F. Shang, J. Tian, and K. Cao, "Dynamic temporal ADS-B data attack detection based on sHDP-HMM," *Computers & Security*, vol. 93, p. 101789, 2020.

[9] M. Pusadan, J. Buliali, and R. Ginardi, "Anomaly detection of flight routes through optimal waypoint," in *Journal of Physics: Conference Series*, vol. 801, IOP Publishing, 2017.

[10] X. Olive and L. Basora, "Identifying anomalies in past en-route trajectories with clustering and anomaly detection methods," 2019.

[11] A. Cretin, B. Legeard, F. Peureux, and A. Vernotte, "Increasing the resilience of ATC systems against false data injection attacks using DSL-based testing," in *International Conference on Research in Air Transportation*, 2018.

[12] A. Cretin, A. Vernotte, A. Chevrot, F. Peureux, and B. Legeard, "Test data generation for false data injection attack testing in air traffic surveillance," in *2020 IEEE International Conference on Software Testing, Verification and Validation Workshops (ICSTW)*, pp. 143–152, IEEE, 2020.

[13] M. Schäfer, M. Strohmeier, V. Lenders, I. Martinovic, and M. Wilhelm, "Bringing up OpenSky: A large-scale ADS-B sensor network for research," in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, pp. 83–94, IEEE, 2014.

[14] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *Journal of the ACM (JACM)*, vol. 17, no. 4, pp. 589–602, 1970.

[15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[16] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.