Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

ADS-B anomaly data detection model based on VAE-SVDD

Peng Luo[*], *Buhong Wang*[1,*], *Tengyao Li*, *Jiwei Tian*

Information and Navigation college, Air Force Engineering University, Xi'an, Shanxi Province, China

ARTICLE INFO

Article history:
Received 7 August 2020
Revised 4 December 2020
Accepted 28 January 2021
Available online 2 February 2021

Keywords:
Air traffic surveillance
ADS-B
Anomaly detection
VAE
SVDD

ABSTRACT

As a key technology of the new generation air traffic surveillance system, ADS-B (Automatic Dependent Surveillance-Broadcast) is vulnerable to cyber security challenges because it lacks data integrity and authentication mechanism. For detecting ADS-B data attacks accurately, an anomaly detection model is proposed which fully considers temporal correlations and distribution characteristics of ADS-B data. First, VAE (Variational AutoEncoder) is used to reconstruct ADS-B data so that the reconstructed values can be obtained. Then, for the sake of solving the adaptive problem of anomaly detection threshold, the difference values between the reconstructed values and the actual values are put into SVDD (Support Vector Data Description) for training, and a hypersphere classifier that can detect ADS-B anomaly data is obtained. In addition, in order to prevent overfitting and underfitting, appropriate reconstructed values are selected which can reduce FPR (False Positive Rate) and FNR (False Negative Rate) of anomaly detection. Experiments show that the VAE-SVDD model can detect ADS-B anomaly data which is generated by attacks such as random position deviation and constant position deviation. Moreover, compared with other machine learning methods, this model is not only more adaptable, but also has a lower FPR and FNR.

1. Introduction

ATM (Air traffic management) is an important system which functions as managing airspace and aircraft. It uses communication, navigation and surveillance technology to ensure the safety and order of air traffic. At present, ATM surveillance technology mainly includes PSR (Primary Surveillance Radar), SSR (Secondary Surveillance Radar), WAM (Wide Area Multilateration) and ADS-B. ADS-B is a new generation of surveillance technology which has the advantages of high accuracy, wide range, low cost and supporting information sharing. Thus, ADS-B devices will be equipped in most countries around the world by 2020 (Mccallie et al., 2011; Li et al., 2020).

However, the ADS-B protocol lacked data integrity and authentication mechanism when it was designed. It broadcasts data in plaintext format, which makes ADS-B data vulnerable to various attacks. Attacks such as jamming and message modification suffered by ADS-B data are discussed, also attack difficulty and severity level are analyzed (Strohmeier et al., 2015a). In our previous work, many types of attacks on ADS-B data were analyzed and modeled (Li and Wang, 2019). It shows that USRP (Universal Software Radio Peripheral) device can be used to implement attacks on ADS-B, which proves the sim-

* Corresponding authors.
E-mail addresses: 1939552724@qq.com (P. Luo), wbhgroup@aliyun.com (B. Wang), ltyleader@litengyao.com (T. Li), tianjiwei2016@163.com (J. Tian).

[1] BuHong Wang is the supervisor of Peng Luo. BuHong Wang provides financial support for the paper.

https://doi.org/10.1016/j.cose.2021.102213
0167-4048/

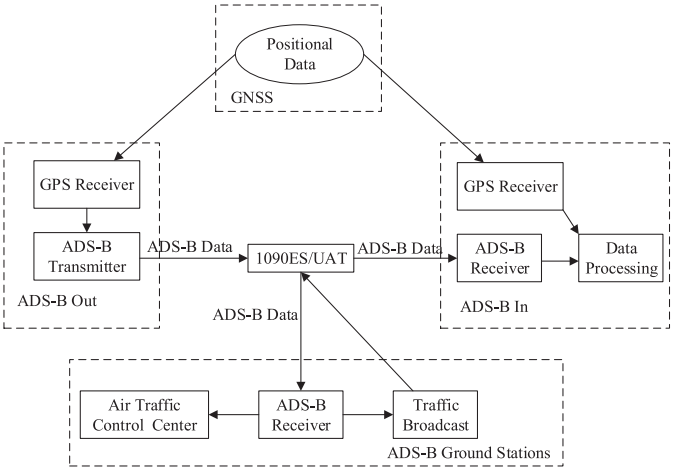

Fig. 1 – ADS-B system overview.

plicity and feasibility of attacks (Costin and Francillon, 2012; Schafer et al., 2013).

Aiming at the security vulnerabilities of ADS-B data, current solutions include encryption methods and non-encryption methods. Encryption methods encrypt ADS-B data or generate additional security information by using symmetric keys (Yang et al., 2018), message authentication code (Kacem et al., 2017), asymmetric keys (Lee et al., 2014), and signatures (Bake et al., 2017). Encryption methods can make ADS-B data secure during transmission. One advantage of encryption methods is that some security protections have been applied to ADS-B data before attackers develop attacks, which can prevent eavesdropping and other attacks (Wesson et al., 2014). However, due to complicated key managements and incompatibility with standard ADS-B protocol, encryption methods are difficult to be applied in practice.

Non-encryption methods check whether ADS-B data is legal by using physical layer information (Strohmeier et al., 2015b), multilateration (Monteiro, 2015) and data fusion (Zhang et al., 2016). Compared with encryption methods, one advantage of non-encryption methods is that standard ADS-B protocol does not need to be modified. Recently, methods based on machine learning are emerging (Li et al., 2019; Habler and Shabtai, 2018). Methods based on machine learning are designed by training historical ADS-B data, and determine the threshold by analyzing reconstruction errors or prediction errors. Methods based on machine learning don't need to add additional ground stations and other surveillance data. However, the threshold determined by manual analyzing reconstruction errors or prediction errors may not be adaptive.

When applying machine learning to ADS-B anomaly data detection, some problems should be paid attention to. First of all, the performance of anomaly detection will be poor if temporal correlations and distribution characteristics of ADS-B data are not fully considered (Li et al., 2019). Secondly, the threshold determined by manual analyzing reconstruction errors or prediction errors may have poor adaptability. Finally, the training process is prone to overfitting and underfitting, which will lead to a high FPR and FNR.

In order to solve the problems above, an anomaly detection model is proposed which can detect ADS-B anomaly data under various attacks. The main contributions of this paper are summarized as following:

(I) VAE is used to reconstruct ADS-B data. Since distribution characteristics of anomaly data are different from that of normal data, VAE is applied to learning distribution characteristics and reconstructing ADS-B data.
(II) In order to solve the problem of poor adaptability of the threshold determined by manual analyzing reconstruction errors, SVDD is used to train the difference values.
(III) In order to make full use of temporal correlations of ADS-B data, the dimension of VAE latent variable is increased so that the latent variable can not only retain ADS-B data at the current moment but also retain ADS-B data at the past and future moment. What is more, the hidden layer of VAE is designed as BiGRU (Bidirectional Gated Recurrent Unit) to make use of temporal correlations.
(IV) Appropriate reconstructed values are selected to reduce FPR and FNR (see section 3.4 and section 4.4.1). The sliding step of the sliding window is set as 1, so each actual value is reconstructed L times (L is the length of the sliding window). The method for selecting the reconstructed value is: after calculating the cosine similarities between the actual value and the L reconstructed values, the reconstructed value which corresponds to the median of the cosine similarities is selected as the final reconstructed value. Appropriate reconstructed values can prevent underfitting and overfitting, thereby reducing FPR and FNR.

The paper is organized as follows. Section 2 introduces the basic knowledge about ADS-B and analyzes the attacks suffered by ADS-B. Section 3 constructs the VAE-SVDD model for detecting ADS-B anomaly data and describes each submodule of the VAE-SVDD model. Section 4 shows the experimental results of the VAE-SVDD model and compares the

Table 1 – ADS-B attack analysis.

ADS-B Data Attack		Implemention Difficulty		
		Low	Medium	High
Severity Level	Low	Eavesdropping	/	/
	Medium	/	Jamming	Message Deletion
	High	/	Message Injection	Message Modification

model with other machine learning methods. Section 5 makes conclusions.

2. Related works

2.1. ADS-B

As a key technology of the new generation air traffic surveillance, ADS-B has the advantages of wide coverage, high accuracy, low cost and support for information sharing. Fig. 1 gives a general overview of the ADS-B system. The aircrafts obtain status information such as position, velocity and heading from GNSS (Global Navigation Satellite System) and other onboard measurement devices. The information is then encapsulated as 112-bit ADS-B data. The aircrafts equipped with ADS-B Out broadcast ADS-B data through the 1090ES (1090 MHz Extended Squitter) or UAT (Universal Access Transceiver) data link. Adjacent aircrafts equipped with ADS-B In can receive broadcast ADS-B data. After ADS-B data is processed, situation awareness is generated and functions including TCAS (Traffic Collision Avoidance System) can be realized. The ground stations receive ADS-B data through the ADS-B receiver. After Air Traffic Control Center processes ADS-B data, it generates the situation awareness of the entire airspace and provides support for air traffic safety and order.

2.2. ADS-B security vulnerabilities

However, ADS-B broadcasts data in plaintext format. Therefore, ADS-B data is vulnerable to eavesdropping, jamming, message injection, message modification, and message deletion. Table 1 shows the difficulty of implementing attacks and the severity level caused by various attacks.

Eavesdropping: ADS-B data can be easily eavesdropped because it is sent in plaintext format. However, eavesdropping will not cause harm directly. It is the basis for other attacks.

Jamming: Jamming is a problem common to all wireless communication. Attackers can implement attacks by sending jamming signals with sufficiently high power on the 1090MHz frequency.

Message Injection: Since the ADS-B system does not have an authentication mechanism, attackers implement message injection attacks by constructing a transmitter that can produce the correct modulated and formatted ADS-B messages. Due to the message injection, a large number of false tracks appear in the air traffic surveillance system, which disrupt the safety and order of air traffic. Therefore, the severity level of message injection is relatively high.

Message Modification: Message modification is implemented by injecting deviations into the actual ADS-B data, which has high severity level. The difficulty of implementing message modification is high because it requires strict time synchronization.

Message Deletion: Message deletion causes the disappearance of the tracks in the air traffic surveillance system. It is quite difficult to implement message deletion because of strict time synchronization. However, message deletion can be easily detected with the support of SSR or WAM. Therefore, the severity level is medium.

2.3. Security solutions to ADS-B data

For ADS-B security vulnerabilities, there are mainly five types of solutions. The five types are encryption methods, physical layer information, multilateration, data fusion and machine learning. Table 2 not only shows the applicable scenarios of various methods, but also analyzes the advantages and disadvantages of various methods.

2.3.1. Encryption methods

For ADS-B security reinforcement, the feasibility of encryption methods is discussed (Strohmeier et al., 2015a). Using symmetric keys to encrypt ADS-B data can avoid message modification with specific purpose (Yang et al., 2018). Another solution for symmetric encryption is to generate MAC(Message Authentication Code) based on ADS-B data (Kacem et al., 2017). One disadvantage of symmetric encryption is that ADS-B data is vulnerable to attacks if symmetric keys are leaked. Asymmetric encryption uses private keys to decrypt ADS-B data, which can prevent key leakage caused by key exchange (Lee et al., 2014). However, aircrafts and ground stations need to save and use public keys of all nodes in the same airspace, which aggravates the communication of ATM. In a word, both symmetric and asymmetric encryption need to modify standard ADS-B protocol, so it is difficult to be applied in practice.

2.3.2. Physical layer information

ADS-B data is attached with some physical layer information in the communication process, which can be used for anomaly detection. Hypothesis testing is used to detect attacks on ADS-B according to the negative correlation between the received signal strength and the distance (Strohmeier et al., 2015b). Strohmeier and Martinovic (2015) proposed a fingerprinting method by analyzing time distributions between two ADS-B messages with specific ADS-B transponders. However, when attackers obtain the prior knowledge of physical layer information through statistical analysis and carefully construct

Table 2 – Comparison of related methods.

Methods	Requiring modifying protocal or additional nodes	Disadvantages and advantages
Encryption	Requires modifying protocal.	Disadvantages: First, ADS-B data is vulnerable to attacks if symmetric keys are leaked. Second, applying for public keys of all nodes in the same airspace aggravates the communication of ATM.
Physical layer information	Relys on ground stations and other entities.	Disadvantages: When attackers obtain the prior knowledge of physical layer information and carefully construct false data injection attack, methods based on physical layer information can hardly detect ADS-B anomaly data.
Multilateration	Requires mutiple ground stations	Disadvantages:Multilateration is mainly deployed on airports and main routes.
Data fusion	Requires SSR receivers	Disadvantages:Methods based on data fusion have limitations and shortcomings owing to the difference of time synchronization and suveillance precision.
Machine learning	Does not require modifying protocal and additional nodes	Advantages:Methods based on machine learning make full use of ADS-B data to build an anomaly detection model which can detect ADS-B anomaly data quickly and accurately.

false data injection attack, methods based on physical layer information can hardly detect ADS-B anomaly data.

2.3.3. *Multilateration*

Comparing the location calculated by TDOA(Time Difference of Arrival) with the demodulated location can identify attacks on ADS-B data. Monteiro(2015) compares the calculated location with the demodulated location, and then uses hypothesis testing to check whether ADS-B data is attacked. Comparing the location calculated by TDOA, AOA (Angle of Arrival), FDOA (Frequency Difference of Arrival) with the demodulated location can judge whether attack behaviors happen (Nijsure et al., 2016). As a backup ATM surveillance technology, multilateration requires multiple ground stations to work together and is mainly deployed on airports and main routes.

2.3.4. *Data fusion*

PSR, SSR, WAM can also be used to check whether ADS-B data is legal. Attack behaviours can be detected when the difference between ADS-B data and other surveillance data beyond the threshold. A novel sensor fusion method with Interacting Multiple Model (IMM) filter to ADS-B, multilateration, and WAM data is proposed, which can improve the reliability of the aircraft position and offer the alternative way to detect attack behaviours (Cho et al., 2013). PHD (Probability hypothesis density) filter is uesd to fuse SSR and ADS-B data, which provides a solution for systematic bias estimation of SSR and shows the novel way to detect ADS-B anomaly data (Zhang et al., 2016). However, methods based on data fusion have limitations and shortcomings owing to the difference of time synchronization and suveillance precision.

2.3.5. *Machine learning*

Methods based machine learning like RNN(Recurrent Neural Network) is often used for anomaly detection of time series (Malhotra et al., 2015; Nanduri and Sherry., 2016). ADS-B data is time-dependent, so it can be regarded as time series with its own characteristics. It is pointed out that using LSTM(Long Short Term Memory) constructs a model by training ADS-B data, which can be used to check the legitimacy of ADS-B data (Li et al., 2019). In addition, the combination of LSTM and Encoder-Decoder can improve the performance of anomaly detection. It shows that the Encoder-Decoder model based on LSTM can reconstruct ADS-B data and detect anomaly data by analyzing reconstruction errors (Habler and Shabtai, 2018; Akerman et al., 2019).It is pointed out that, compared with RNN, using LSTM as the hidden layer of Encoder-Decoder can improve the performance of anomaly detection (Wang et al., 2020). Methods based on Encoder-Decoder have a low FNR and good performance of anomaly detection. However, the threshold determined by manual analyzing reconstruction errors may not be adaptive. In our previous work, a HTM model is built to predict ADS-B data and detect ADS-B anomaly data by analyzing prediction errors (Li et al., 2019). The advantage of the HTM model is predicting and learning at the same time, so the online performance of anomaly detection is better. However, the HTM model has an obvious drawback. HTM uses a small amount of ADS-B data(about 120 ADS-B data) in the past for online prediction, so if the ADS-B data in the past are attacked, the performance of anomaly detection will be greatly reduced. It means that HTM is not suitable for detecting long duration attacks.

2.4. *VAE (variational autoencoder)*

VAE is a generation model for distribution estimation and sample generation (Kingma and Welling, 2014). The general structure of VAE is shown in Fig. 2. It is assumed that the actual input data Y follows an unknown distribution $p_{actu}(Y)$, the reconstructed data $\widehat{Y}$ follows an estimated distribution $p_{recons}(\widehat{Y})$. The goal of VAE is to make the distribution $p_{recons}(\widehat{Y})$ similar to $p_{actu}(Y)$, and make $\widehat{Y}$ similar to Y. In order to achieve this goal, VAE uses Encoder and Decoder to fit two probability density distributions. Encoder fits $q(Z|Y)$, which is an approximate inference of the posterior distribution of the latent variable Z. The target of the approximate inference is a normal distribution with the mean μ and the variance σ^2. Decoder fits

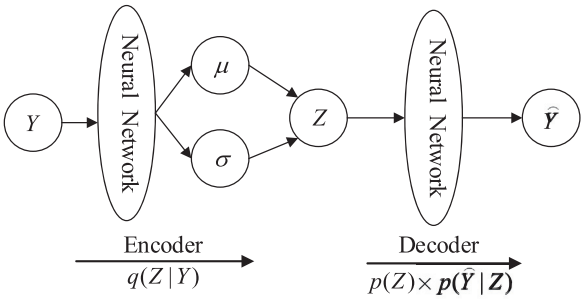

Fig. 2 – VAE network structure.

$p(Z) \times p(\widehat{Y}|Z)$, which generates the probability distribution of the reconstructed data according to the variational probability distribution.

Currently, VAE is being used in the field of anomaly detection. Xu et al. (2018) uses VAE to detect anomaly data for seasonal KPIs in web applications. It is pointed out that VAE allows small noise and randomness of latent variables. Therefore, it can prevent overfitting. Chen et al. (2019) uses the VAE-LSTM method for anomaly detection and robust prediction of time series.

When VAE is used for ADS-B anomaly data detection, two issues should be paid attention to. First, considering temporal correlations of ADS-B data, Encoder and Decoder need to be designed as RNN (Recurrent Neural Network). Second, in order to reduce reconstruction errors, the latent variables not only need to retain ADS-B data at the current moment, but also need to retain ADS-B data at the past and future moment.

3. VAE-SVDD anomaly detection model

3.1. *Problem definition*

A time series $X =< X_1, X_2, \cdots, X_M >$ is defined as the original ADS-B data received by the model, and M is the length of X. $X =< X_1, X_2, \cdots, X_M >$ is a sequence of d-dimensional vectors where $X_k = (x_1, x_2, \cdots, x_d)$ is a d-dimensional vector at time t_k, for $1 \leq k \leq M$. Each dimension represents a feature, which includes latitude, longitude, altitude, velocity, and heading. The general framework of VAE-SVDD model is shown in Fig. 3.

Training phase: First, normal ADS-B time series $X =< X_1, X_2, \cdots, X_M >$ is input. The normalized time series $Y =< Y_1, Y_2, \cdots, Y_M >$ is obtained in the data preprocessing module. Then $Y =< Y_1, Y_2, \cdots, Y_M >$ is reconstructed in the VAE reconstruction module. In the reconstructed values selection module, the appropriate reconstructed value $\widehat{Y}_{k,med}$ ($\widehat{Y}_{k,med}$ is defined in Section 3.4, which represents the reconstructed value corresponding to the median of cosine similarities) is selected by calculating cosine similarities and then the appropriate reconstructed time series $\widehat{Y}_{med} =< \widehat{Y}_{1,med}, \widehat{Y}_{2,med}, \cdots, \widehat{Y}_{M,med} >$ is obtained. In the threshold calculation module, the difference value between the reconstructed value and the actual value $\{D_k | D_k = \widehat{Y}_{k,med} - Y_k\}$ is put into SVDD for training (Tax and

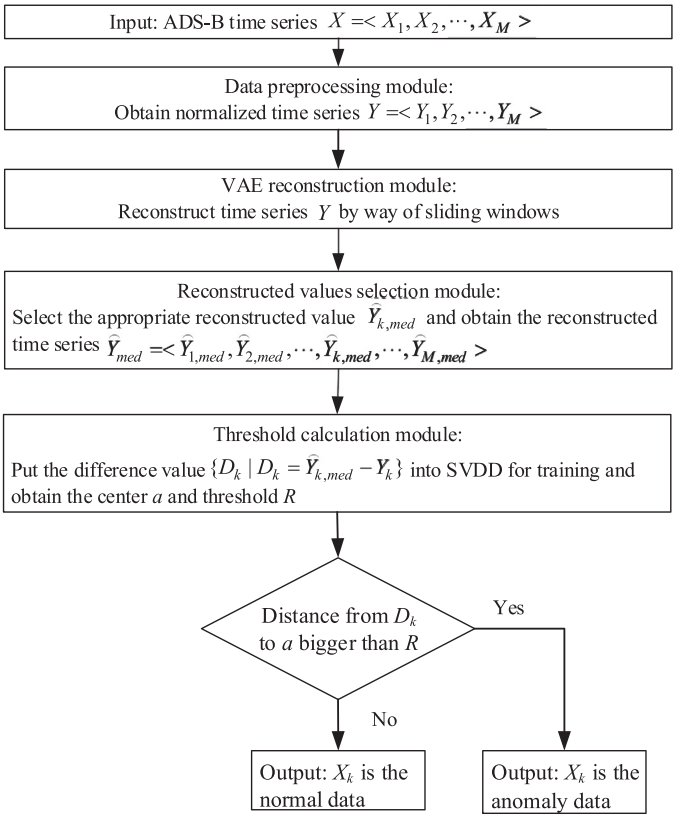

Fig. 3 – VAE-SVDD general framework.

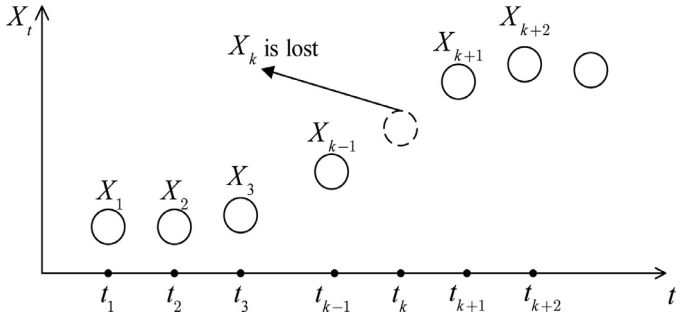

Fig. 4 – Schematic of interpolation method.

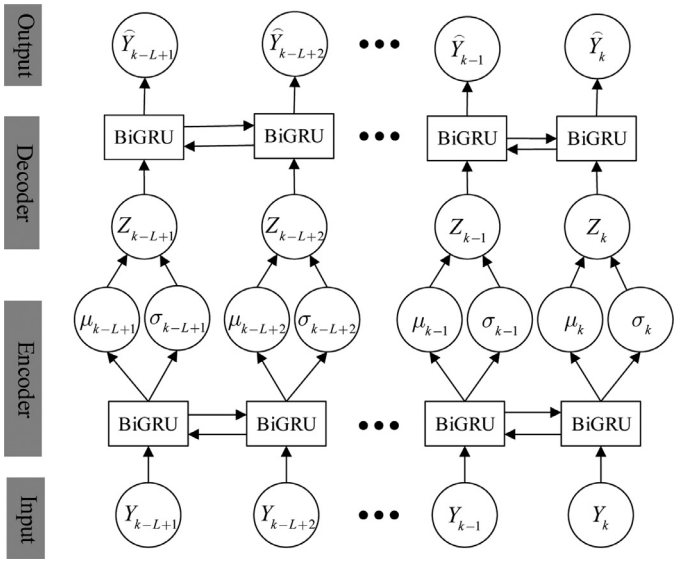

Fig. 5 – VAE reconstruction structure.

Duin, 2004), and then the center a and radius$R$ of SVDD hypersphere are obtained.

Test phase: If the distance from D_k to a is bigger than R, the test sample X_k is anomaly data. If the distance D_k is less than R, the test sample X_k is normal data.

3.2. *Data preprocessing module*

As ADS-B data has a low packet loss rate, it is necessary to preprocess ADS-B data before training. The interpolation method is used to complete the lost ADS-B data. Fig. 4 is a schematic of the interpolation method. If ADS-B data X_k is lost at time t_k, the slope is calculated as following:

$$b = \frac{X_{k+1} - X_{k-1}}{t_{k+1} - t_{k-1}} \tag{1}$$

The lost ADS-B data X_k is completed according to the slope b :

$$X_k = X_{k-1} + b \times (t_k - t_{k-1}) \tag{2}$$

The latitude, longitude, altitude and velocity of ADS-B data are selected as features for training. The time series $H =< H_1, H_2, \cdots, H_k, \cdots, H_M >$ represents a sequence of the selected features where H_k is a four-dimensional vector at time t_k, for $1 \le k \le M$. $H_k^1, H_k^2, H_k^3, H_k^4$ is latitude, longitude, altitude, velocity respectively at time t_k. In order to treat each feature equally and to make the training process converge as soon as possible, the feature sequence $H =< H_1, H_2, \cdots, H_k, \cdots, H_M >$ should be normalized. The normalized equation is shown as following:

$$Y_k^i = \frac{H_k^i - \min(H^i)}{\max(H^i) - \min(H^i)}, 1 \le i \le 4 \tag{3}$$

H_k is the k-th data of the sequence H, the superscript i represents the i-th feature, $\min(H^i)$ is the minimum value of H^i, and $\max(H^i)$ is the maximum value of H^i.

After preprocessing the ADS-B time series $X =< X_1, X_2, \cdots, X_M >$, the normalized sequence $Y =< Y_1, Y_2, \cdots, Y_M >$ is obtained.

3.3. *VAE reconstruction module*

ADS-B data is a time series with multiple features. The features such as latitude, longitude, altitude and velocity at any time are related to the past and future ADS-B data. In order to

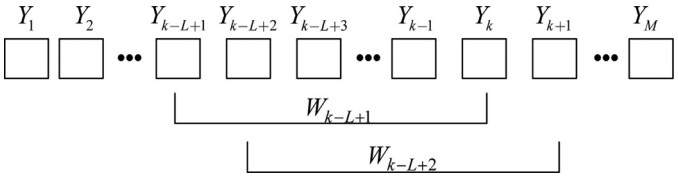

Fig. 6 – Schematic of sliding window.

reconstruct the ADS-B time series, VAE reconstruction structure is designed in Fig. 5.

Input layer: The input layer inputs sequence Y by the sliding window mechanism. As shown in Fig. 6, the sliding window $W_{k-L+1} =< Y_{k-L+1}, Y_{k-L+2}, \cdots, Y_k >$ is defined as the sequence on the interval $[t_{k-L+1}, t_k]$, and the length of the sliding window is L. The sliding step is set to 1, so the next sliding window $W_{k-L+2} =< Y_{k-L+2}, Y_{k-L+3}, \cdots, Y_{k+1} >$ represents the sequence on the interval $[t_{k-L+2}, t_{k+1}]$. In the same way, $W_{k-L+3} =< Y_{k-L+3}, Y_{k-L+4}, \cdots, Y_{k+2} >$ represents the sequence on the interval $[t_{k-L+3}, t_{k+2}]$. Therefore, Y_k is reconstructed L times in total from the sliding window W_{k-L+1} to the sliding window W_k, which is represented as $\{\widehat{Y}_{k,k-L+1}, \widehat{Y}_{k,k-L+2}, \cdots, \widehat{Y}_{k,k}\}$. To be specific, $\widehat{Y}_{k,k-L+1}$ is the reconstructed value of Y_k in the sliding window W_{k-L+1}, $\widehat{Y}_{k,k-L+2}$ is the reconstructed value of Y_k in the sliding window W_{k-L+2}, and $\widehat{Y}_{k,k}$ is the reconstructed value of Y_k in the sliding window W_k.

Encoder layer: The Encoder layer uses BiGRU (Bidirectional Gated Recurrent Unit) to capture temporal correlations of $Y =< Y_1, Y_2, \cdots, Y_M >$. BiGRU is the RNN (Recurrent Neural Network) that adds the bidirection mechanism on the basis of GRU (Chung et al., 2014). Compared with GRU, BiGRU has the advantage of not only considering the past ADS-B data but also the future ADS-B data. The function realized by BiGRU is represented as $q(Z|Y)$, and the number of BiGRU neurons is set to 20. It is assumed that the posterior distribution of the latent variable Z satisfies the equation $p(Z|Y) = N(\mu, \sigma^2)$. The goal of Encoder is to make $q(Z|Y)$ similar to $p(Z|Y)$.

Decoder layer: The function realized by the Decoder layer is represented as $p(Z) \times P(\widehat{Y}|Z)$. To be specific, take a sample ε_1 from $p(Z|Y)$ first, and then fit $P(\widehat{Y}|Z)$ by BiGRU. The number of BiGRU neurons is set to 20. Sampling ε_1 from $p(Z|Y)$ is equivalent to sampling ε_2 from the standard normal distribution $N(0,1)$ since the posterior distribution of the latent variable Z satisfies the equation $p(Z|Y) = N(\mu, \sigma^2)$. Therefore, ε_1 satisfies the equation as following:

$$\varepsilon_1 = \mu + \varepsilon_2 \times \sigma \tag{4}$$

In this way, the sampling process of $p(Z|Y)$ becomes sampling from $N(0,1)$. After the sampling process is completed, BiGRU is used to fit $P(\widehat{Y}|Z)$.

Output layer: There is no need to use latent variables to generate samples, so mean square error $MSE(\widehat{Y}, Y)$ is selected as the loss function *loss*.

3.4. *Reconstructed values selection module*

Due to different distribution characteristics of the data in different sliding windows, the reconstruction errors of different sliding windows are also different. As a result, the numerical values of the elements in $\{\widehat{Y}_{k,k-L+1}, \widehat{Y}_{k,k-L+2}, \cdots, \widehat{Y}_{k,k}\}$ are different. Calculate the cosine similarity between Y_k and each element in $\{\widehat{Y}_{k,k-L+1}, \widehat{Y}_{k,k-L+2}, \cdots, \widehat{Y}_{k,k}\}$ as following:

$$\cos(Y_k, \widehat{Y}_{k,i}) = \frac{\sum_{j=1}^{4} (Y_k^j \times \widehat{Y}_{k,i}^j)}{\sqrt{\sum_{j=1}^{4} (Y_k^j)^2} \times \sqrt{\sum_{j=1}^{4} (\widehat{Y}_{k,i}^j)^2}}, k - L + 1 \le i \le k, 1 \le j \le 4 \tag{5}$$

In Eq. (5), the superscript *j* represents the j-th dimension of Y_k, and the four dimensions of Y_k are latitude, longitude, altitude and velocity.

The minimum value, the maximum value, and the median value in$\{\cos(Y_k, \widehat{Y}_{k,i}) | k - L + 1 \le i \le k\}$ is recorded as $\cos(Y_k, \widehat{Y}_{k,\min})$, $\cos(Y_k, \widehat{Y}_{k,\max})$ and $\cos(Y_k, \widehat{Y}_{k,med})$ respectively. As a result, $\widehat{Y}_{k,med}$ is selected as the reconstructed value of Y_k. The reasons include the following two aspects. (1) If $\widehat{Y}_{k,\min}$ is selected as the reconstructed value of Y_k, it is easy to cause the phenomenon of underfitting. As the aircraft has maneuvering states such as climb, turning and descent, it is easy to result in high FPR of ADS-B normal data. (2) If $\widehat{Y}_{k,\max}$ is selected as the reconstructed value of Y_k, it is easy to cause overfitting which results in high FNR of ADS-B anomaly data. Therefore, selecting $\widehat{Y}_{k,med}$ as the reconstructed value can prevent overfitting and underfitting.

In the reconstructed values selection module, a appropriate reconstructed value $\widehat{Y}_{k,med}$ is selected, and then the reconstructed sequence $\widehat{Y}_{med} =< \widehat{Y}_{1,med}, \widehat{Y}_{2,med}, \cdots, \widehat{Y}_{M,med} >$ of the sequence $Y =< Y_1, Y_2, \cdots, Y_M >$ is obtained.

3.5. *Threshold calculation module*

SVDD (Support Vector Data Description) is an unsupervised machine learning method that can solve one-class classification problems (Tax and Duin, 2004). SVDD aims to find a hypersphere that contains as many training samples as possible. Aiming at ADS-B anomaly data detection, the difference values $D = \{D_k = \widehat{Y}_{k,med} - Y_k | 1 \le k \le M\}$ are regarded as training samples, and then the difference values $D = \{D_k = \widehat{Y}_{k,med} - Y_k | 1 \le k \le M\}$ are put into the SVDD. As a result, the threshold of ADS-B anomaly data detection can be obtained.

Table 3 – Results of sample classification.

Actual Value/Sample	Classification result	
	Reconstructed Normal Value/Sample	Reconstructed Anomaly Value/Sample
Actual Normal Value/Sample	TP	FN
Actual Anomaly Value/Sample	FP	TN

SVDD can be expressed as solving the following optimization problems:

$$\min F(R, a) = R^2 + C\sum_{i=1}^{n}\xi_k$$

$$\text{s.t.}\begin{cases}\|D_k - a\|^2 \le R^2 + \xi_k, (k = 1, 2, \cdots, M)\\ \xi_k \ge 0\end{cases} \tag{6}$$

In Eq. (6), a is the center of the SVDD hypersphere. R is the radius of the hypersphere, and R is also the threshold. ξ_k is the slack variable which is used to measure the few anomaly data outside the hypersphere. C is the penalty coefficient which is used to control the volume of the hypersphere, and the value of C is usually 1. Use the Lagrangian multiplier method to solve this optimization problem and get the Lagrangian function as following:

$$L(R, a, \xi_k) = R^2 + C\sum_{k=1}^{M}\xi_k - \sum_{k=1}^{M}\lambda_k\left[R^2 + \xi_k - \|D_k - a\|^2\right] - \sum_{k=1}^{M}\beta_k\xi_k \tag{7}$$

In Eq. (7), λ_k and β_k are Lagrange multipliers. The distance from D_k to a is recorded as $g(D_k)$, and $g(D_k)$ can be calculated as following:

$$g(D_k) = \|D_k - a\| = \sqrt{(D_k, D_k) - 2\sum_{i=1}^{M}\lambda_i(D_i, D_k) + \sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j(D_i, D_j)} \tag{8}$$

The threshold R can be calculated as following:

$$R = \sqrt{(D_s, D_s) - 2\sum_{i=1}^{M}\lambda_i(D_i, D_s) + \sum_{i=1}^{M}\sum_{j=1}^{M}\lambda_i\lambda_j(D_i, D_j)} \tag{9}$$

In Eq. (9), D_s is a support vector, which means that D_s is on the sphere of the hypersphere.

Therefore, if $g(D_k) > R$, it indicates that the distance from D_k to a is bigger than R, and then X_k is ADS-B anomaly data. If $g(D_k) < R$, it indicates that the distance from D_k to a is smaller than R, and then X_k is ADS-B normal data.

In order to make the training samples linearly separable, a kernel function is needed to map the samples from the original space to a higher-dimensional feature space (Ali and Smith, 2005). The commonly RBF (Radial Basis Function) kernel function is used as following:

$$K_{RBF} = \exp(-\gamma \times \left\|D_i, D_j\right\|^2), \quad \gamma > 0 \tag{10}$$

In Eq. (10), D_i and D_j are two samples of SVDD. γ is the kernel parameter, and γ is usually set as the reciprocal of the feature number. The four features are latitude, longitude, altitude and velocity. Thus the value of γ is 0.25.

3.6. *Evaluation index*

Table 3 describes the results of sample classification. TP (True Positive) refers to the number of samples that are actual normal values and are correctly classified as normal samples. FN (False Negative) refers to the number of samples that are actual normal values and are incorrectly classified as anomaly samples. FP (False Positive) refers to the number of samples that are actual anomaly values and are incorrectly classified as normal samples. TN (True Negative) refers to the number of samples that are actual anomaly values and are correctly classified as anomaly samples.

FPR (False Positive Rate), FNR (False Negative Rate), ER (Error Rate), F1_score are used as the evaluation index of the anomaly detection model. FPR, FNR, ER, F1_score are defined as following:

$$\begin{cases}\text{FPR} = \frac{\text{FN}}{\text{TP+FN}}\\ \text{FNR} = \frac{\text{FP}}{\text{FP+TN}}\\ \text{ER} = \frac{\text{FN+FP}}{\text{TP+FN+FP+TN}}\\ \text{F1\_score} = \frac{2\text{TP}}{2\text{TP+FP+FN}}\end{cases} \tag{11}$$

4. Experimental analysis

4.1. *Data collection*

For the experiment, 50 flights ADS-B data is extracted from datasets of OPENSKY project as the training samples, and 20 flights ADS-B data is extracted as the test samples (Strohmeier et al., 2015c). Each flight contains ADS-B data from 200 to 1500. Considering the limitations of the actual environments, ADS-B attack data is difficult to obtain. ADS-B attack data is generated by artificial construction in the test samples. Considering the stealth and effectiveness, five common types of attacks are constructed based on original ADS-B data which is shown in Table 4.

Therefore, five types of attack patterns are modeled which are based on the actual ADS-B data: constant position deviation attack, random position deviation attack, velocity drift attack, DOS attack and flight replacement attack.

A sample flight containing 360 ADS-B data is selected randomly for illustration. Fig. 7 describes the track (latitude, longitude and altitude) of this example flight, and Fig. 8 describes

Table 4 – Experimental dataset.

Data attack	Data construction methods
Constant position deviation attack	Inject the constant 1 into the actual latitude and longitude.
Random position deviation attack	Inject Gaussian noise with a mean value 0 and a variance 0.5 into the actual latitude and longitude.
Velocity drift attack	Enlarge the velocity with a multiple of 2m/s gradually.
DOS attack	The aircraft track disappears in the air traffic surveillance system.
Flight replacement attack	Replace the actual flight track with another false flight track.

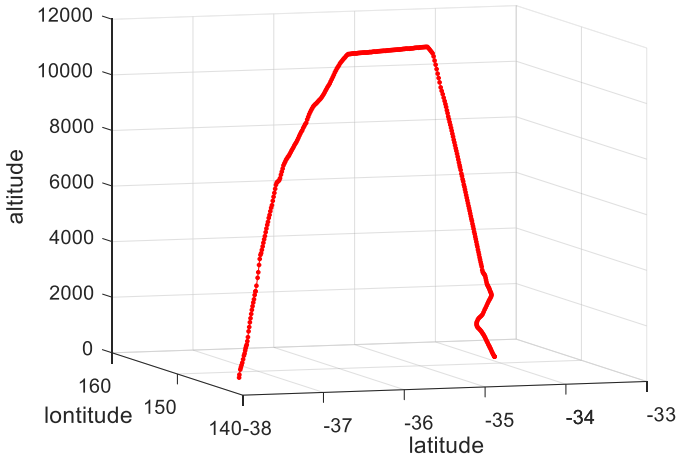

Fig. 7 – Flight track.

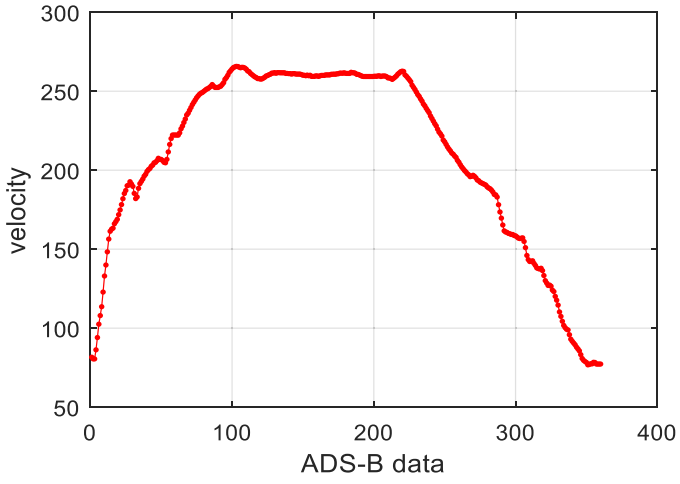

Fig. 8 – Flight velocity.

the velocity of this example flight. In order to verify the detection performance of ADS-B anomaly data and to illustrate that the model doesn't misreport the normal ADS-B data as anomaly data, the first 160 ADS-B data isn't injected with attack data and the last 200 ADS-B data is injected with attack data. The first 160 ADS-B data contains the aircraft's take-off, climb, turning and cruise phase. The last 200 ADS-B data contains the aircraft's cruise, turning and descent phase.

(1) **Constant position deviation attack** (Message modification): The attacker implements the constant position deviation attack by adding a constant deviation vector to the actual ADS-B data. As shown in Fig. 9, this attack results in the track deviating to a specific direction. The first 160 ADS-B data is not tampered. For the last 200 ADS-B data, the constant 1 is added to the actual latitude and longitude.

(2) **Random position deviation attack** (Message modification): The attacker implements the constant position deviation attack by adding the Gaussian noise to the actual ADS-B data. As shown in Fig. 10, this attack results in the track fluctuating randomly around the actual track. The first 160 ADS-B data is not tampered. For the last 200 ADS-B data, Gaussian noise with a mean value of 0 and a variance of 0.5 is added to the longitude and latitude.

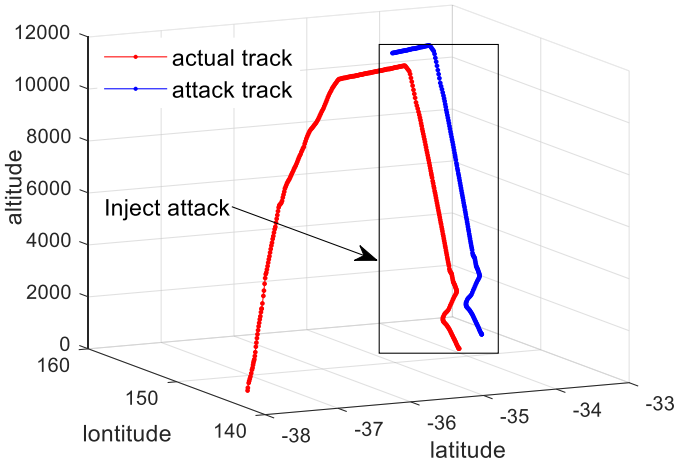

Fig. 9 – Constant position deviation.

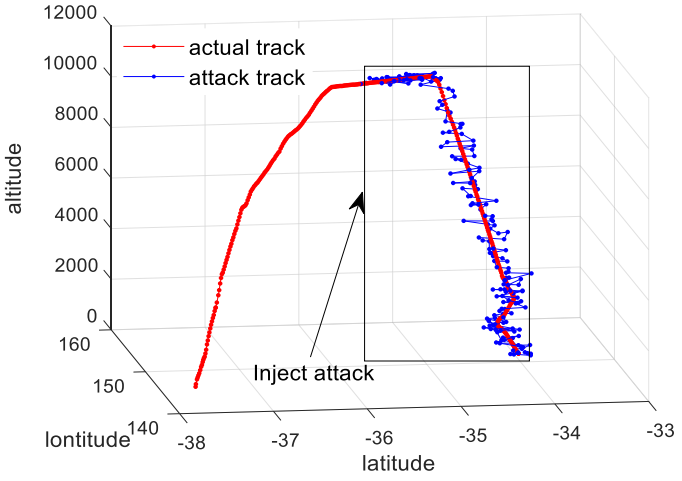

Fig. 10 – Random position deviation.

(3) **Velocity drift attack** (Message modification): The attacker implements the velocity drift attack by gradually changing the value of the velocity in the actual ADS-B data. As shown in Fig. 11, this attack results in velocity drifting away from the actual velocity, and the drift gradually increases. The first 160 ADS-B data is not tampered. For the last 200 ADS-B data, the velocity changes with a multiple of 2m/s. Specifically, the velocity of 161-th ADS-B data increases by 2m/s, the velocity of 162-th ADS-B data increases by 4m/s, the velocity of 163-th ADS-B data increases by 6m/s, and so on.

(4) **DOS attack** (Message deletion): The attacker implements the DOS attack at the physical layer through constructive or destructive interference (Strohmeier et al., 2015a). As shown in Fig. 12, the ADS-B data can not be surveilled because of DOS attack. The first 160 ADS-B data is not tampered. For the last 200 ADS-B data, the aircraft track disappears in the air traffic surveillance system due to DOS attack.

(5) **Flight replacement attack** (Message injection): The attacker implements the flight replacement by replacing the actual flight track with another false flight track, which results in target missing and mistaken airspace intelligence. As shown in Fig. 13, the first 160 ADS-B data is not injected with false flight track. For the last 200 ADS-B data, the false flight track injected by the attacker also contains the cruise, turning and descent phase. The false flight track contains 200 ADS-B data.

4.2. *Dimension of latent variable*

When VAE is applied to fields such as image generation, the dimension of latent variable is generally smaller than that of the input images due to a lot of redundant information. However, the latitude, longitude, altitude and velocity information of ADS-B data are independent of each other, and there is almost no redundant information. Since the sequence $Y =< Y_1, Y_2, \cdots Y_k, \cdots, Y_M >$ is the time series, the VAE reconstruction module must make full use of temporal correlations of the sequence $Y =< Y_1, Y_2, \cdots Y_k, \cdots, Y_M >$ to have a better reconstruction effect. It is proposed to increase the dimension of the latent variable Z_k. The purpose is to make Z_k not only retain the latitude, longitude, altitude and velocity feature information at the current moment, but also retain the feature

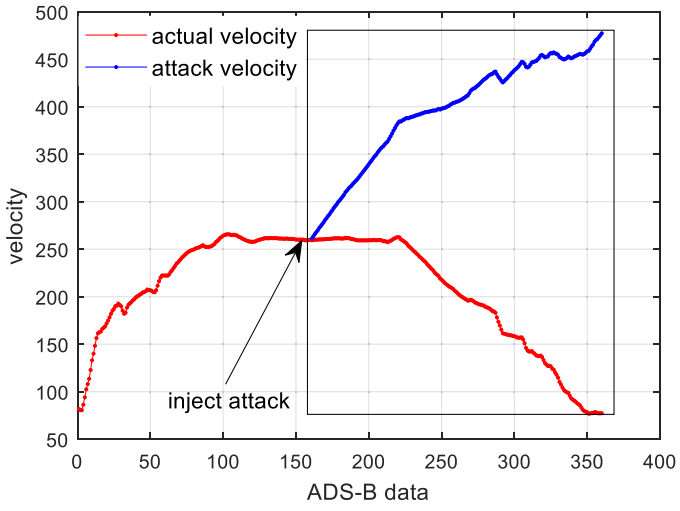

Fig. 11 – Velocity draft attack.

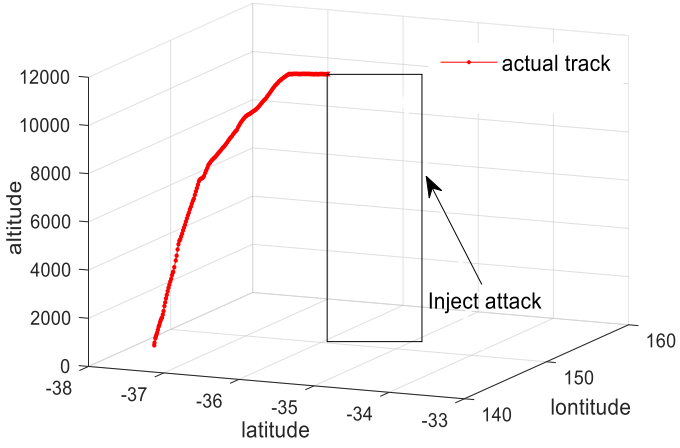

Fig. 12 – DOS attack.

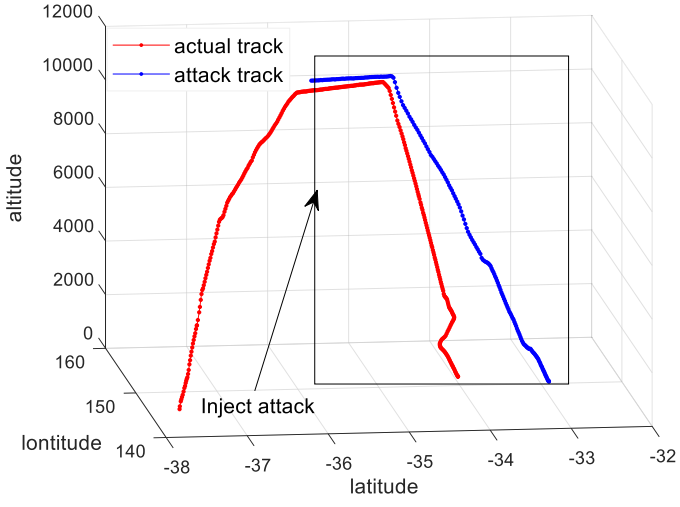

Fig. 13 – Flight replacement.

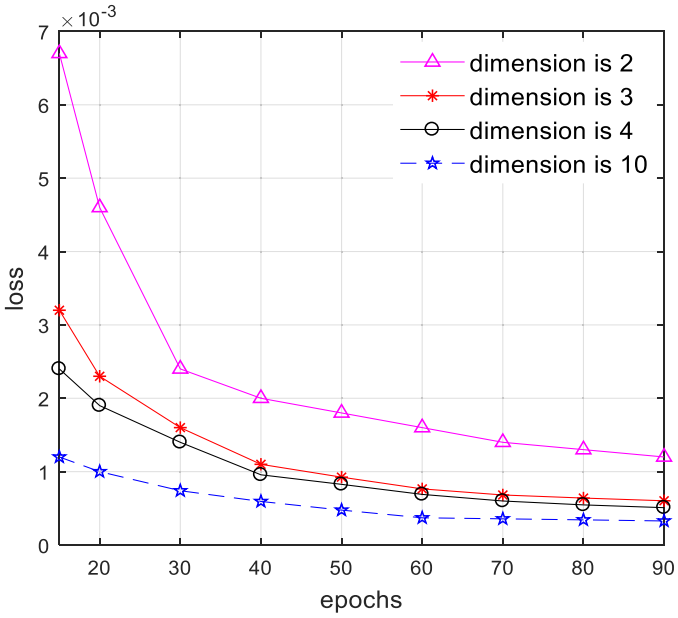

Fig. 14 – Dimension of latent variable.

information of past and future moments. As a result, the VAE reconstruction module has a better reconstruction effect. For different dimensions of latent variable Z_k, Fig. 14 shows the relation between the loss function *loss* and training times *epochs*. The following conclusions are drawn by analyzing Fig. 14:

(1) When the dimension of the latent variable Z_k is 2 or 3, the loss function *loss* is relatively large. The input data $Y =< Y_1, Y_2, \cdots Y_k, \cdots, Y_M >$ has 4 features. When the dimension of Z_k is less than 4, the effect of VAE can be understood as dimensionality reduction. However, the four features of latitude, longitude, altitude and velocity are independent of each other, and there is no redundant information. Therefore, dimensionality reduction will lose too much feature information, resulting in a bad reconstruction effect.
(2) When the dimension of the latent variable Z_k is 4, the loss function *loss* is still relatively large. Since the four features of the input data $Y =< Y_1, Y_2, \cdots Y_k, \cdots, Y_M >$ are independent of each other, Z_k can only retain all the feature information of Y_k at the current moment, but cannot retain the feature information at past moments (Y_{k-1},Y_{k-2},Y_{k-3}, et al.) and future moments (Y_{k+1},Y_{k+2},Y_{k+3}, et al.). Therefore, the reconstruction effect is not good.
(3) When the dimension of the latent variable is 10, the loss function *loss* is relatively small. When the dimension is greater than 4, Z_k can not only retain feature information of Y_k at the current moment, but also retain feature information at past moments (Y_{k-1},Y_{k-2},Y_{k-3}, et al.) and future moments (Y_{k+1},Y_{k+2},Y_{k+3}, et al.). Therefore, the reconstruction effect is relatively good. We also conduct experiments where the dimension is 20 and 30 respectively, and the results show that: compared with the dimension 10, the reduction effect of the loss function *loss* is not obvious, but the training time is still increasing rapidly. Therefore, the selected dimension is 10.

4.3. *Experimental results*

In order to obtain a good reconstruction effect and make training time short, it is necessary to select an appropriate *epochs*. If *epochs* is too small, the reconstruction effect will be poor. If *epochs* is too large, training time will be too long. Fig. 15 shows the curves of the loss function *loss* and training time varying with *epochs*. Training time refers to the time that it takes to train one flight (the average number of ADS-B data contained in one flight is about 500). When *epochs* is beyond 60, the reduction effect of the loss function *loss* is not obvious, but training time is still increasing rapidly. Therefore, the selection of *epochs* as 60 can ensure the optimal reconstruction effect and training time.

Similarly, under the condition of $batch\_size = 128$ and length of the sliding window $L = 15$, training time is relatively short and VAE reconstruction module has a good reconstruction effect. The experimental results of anomaly detection for the above five attacks are shown from Fig. 16 to 20.

The result of detecting the constant position deviation attack is depicted in Fig. 16. In the first 160 test samples, there are 11 samples whose distance to the center a is greater than R, thus FPR is 6.875%. In the last 200 test samples, there are 13 samples whose distance to the center a is less than R, thus FNR is 6.5%. In the 360 test samples, 24 samples are misclassified, thus ER is 6.67% and F1_score is 92.55% .

It shows that FPR, FNR, ER, F1_score of random position deviation attack is 3.75%, 8%, 6.11%, 93.33% respectively in Fig. 17. It shows that FPR, FNR, ER, F1_score of velocity draft attack is 3.75%, 7.5%, 5.833%, 93.62% respectively in Fig. 18. It shows that FPR, FNR, ER, F1_score of DOS attack is 8.75%, 0%, 3.89%, 95.42% respectively in Fig. 19. It shows that FPR, FNR, ER, F1_score of flight replacement attack is 6.25%, 13.5%, 10.28%, 89.02% respectively in Fig. 20.

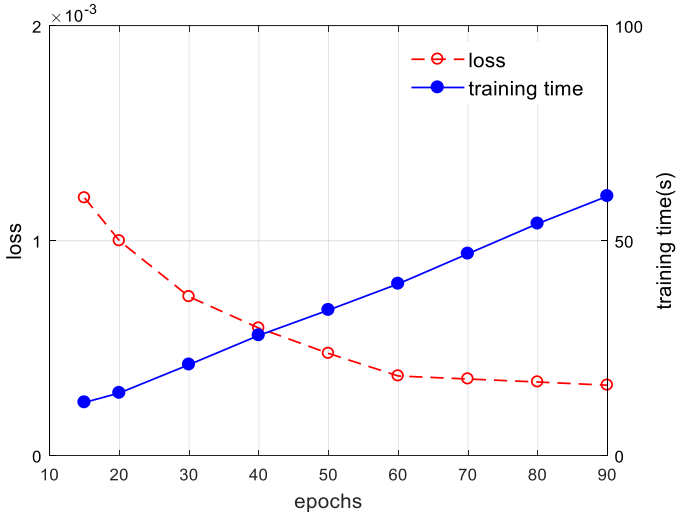

Fig. 15 – Selection of *epochs*.

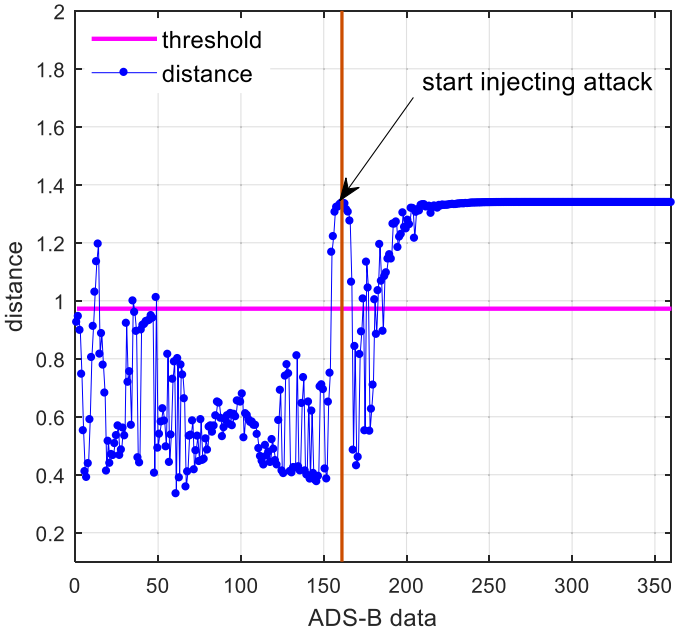

Fig. 16 – Detection of constant position deviation.

20 test flights are tested for anomaly detection, and the average value of FPR, FNR and ER of the 20 experiments are taken as the experimental result. Table 5 shows the experimental results of anomaly detection. It can be seen from Table 5: (1) For the five types of attacks, ER is less than 10% and F1_score is greater than 90%. (2) FNR of DOS attack is the lowest, which is only 0.58%. This is because DOS attack causes the track to disappear, and it is the easiest to detect the DOS attack objectively. (3) FNR of flight replacement attack reaches 11.29%, which is relatively high. This is because the attacker masters the flight rules and replaces the actual track with the false track that also contains cruise, turning and descent phase. Flight replacement attack is highly concealed and relatively difficult to detect. (4) FNR of velocity draft attack reaches 8.85%, which is relatively high. This is because the initial velocity draft caused by the attack is not obvious enough. Objectively speaking, the initial velocity draft is not easy to detect.

4.4. *Comparative experiment*

4.4.1. *Reconstructed values selection*

As shown in Eq. (5), the minimum value, the minimum value, the median in $\{\cos(Y_k, \widehat{Y}_{k,i}) | k - L + 1 \leq i \leq k\}$ is $\cos(Y_k, \widehat{Y}_{k,\min})$,$\cos(Y_k, \widehat{Y}_{k,\max})$,$\cos(Y_k, \widehat{Y}_{k,med})$ respectively. Selecting $\widehat{Y}_{k,med}$ as the reconstructed value of Y_k can prevent underfitting and overfitting, which can reduce FPR and FNR.

Fig. 21 shows the result of detecting constant position deviation attack when $\widehat{Y}_{k,\min}$ is the reconstructed value of Y_k

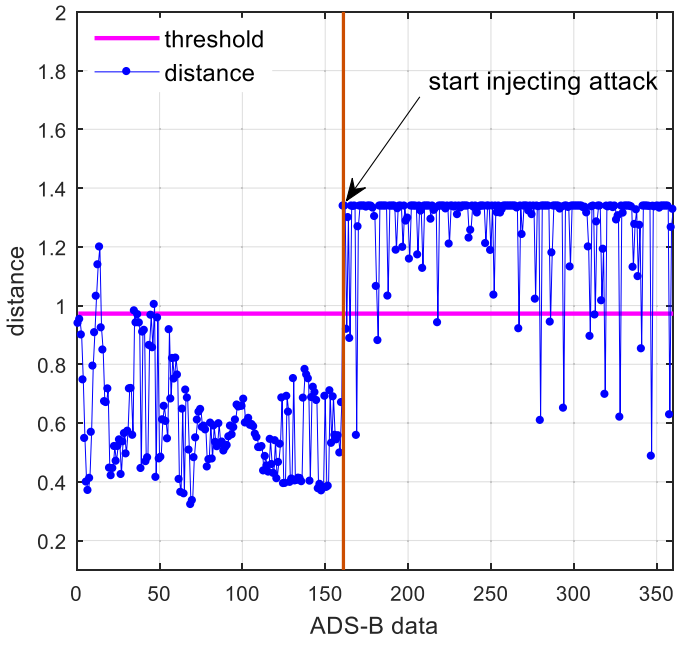

Fig. 17 – Detection of random position deviation.

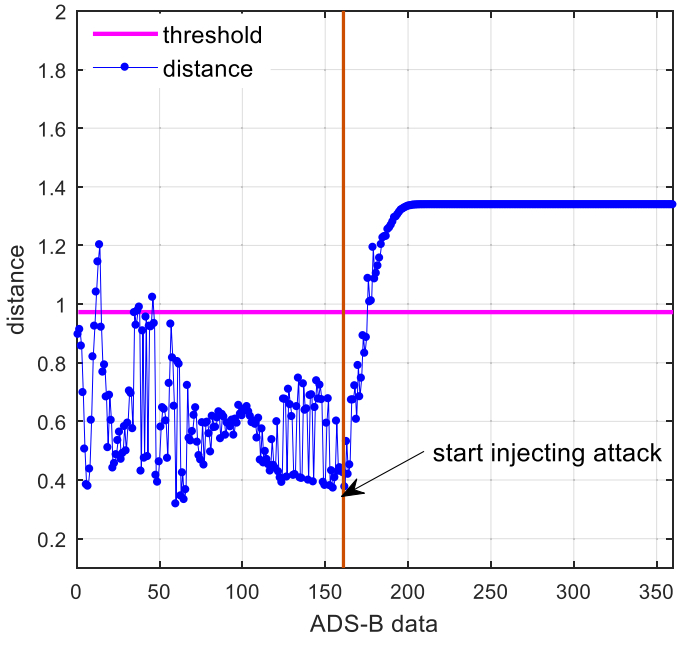

Fig. 18 – Detection of velocity draft attack.

Table 5 – Experimental result of anomaly detection(%).

Evaluation index	Constant position deviation attack	Random position deviation attack	Velocity draft	DOS attack	Flight replacement
FPR	6.39	4.93	3.44	7.32	6.37
FNR	8.23	7.70	8.85	0.58	11.29
ER	7.41	6.47	6.44	3.57	9.10
F1_socre	91.82	92.89	93.01	95.84	90.14

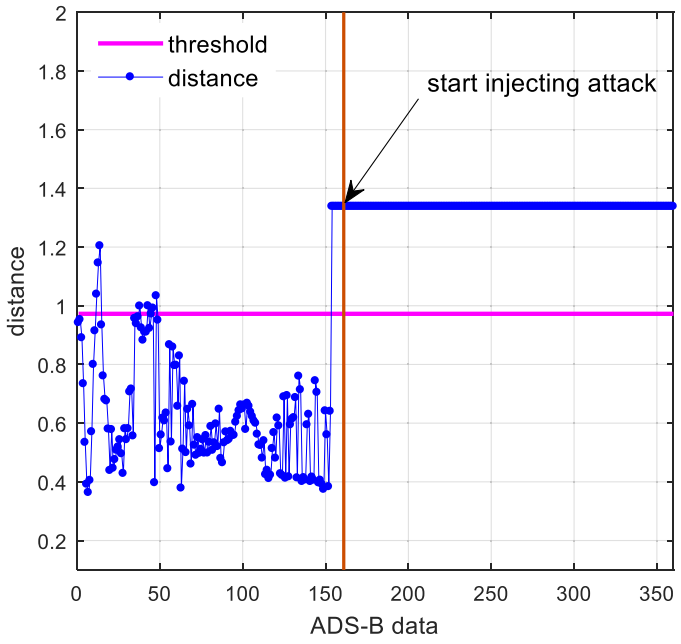

Fig. 19 – Detection of DOS attack.

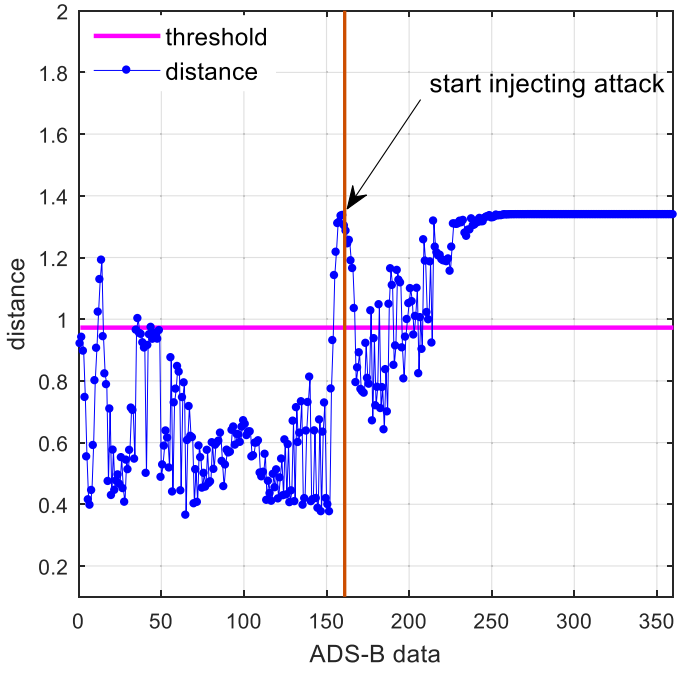

Fig. 20 – Detection of flight replacement.

(The attack method is consistent with Section 4.1. The first 160 ADS-B data is not attacked, and the constant 1 is added to the latitude and longitude of the last 200 ADS-B data.). The results show that in the first 160 test samples, there are 24 samples whose distance to the center a is greater than R, thus FPR is 15%. In the last 200 test samples, there are 6 samples whose distance to the center a is less than R, thus FNR is 3%. FPR of normal data is high. This is because selecting $\widehat{Y}_{k,min}$ as the reconstructed value of Y_k leads to underfitting. Therefore, normal samples with high mobility are easily misreported as anomaly data. As a comparison, when $\widehat{Y}_{k,med}$ is selected as the reconstructed value, FPR of constant position deviation attack is 6.39%, which is reduced.

Fig. 22 shows the result of detecting constant position deviation attack when $\widehat{Y}_{k,max}$ is the reconstructed value of Y_k. The

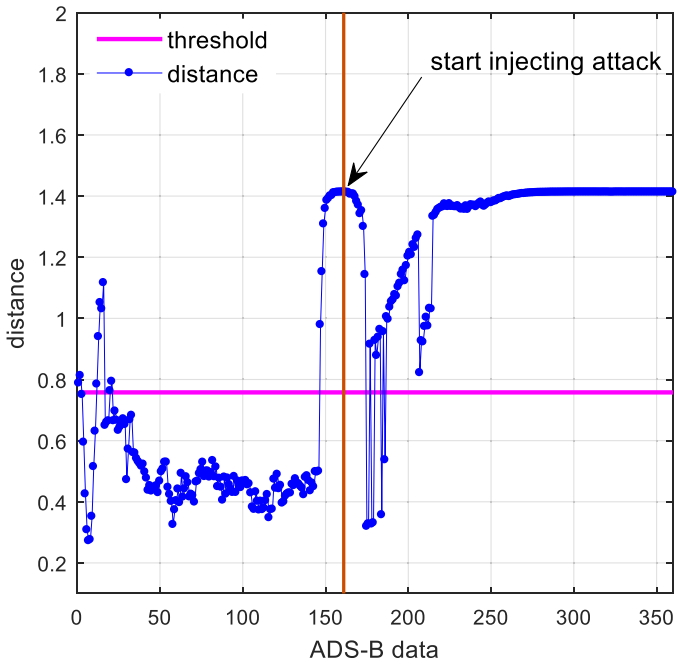

Fig. 21 – Detection of constant attack(underfitting).

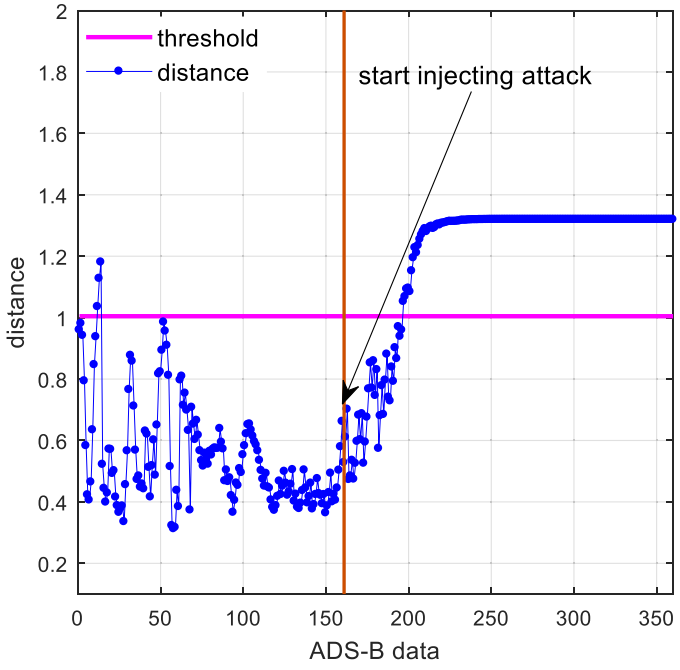

Fig. 22 – Detection of constant attack(overfitting).

results show that in the first 160 test samples, there are 3 samples whose distance to the center a is greater than R, thus FPR is 1.875%. In the last 200 test samples, there are 36 samples whose distance to the center a is less than R, thus FNR is 18%. FNR of anomaly data is high. This is because selecting $\widehat{Y}_{k,max}$ as the reconstructed value of Y_k leads to overfitting. As a comparison, when $\widehat{Y}_{k,med}$ is selected as the reconstructed value, FNR of constant position deviation attack is 8.23%, which is reduced.

4.4.2. *Compared with other machine learning methods*

For comparison, other common unsupervised machine learning methods are selected, which are also used for ADS-B anomaly data detection. These methods include IForest (Liu et al., 2012), the neural network composed of GRU, the neural network composed of LSTM, LSTM-Encoder-Decoder (Habler and Shabtai, 2018).

Table 6 – Comparison of machine learning methods(%).

	Evaluation index	IForest	GRU	LSTM	LSTM-Encoder-Decoder	VAE-SVDD
Constant position deviation attack	FPR	11.26	6.71	6.92	7.32	6.39
	FNR	45.58	84.57	82.10	11.33	8.23
	ER	30.27	49.96	48.68	9.54	7.41
	F1_score	72.23	62.40	61.90	89.61	91.82
Random position deviation attack	FPR	11.26	6.71	6.92	7.32	4.93
	FNR	71.50	26.59	26.74	11.83	7.70
	ER	44.72	17.59	17.92	9.82	6.47
	F1_score	63.82	82.37	80.81	89.34	92.89
Velocity draft	FPR	11.26	6.71	6.92	7.32	3.44
	FNR	53.52	15.60	14.09	8.07	8.85
	ER	29.73	11.65	10.90	7.73	6.44
	F1_score	70.46	87.68	86.87	91.42	93.01
DOS attack	FPR	11.26	6.71	6.92	7.32	7.32
	FNR	0.83	0.62	0.75	0.64	0.58
	ER	34.68	3.32	3.49	3.61	3.57
	F1_score	93.52	96.12	94.34	95.80	95.84
Flight replacement	FPR	11.26	6.71	6.92	7.32	6.37
	FNR	34.44	19.38	17.54	18.51	11.29
	ER	23.53	13.75	12.82	13.53	9.10
	F1_score	81.35	85.78	85.13	85.89	90.14

Table 7 – Average detection time.

Evaluation index	IForest	GRU	LSTM	LSTM-Encoder-Decoder	VAE-SVDD
Detection time(ms)	4.8	11.1	13.9	36.1	46.3

As a traditional machine learning method, IForest is mainly used for unsupervised anomaly detection. The trees are set to 200, and the sampling size of each tree is set to 256.

As improved versions of RNN(Recurrent Neural Network), GRU and LSTM are mainly used for time series prediction and anomaly detection. The neural network is composed of three layers of GRU/LSTM, one Dropout layer and one fully connected layer. The loss function is MSE (Mean Square Error). Use the cosine similarity to determine the threshold (The threshold selection method is depicted as following: After calculate the cosine similarity between the ADS-B predicted value and the actual value, sort the cosine similarity from small to large. The 6% smallest cosine similarity is selected as the threshold).

Network structure of Lstm-Encoder-Decoder is the same as that of the paper written by Habler and Shabtai, 2018. LSTM is taken as the hidden layer, the sliding window length is set as 15, and the loss function *loss* is mean square error (MSE). The threshold is selected as follows: First, the anomaly score is calculated according to cosine similarity. Second, the anomaly score is sorted from the smallest to the largest. Third, the 95% largest anomaly score is selected as the threshold.

Table 6 shows FPR, FNR and ER of various machine learning methods. In order to compare the complexity of various methods, Table 7 shows the average time to detect 100 ADS-B test samples.The following conclusions can be drawn by analyzing Tables 6 and 7.

(1) FPR, FNR and ER of IForest are high. This is because the latitude, longitude, altitude, and velocity features of ADS-B data vary with time. However, IForest only considers these features at the current moment, and does not consider temporal correlations of the ADS-B data. In terms of detection time, IForest has the advantage of the shortest detection time. It only takes 4.8ms to detect 100 ADS-B test samples.

(2) GRU and LSTM cannot be used to detect constant position deviation attack. When GRU and LSTM were used to detect constant position deviation attack, FNR reaches over 80%. This is because as the attack time grows, GRU and LSTM use the attack data to make predictions. However, the attack data only adds a small constant to the actual latitude and longitude. The attack data still conforms to the flight rules, and the latitude and longitude are still within the flight range. As a result, GRU and LSTM can hardly detect constant position deviation attack.

(3) Compared with LSTM-encoder-Decoder, the VAE-SVDD model has better performance of anomaly detection. The reasons are that distribution characteristics of ADS-B data are considered, BiGRU is used as the hidden layer of VAE, appropriate reconstruction values are selected, and SVDD is used to make the threshold better. However, these optimization mechanisms also increase detection time. Compared with LSTM-encoder-Decoder, the detection time of VAE-SVDD is about 10ms more.

(4) FPR of VAE-SVDD is slightly different, ranging from 3.44% to 6.39%. There are two main reasons: First, VAE latent variables are randomly sampled, so the reconstructed values are slightly different. Second, reconstructed values selection module has an impact on the detection of L (L is the length of sliding window) samples before the attack. For

example, FPR of DOS attack is 7.32% because reconstructed values selection module misjudges a small number of normal samples as anomaly samples before the attack.

(5) VAE-SVDD is suitable for the above five types of attack, and ER is below 10%. Compared with IForest, GRU and LSTM, VAE-SVDD model has lower FPR, FNR and ER. The adaptability of VAE-SVDD model is also better. In terms of detection time, it takes 46.3ms for the VAE-SVDD model to test 100 ADS-B data, which is slightly longer. The long detection time mainly comes from the BiGRU hidden layer of VAE, the time cost of reconstructed values selection module, and the time cost of SVDD threshold calculation.

5. Discussion

The proposed model is validated on five attack patterns,which concludes that the efficiency of the model is proved. However, there are some limitations in terms of developing the model further, which is the critical factor to improve the performance of the model.

(1) If the training data is attacked, the VAE-SVDD model is not reliable. The model is established on the basis of training a large number of normal historical ADS-B data. If the training data is attacked, the model cannot detect ADS-B anomaly data.

(2) If a large amount of ADS-B data is lost caused by noise or poor communication environment, the detection performance will be greatly reduced. As ADS-B data is time-dependent, a large amount of packet loss will make the model unable to better describe the characteristics of ADS-B time series, resulting in the reduction of detection performance. Typically, the selected flight has a packet loss rate of no more than 5%.

(3) In order to check the legitimacy of ADS-B data in one flight, the maximum and minimum values of the longitude, latitude, altitude and velocity of the flight need to be known in advance through the flight plan. If the maximum and minimum values are not clear, data preprocessing module will hardly normalize ADS-B data.

(4) If ADS-B anomaly data is sufficiently similar to ADS-B anomaly data, it is very difficult for VAE-SVDD model to detect ADS-B anomaly data. If attack behavior is very stealthy and conforms to flight rules (meaning that the characteristics of ADS-B anomaly data and ADS-B anomaly data are almost the same), it will lead to VAE-SVDD model difficult to check the legitimacy of ADS-B data. In fact, attackers should consider the effectiveness and stealth of attack behavior at the same time, and choose an appropriate value between them. The purpose of effectiveness is to disturb ATC(Air Traffic Control) situation, which requires attack behavior to be large enough. Stealth means that attacks are hard to detect, which requires that attack behavior to be small enough.

(5) VAE-SVDD model could not be obtained through real-time training, but the model could detect ADS-B anomaly data in real time. As VAE-SVDD model does not support online learning, the model needs to be updated regularly in order to accurately detect ADS-B anomaly data. However, the frequency of sending ADS-B data is 2 times per second and it takes 46.3 ms for VAE-SVDD model to detect 100 ADS-B data, so VAE-SVDD model could detect ADS-B anomaly data in real time.

6. Conclusion

In this paper, an ADS-B anomaly data detection method is proposed based on VAE-SVDD model. First, considering distribution characteristics of ADS-B data, VAE is used to reconstruct the ADS-B data. Then, SVDD is used to solve the adaptive problem of threshold. In addition, by increasing the dimension of VAE latent variable and selecting appropriate reconstructed values, FPR (False Positive Rate) and FNR (False Negative Rate) are reduced. In the experiments of detecting 5 common attacks, we verify the good detection performance of VAE-SVDD model. The advantages of VAE-SVDD model are low FPR and FNR, real-time detection of anomaly data and no need to change ADS-B protocol.

In the future, we will continue our work in the following three aspects. First, we plan to combine SSR data and ADS-B data to improve the performance of anomaly detection against high-stealth attacks (such as flight replacement attack). Second, in order to solve the problem of high FPR under high maneuvering state, we plan to add ADS-B features (such as heading) in training and adjust the network architecture or hyper-parameters of VAE-SVDD model. Third, since five common types of attacks cannot represent all attack behaviors, the performance of VAE-SVDD model for other complex attack behaviors needs to be further studied in the future.

Declaration of Competing Interest

The authors declare that they do not have any financial or non-financial conflict of interest in this paper.

CRediT authorship contribution statement

Peng Luo: Conceptualization, Methodology, Formal analysis, Software, Validation, Visualization, Writing - original draft. **Buhong Wang:** Supervision. **Tengyao Li:** Writing - review & editing. **Jiwei Tian:** Writing - review & editing.

REFERENCES

Ali S, Smith KA. Kernal width selection for svm classification: a meta-learning approach. Int. J. Data Warehous. Min. 2005;1(4):78–97. doi:10.4018/978-1-59904-951-9.ch209.

Akerman, S., Habler, E., Shabtai, A., 2019. VizADS-B: analyzing sequences of ADS-B images using explainable convolutional LSTM encoder-decoder to detect cyber attacks. Arxiv.

Baek J, Hableel E, Byon Y, Wong D, Jang K, Yeo H. How to protect ads-b: confidentiality framework and efficient realization based on staged identity-based encryption. IEEE Trans. Intell. Transp. Syst. 2017:1–11.

Chen, R.Q., Shi, G.H., Zhao, W.L., Liang, C.H., 2019. Sequential vae-lstm for anomaly detection on time series. Arxiv.

Chung, J., Gulcehre, C., Cho, K.H., Bengio, Y., 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. Eprint Arxiv.
Cho T, Lee C, Choi S. Multi-sensor fusion with interacting multiple model filter for improved aircraft position accuracy. Sensors 2013;13(4).
Costin A, Francillon A. Ghost in the air(traffic): on insecurity of ADS-B protocol and practical attacks on ADS-B devices. Proceedings of Black Hat USA, 2012.
Habler E, Shabtai A. Using LSTM encoder-decoder algorithm for detecting anomalous ADS-B messages. Comput. Sec. 2018;78:155–73. doi:10.1016/j.cose.2018.07.0004.
Kacem T, Barreto A, Wijesekera D, Costa P. Ads-bsec: a Novel Framework to Secure ads-b. ICT Express; 2017.
Kingma DP, Welling M. Auto-Encoding Variational Bayes. Arxiv; 2014.
Lee S-H, Kim Y-K, Han J-W. Protection method for data communication between ads-b sensor and next-generation air traffic control systems. Information 2014;5(4):622–33.
Li T, Wang B. Sequential collaborative detection strategy on ads-b data attack. Int. J. Crit. Infrastruct. Protect. 2019;24:78–99. doi:10.1016/j.ijcip.2018.11.003.
Li T, Wang B, Shang F, Tian J, Cao K. Online sequential attack detection for ADS-B data based on hierachical temporal memory. Comput. Sec. 2019;87 10.1016.j.cose.2019.101599.
Li T, Wang B, Shang F, Tian J, Cao K. Dynamic temporal ADS-B data attack detection based on sHDP-HMM. Comput. Secur. 2020;93. doi:10.1016/J.COSE.2020.101789.
Liu FT, Ting KM, Zhou ZH. Isolation-based anomaly detection. Acm Trans. Know. Discov. Data 2012;6(1):1–39. doi:10.1145/2133360.2133363.
Malhotra P, Vig L, Shroff G, Agarwal P. In: 23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Long short term memory networks for anomaly detection in time series; 2015.
Mccallie D, Butts J, Mills R. Security analysis of the ADS-B implementation in the next generation air transportation system. Int. J. Crit. Infrastruct. Protect. 2011;4(2):78–87. doi:10.1016/j.ijcip.2011.06.001.
Monteiro M. In: Digital Avionics Systems Conference. Detecting malicious ADS-B broadcasts using wide area multilateration. IEEE; 2015. doi:10.1109/DASC.2015.7311579.
Nanduri A, Sherry L. In: Integrated Communications Navigation & Surveillance. Anomaly detection in aircraft data using recurrent neural networks (RNN). IEEE; 2016. doi:10.1109/ICNSURV.2016.7486356.
Nijsure YA, Kaddoum G, Gagnon G, Gagnon F, Yuen C, Mahapatra R. Adaptive air-to-ground secure communication system based on ads-b and wide area multilateration. IEEE Trans. Vehic. Technol. 2016;65(5):3150–65. doi:10.1109/TVT.2015.2438171.
Schafer M, Lenders V, Martinovic I. Experimental analysis of attacks on next generation air traffic communication. In: International Conference on Applied Cryptography and Network Security; 2013. p. 253–71. doi:10.1007/978-3-642-38980-1_16.
Strohmeier M, Lenders V, Martinovic I. On the security of the automatic dependent surveillance-broadcast protocol. IEEE Commun. Surv. Tutor. 2015a;17(2):1066–87. doi:10.1109/COMST.2014.2365951.
Strohmeier M, Martinovic I. On passive data link layer fingerprinting of aircraft transponders. ACM Worksh. Cyber-Phys. Syst. Privacy 2015. doi:10.1145/2808705.2808712.
Strohmeier M, Lenders V, Martinovic I. Intrusion Detection for Airborne Communication Using PHY-Layer; 2015b. doi:10.1007/978-3-319-20550-2_4.
Strohmeier M, Schfer M, Fuchs M, Lenders V, Martinovic I. In: Digital Avionics Systems Conference, 2015 IEEE/AIAA 34th. OpenSky: a swiss army knife for air traffic security research. IEEE; 2015c. pp. 4A1-1-4A1-14. doi:10.1109/DASC.2015.7311411.
Tax DMJ, Duin RPW. Support vector data description. Mach. Learn. 2004;54(1):45–66. doi:10.1023/B:Mach.0000008084.60811.49.
Wang E, Song Y, Xu S, Guo J, Hong C, Qu P, Pang T, Zhang J. ADS-B anomaly data detection model based on deep learning and difference of gaussian approach. Trans. Nanjing Univ. Aeronaut. Astronaut. 2020;37(4):550–61.
Wesson KD, Humphreys TE, Evans BL. Can Cryptography Secure Next Generation Air Traffic Surveillance?. IEEE Security & Privacy; 2014.
Xu H, Feng Y, Chen J, Wang Z, Qiao H, Chen W. Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. In: WWW 2018: The 2018 Web Conference; 2018. p. 187–96. doi:10.1145/3178876.3185996.
Yang H, Zhou Q, Yao M, Lu R, Zhang X. A practical and compatible cryptographic solution to ads-b security. IEEE Internet Things J. 2018;6(2):3322–34.
Zhang T, Wu R, Lai R, Zhang Z. Probability hypothesis density filter for radar systematic bias estimation aided by ads-b. Signal Process. 2016;120(MAR.):280–7.

Peng Luo received the Bachelor's degree in computer science and technology from Hunan Normal University, Changsha, China, in 2018 and is pursing for the master's degree in Air Force Engineering University. His current research interest is machine learning and anomaly detection on ADS-B data.

Buhong Wang received the M.S. and Ph.D. degrees in signal and information prcessing from Xidian University, Xi'an, China, in 2000 and 2003, respectively, where his Ph.D. degree thesis "On Some Crucial Aspects of High-Resolution Direction of Arrival Estimation" was honored as the "Excellent Doctoral Dissertation of Shaanxi Province." From 2003 to 2005, with the support of the National Post-Doctoral Science Foundation, he was a Post-Doctoral Fellow with the Post-Doctoral Technical Innovation Center, Nanjing Research Institute of Electronics Technology, Nanjing, China. From 2006 to 2008, he was an Associate Professor with the School of Electronic Engineering, Xidian University. From 2009 to 2010, he was a Research Fellow with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. Since 2012, he has been a Professor with the Information and Navigation College, Air Force Engineering University, Xi'an. His current research interests include cyber security and cyber physical system.

Tengyao Li received the M.S. degree in computer science and technology from Air Force Engineering University, in 2016 and is pursuing for the Ph.D. degree in Air Force Engineering University. His current research interest is attack detection and resilient recovery on ADS-B data.

Jiwei Tian received the M.S. degree in information and communication engineering from Air Force Engineering University, in 2017 and is pursuing for the Ph.D. degree in Air Force Engineering University. His current research interest is machine learning and adversarial attack on cyber physical system.