DoS attack detection model of smart grid based on machine learning method

Wang Zhe, Cheng Wei*, Li Chunlin
Software College
Northeastern University
Shenyang, China
497034680@qq.com
*Corresponding author: Cheng Wei

***Abstract*—In recent years, smart grid has gradually become the common development trend of the world's power industry, and its security issues are increasingly valued by researchers. Smart grids have applied technologies such as physical control, data encryption, and authentication to improve their security, but there is still a lack of timely and effective detection methods to prevent the grid from being threatened by malicious intrusions. Aiming at this problem, a model based on machine learning to detect smart grid DoS attacks has been proposed. The model first collects network data, secondly selects features and uses PCA for data dimensionality reduction, and finally uses SVM algorithm for abnormality detection. By testing the SVM, Decision Tree and Naive Bayesian Network classification algorithms on the KDD99 dataset, it is found that the SVM model works best.**

Keywords—DoS attacks, Machine Learning, Smart Grid, SVM, Software Security

I. INTRODUCTION

With the advent of the information age, the industrial industry is undergoing a major transformation. Under such a development background, the concept of smart grid emerged as the times require, and it has been widely recognized on a global scale, becoming the common development trend of the world power industry [1]. However, the incident of smart grid intrusion has occurred from time to time. For example, on January 6, 2016, the Ukrainian power grid system was attacked by hackers, and hundreds of households were forced to interrupt power supply. This is the first cyber-attack that has led to power outages in history [2]. This attack on industrial control systems is undoubtedly a milestone. Therefore, large power grids in various countries need to strengthen security and avoid serious network attacks.

II. LITERATURE REVIEW

Smart grid networks require reliable local detection schemes to detect intrusions. In the literature [3], Eunsuk Kang et al. used a Markov chain-based game analysis model to propose a real-time detection scheme for false data injection attacks in smart grids. This scheme lacks coverage of common attack types of the grid. In the literature [4], an improved CUSUM intrusion attack detection method based on Bloom Filter address statistics for dynamic threshold update is proposed. The method collects grid traffic data in the data collection phase, and then performs intrusion behavior judgment. The attack detection method based on alarm data fusion has proposed by Yanan Sun [5].It conforms to the heterogeneous characteristics of the grid, thus eliminating the deviation of the simple physical layer detection method or the simple information layer detection method. But this method can only determine that the line is abnormal and cannot locate the specific attacked node. And this method requires additional host devices to deploy Snort, and the security of the host device will also affect the detection results. In this paper [6], the vulnerability of the route maintenance phase of the wireless Mesh network DSR (Dynamic Source Routing) protocol is used, and the corresponding attack method is designed for the network protocol. Wang K et al. [7] proposed using the Bayesian honeypot game model to solve the shortcomings of the traditional honeypot technology detection dynamic attack ability.

In summary, this paper proposes a DoS attack intrusion detection model based on machine learning for the defects of the existing intrusion detection system of smart grid. The model mainly uses the machine learning algorithm to detect the DoS attack behavior or normal behavior, and improves the security of the smart grid.

III. SMART GRID DOS INTRUSION DETECTION METHOD

According to the design pattern of smart grid and the characteristics of DoS attack, this paper designs a smart grid intrusion detection structure based on machine learning, as shown in Figure 1.

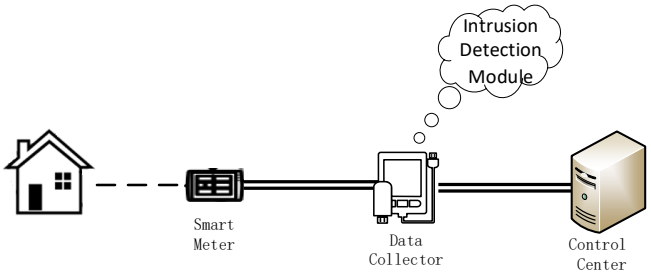

Fig. 1. Smart grid intrusion detection architecture diagram

Aiming at the existing design structure of smart grid, this paper uses the attack vulnerabilities of smart meters and data servers to add data acquisition and intrusion detection modules between smart meters and data servers. Real-time data acquisition and detection in the smart grid is realized. When DoS attack behavior is detected, the alarm system is activated to perform alarm processing.

The principle of DoS attack generally utilizes the lack of effective authentication mechanism for management frames and control frames and the defects of CSMA/CA mechanism [8]. An attacker can send a normal connection request through a large number of forged illegal management frames and control frames, which will significantly increase the probability of the attack node accessing the wireless attack communication channel, thereby making the wireless access point unable to provide normal service or access due to access overload. The purpose of continuously occupying the communication channel for a long time affects the normal communication between other legitimate clients and the wireless access point.

The typical attack methods of DoS are smurf, neptune, teardrop, pod, land and so on. DoS attacks make the resources of smart grid unavailable. It achieves the goal of attack by blocking communication or consuming network bandwidth. By repeatedly sending a large number of useless requests to smart meters or collectors, the devices can not process data from normal nodes, which occupies their storage space and network bandwidth.

In this paper, the attack detection algorithm based on support vector machine is selected. Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. The support vector machine adopts the principle of structural risk minimization, which ensures good generalization ability and improves the prediction accuracy. Through the support vector machine theory, an optimal classification hyperplane with strict constraints can be formed, and the data aggregation points are separated on both sides of the hyperplane, as shown in Figure. 2.

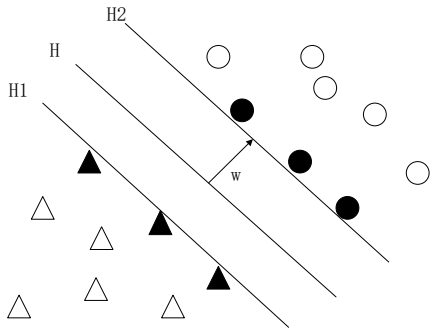

Fig. 2. Schematic diagram of the super plane

SVM is a supervised learning algorithm. The goal of SVM is to maximize the boundary range between two separate classes while minimizing classification errors. First, we need to assume that $T = \{(x,y)|(x_1, y_1), \dots, (x_i, y_i)\}$ is the given training set. Among them, $x_i \in R^n$, $y_i \in \{-1,1\}$. Then there is a plane in a high-dimensional space, generally called a hyperplane, which can linearly divide the above training set. This hyperplane can be expressed as $\omega \cdot x + b = 0$, Where ω is the unit normal vector. There is an optimal hyperplane in this kind of hyperplane. In addition to accurately dividing the input training set, it can ensure that the two nearest alien vectors on the two sides of the hyperplane are the farthest. The two heterogeneous vectors are support vectors. The distance between them is the classification interval. Therefore, a hyperplane is uniquely determined by a set of support vectors that satisfy the above characteristics. The classification spacing between H1 and H2 is $\frac{2}{||\omega||}$. In order to maximize the classification spacing, it can be converted to $||\omega||^2$ minimum.

Aiming at the real-time characteristics of DoS attacks, this paper designs a dynamic DoS intrusion detection model based on machine learning, as shown in Figure 3. The detection model mainly includes: a data acquisition module for collecting smart network data packets, a feature extraction module, an intrusion detection module for identifying DoS attacks, and an attack detection response module for outputting and alarming results.

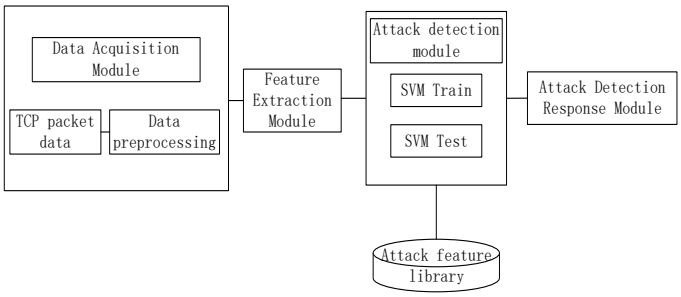

Fig. 3 Attack detection model

IV. Experiments

A. Datasets and Processing

For the research of intrusion detection, a large amount of effective datasets is required. Data can be collected through some capture tools, such as Labium under Windows, or dedicated software such as snort, to capture data packets and generate connection records as a dataset. Here, this paper introduces KDDCup99 [9] called the Network Intrusion Detection Dataset used in the research of intrusion detection technology based on data mining.

TABLE I. Intrusion Types of KDDCup99 Dataset

Attack Class	Implication	Specify attack way
Normal	Normal record	normal
DoS	Denial of service attack	back, land, neptune, pod, smurf, teardrop
Probing	Surveillance and other detection activities	ipsweep, nmap, portsweep, satan
R2L	Illegal access from a remote machine	ftp_write, guess_passwd, imap, multihop, phf, spy, warezxlied, warezmaster
U2R	Ordinary user's illegal access to local superuser privileges	buffer_oberflow, loadmobile, perl, rootkit

The KDDCup99 training dataset contains one normal type called 'normal' and 22 attack types, as shown in TABLE I. The attack types can be summarized into four classes: DoS, Probing, R2L, U2R. There are 15 types of attacks only appears in test dataset. At the same time,

training set and test set have different probability distributions, which makes the intrusion detection more realistic.

One dataset called kddcup_data_10_percent is generally used in the experiment, which is a 10% of KDDCup99. We select the DoS attack data to train the DoS attack detection model, which includes six types of attacks, namely back, land, neptune, pod, smurf, teardrop. DoS type data contains 488,736 training sample and 290,446 test sample.

Original train and test datasets contain numerical features and text features, the numerical features include continuous and discrete features. Text features need to be transformed to vector, encoded according to the categories counts of each feature, so that each category is uniquely encoded. For continuous numerical features, it is standardized and normalized as follows. For the numerical discrete features, mainly from the text feature and the two-category feature, the discrete feature is not standardized and used as a feature directly.

Suppose training data have N network connection records, and feature vectors in each connection records written as X_{ij} ($1 \le i \le n, 11 \le j \le 32$). Suppose X'_{ij} as standardization results of X_{ij}. We can get following formulation about standardization:

$$X'_{ij} = \frac{X_{ij} - AVG_j}{STAD_j} \quad (1)$$

$$AVG_j = \frac{1}{n}(X_{1j} + X_{2j} + \cdots + X_{nj}) \quad (2)$$

$$STAD_j = \frac{1}{n}(|X_{ij} - AVG_j| + |X_{2j} - AVG_j| + \cdots + |X_{nj} - AVG_j|) \quad (3)$$

Among them: AVG_j is mean value, $STAD_j$ is average absolute deviation, which is better robustness to isolated points than $\sigma_A = \sqrt{\frac{\sum(A-\bar{A})^2}{n-1}}$. In the above calculation, the following judgment is required:

- $if\ AVG_j$ equals 0,then $X'_{ij} = 0$.
- $if\ STAD_j$ equals 0,then $X'_{ij} = 0$.

Normalize each of values to the [0,1] interval. Suppose X'_{ij} is normalization results of X_{ij}, and then:

$$X'_{ij} = \frac{X'_{ij} - X_{min}}{X_{max} - X_{min}},$$

$$X_{min} = \min\{X'_{ij}\},$$

$$X_{max} = \max\{X'_{ij}\},$$

$$for\ 1 \le i \le n, 11 \le j \le 32. \quad (4)$$

It tends to fall into dimensional disasters when doing feature extraction and processing with high-dimensional feature vectors. As the data dimension increases, the number of samples required for algorithm increases exponentially. In some applications, it is very disadvantageous, and requires more memory and processing power. In addition, as the dimension increases, the sparseness of the data will become higher and higher.

Therefore, we need to reduce feature dimensions. Feature reduction includes feature selection and feature extraction. We have compared and analyzed the effective of different feature selection and feature extraction techniques. Finally, PCA is selected to get 29 features for model training, shown as TABLE II.

TABLE II. FEATURES AFTER FEATURE REDUCTION

Discrete Features	Continuous Features
protocol_type, service, flag, land, logged_in, root_shell, su_attempted, Is_host_login, Is_guest_login	Duration,src_bytes,dst_bytes, wrong_fragment,urgent,hot, num_failed_login,num_root, num_acces_files, num_outbound_cmds, count,srv_count,serror_rate, srv_serror_rate, rerror_rate,srv_rerror_rate, num_file_creations,num_shells, same_srv_rate,diff_srv_rate

B. Experiment Results

According to the actual situation, we selects the accuracy, precision, recall rate, F1-Score to evaluate the effective of model. Formulation as following:

- Accuracy:

$$A = \frac{TP + TN}{TP + FN + FP + TN}$$

- Precision:

$$P = \frac{TP}{TP + FP}$$

- Recall:

$$R = \frac{TP}{TP + FN}$$

- F1 Score:

$$F1\ Score = \frac{2 * P * R}{P + R}$$

Among them: TP is the number of attack samples that are correctly classified, FN is the number of normal samples that are misclassified as attacks; FP is the number of attack samples that are misclassified as normal; TN is the number of normal samples that are correctly classified. The 'Accuracy' reflects the proportion of correctly classified

samples in the whole sample; 'Precision' reflects the proportion of correctly classified attack samples in all attack samples; 'Recall' reflects the proportion of correctly classified attack samples in all attack samples; 'F1 Score' was called balanced F score as well, which is used to define the harmonic mean of the precision and recall rate.

C. Results Analysis

We use the accuracy, precision, recall and F1 Score indicators to compare three different detection algorithms: Support Vector Machine (SVM), Decision Tree and Naive Bayesian Network. The intrusion detection performance of DoS attacks is shown in Figure 4. As can be seen from the figure, the SVM algorithm achieves the best results.

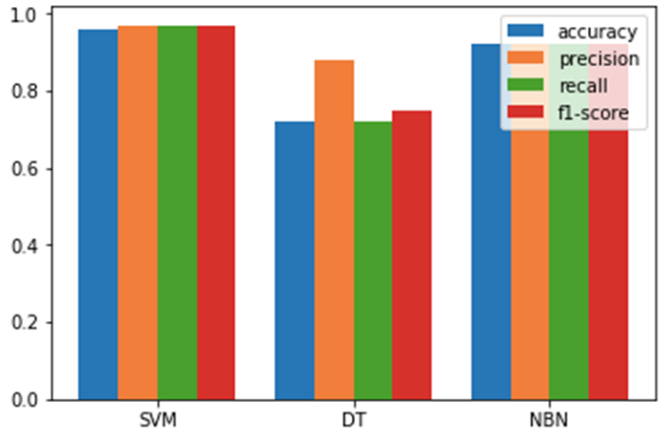

Fig. 4 Results for Different Algorithms

In order to evaluate the classification performance of an algorithm on different categories, we compare the accuracy of each algorithm to the attack and normal categories, so as to recall rate and F1 Score, as shown in TABLE III, TABLE IV and TABLE V. It can be seen that the SVM algorithm has higher precision and recall rate than other algorithms for different detection categories.

TABLE III. SVM DETECTION RESULTS

Classes	Precision	Recall	F1_Score	Support
0	0.97	0.99	0.98	229853
1	0.97	0.87	0.92	60593
Avg/Total	0.97	0.97	0.97	290446

TABLE IV. DECISION TREE DETECTION RESULTS

Classes	Precision	Recall	F1_Score	Support
0	1.00	0.65	0.79	229853
1	0.43	0.99	0.60	60593
Avg/Total	0.88	0.72	0.75	290446

TABLE V. NAIVE BAYESIAN NETWORK DETECTION RESULTS

Classes	Precision	Recall	F1_Score	Support
0	0.96	0.94	0.95	229853
1	0.80	0.84	0.82	60593
Avg/Total	0.92	0.92	0.92	290446

V. CONCLUSION

Aiming at the problem of smart grid intrusion detection, this paper proposes a smart grid DoS attack detection model based on machine learning. The method collects network communication data between the smart meter and the data server in real time. Through feature selection and PCA dimension reduction to select more representative features, the SVM classifier trained model is used to identify and detect DoS attacks. Experiments on the KDD99 dataset show that the SVM classification model is better than Naive Bayesian Network and Decision Tree classification models.

This method has higher classification detection rate and accuracy, which can effectively improve the security of the smart grid DoS intrusion detection system. In the future work, the classification model calculation method will be further expanded and improved to obtain better classification detection performance.

REFERENCES

[1] Vidyaev I G, Ivashutenko A S, Samburskaya M A. «Smart Grid» Concept As A Modern Technology For The Power Industry Development[C]// 2017:012173.

[2] Huang H B, Hong L, Chang-Yue Y U, et al. Analysis on Ukraine Power Grid Blackout and Its Enlightenment of ICS in China[J]. Standard Science, 2016.

[3] Jianye Hao, Eunsuk Kang, Jun Sun, Zan Wang, "An Adaptive Markov Strategy for Defending Smart Grid False Data Injection from Malicious Attackers", IEEE Transactions on Smart Grid. Sept. 2016.

[4] Jiaxuan Fei,Tao Zhang,Yuanyuan Ma,Cheng Zhou. A DDoS attack detection method for power grid industrial control system based on BF-DT-CUSUM algorithm[J]. Telecommunications Science.2015 (12).

[5] Yanan Sun, Xiaohon Guan, Ting Liu, Yang Liu, "A cyber-physical monitoring system for attack detection in smart grid", Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on, Turin, Italy, Dec. 2014.

[6] Yi P, Zhu T, Zhang Q, et al. A denial of service attack in advanced metering infrastructure network[C]// IEEE International Conference on Communications. IEEE, 2015:1029-1034.

[7] Wang K, Du M, Maharjan S, et al. Strategic Honeypot Game Model for Distributed Denial of Service Attacks in the Smart Grid[J]. IEEE Transactions on Smart Grid, 2017, PP(99):1-1.

[8] Pooja B, Pai M M M, Pai R M, et al. Mitigation of insider and outsider DoS attack against signature based authentication in VANETs[C]// Computer Aided System Engineering. IEEE, 2014:152-157.

[9] Saxena H, Richariya V. Intrusion Detection in KDD99 Dataset using SVM-PSO and Feature Reduction with Information Gain[J]. International Journal of Computer Applications, 2014, 98(6):25-29

[10] Sousa, P. H. F.; Nascimento, N. M. M.; Almeida, J. S.; Rebouças Filho, P. P. and Albuquerque, V. H. C. (2019). Intelligent Incipient Fault Detection in Wind Turbines based on Industrial IoT Environment. Journal of Artificial Intelligence and Systems, 1, 1–19.