are complicated if being expressed in models, for example, thousands of leaves on a tree will have millions of faces in models, as a result, operational speed of the software will be greatly reduced if these complicated garden models are imported into Quest3D, and then the expected effects or virtual roaming will hardly be realized. Therefore, to achieve high simulated or artistic effect of garden landscapes in Quest3D, further studies are required to simplify the method of modeling.

References

1. Bishop, I. D., Ye, W. S., & Karadaglis, C., Experiential approaches to perception response in virtual worlds. Landscape and Urban Planning, 2001.54(1), pp.117-125.
2. Frade, M., De Vega, F. F., & Cotta, C., Modelling video games' landscapes by means of genetic terrain programming-a new approach for improving users' experience. In Applications of evolutionary computing, Springer Berlin Heidelberg, 2008, pp. 485-490.
3. YANG, J. S., DU, Z. Q., PENG, Z. H., HUANG, J. N., & CHEN, Y. X., Modeling technology for 3D landscape models of cybercity. Journal of Wuhan University of Hydraulic and Electric Engineering, 2003, 3, pp. 008.
4. Wang, S., Mao, Z., Zeng, C., Gong, H., Li, S., & Chen, B., June. A new method of virtual reality based on Unity3D. In Geoinformatics, 2010 18th International Conference on IEEE, 2010, pp. 1-5.
5. Oh, J. Y., Stuerzlinger, W., & Danahy, J., June. Sesame: towards better 3d conceptual design systems. In Proceedings of the 6th conference on Designing Interactive systems, 2006, pp. 80-89
6. Ying, Z. X. P. C. G. H. D. A., & XueTao, Z. A. H., Design of a virtual simulation platform based on 3ds Max with implementation based on Virtools. Journal of Beijing University of Chemical Technology (Natural Science Edition), 2009, 21p.

A Layered Bayesian Network Intrusion Detection Algorithm

Wang Xingzhu[1]

Furong College Hunan, University of Arts and Science, Hunan Changde, 415000, China E-mail: Wangxzhu@sina.com

Corresponding author is Wang Xingzhu

Abstract

Aiming at the problem that the network intrusion events frequently happen in the current internet era and the fallout ratio and omission ratio of the intrusion detection system are two high, it proposes a Bayesian intrusion detection algorithm based on the multilayer feature extraction of Gabor. First, aiming at the problem that the degree of accuracy of the Bayesian classification algorithm is low, the multilayer Bayesian classification recognition algorithm is utilized and combined to propose the Gabor multilayer feature extraction algorithm, and based on this

point, the Bayesian algorithm of Gabor multilayer feature extraction is designed and realized. After that, according to the feature of the intrusion detection data, the algorithm flow of the intrusion detection system is designed on the basis of this improved algorithm and the effective improvement of the success rate of intrusion detection is realized. The experimental result reveals that the Bayesian intrusion detection algorithm based on the multilayer feature extraction of Gabor could effectively improve the precision ratio and recall ratio of the intrusion detection and the expected requirements are reached.
Keywords: NETWORK SECURITY, FEATURE EXTRACTION, INTRUSION DETECTION, CLASSIFICATION RECOGNITION, BAYESIAN CLASSIFICATION.

1. Introduction

The 21st century is the age of network, the network has gradually integrated into people's daily life and become one of the most convenient tools to obtain information, and people's reliance for the network is increasing. The establishment of efficient intrusion detection system to prevent and respond to the increasing network intrusion behaviors is of great significance for maintaining the security of the information system and protecting the normal production and living order of the society. The concept of intrusion detection system (intrusion detection system, IDS for short) is automatic system to detect the network intrusion behaviors and it is the premise of security protection for the intrusion has been widely used in many security systems.

In order to improve the detection rate of intrusion system, reduce the fallout ratio and omission ratio, and then improve the level of intrusion detection technology, many domestic and overseas scholars and the related safety technical personnel have conducted extensive research on intrusion detection algorithm. For example, the intrusion detection system on the basis of SVM algorithm is designed and the limitation of the traditional machine learning algorithm is solved in the literature based on the advantage of the algorithm of SVM in the classification of small samples and nonlinear data. It makes the classification detection for the intrusion data in the literature though the feature extraction of the intrusion data and the nonlinear approximation feature of the neural network model. It adopts the K-Means algorithm to reduce the sample size of the neural network training, improve the learning ability of the algorithm, reduce the calculation strength of the competitive learning, and apply it into the intrusion detection data through multilayer network structure with back propagation learning mechanism in literature. It uses CACC discretization algorithm to improve the expression ability of the training data, and then uses the Naive Bayes and K-Means clustering hybrid approach for the intrusion detection, etc in literature.

For the intrusion data, the previous intrusion detection algorithms often adopt the processing mode of complanation, and it could not reflect the different effects of the intrusion data with different levels or different time periods predicted on the intrusion data in the next step. However, in the practical application, the effect of the recent intrusion data is obviously higher than the former intrusion data. The data with a higher level will carry more global information, while the lower level data will carry more local information. In consideration of this point, this paper proposes a kind of layered Bayesian network intrusion detection algorithm to distinguish the feature information brought by the data with different levels and realize the effective promotion of the performance of the algorithm.

2. Multilayer Gabor Feature Extraction

2.1. Multilayer Gabor Filter Bank

Within the scope of space - frequency domain, the standard Gabor filter can be defined as:

$$\begin{cases} G(x,y,w_0,\sigma,r,\theta)=\dfrac{1}{\sqrt{\pi}r\sigma}e^{t} \\ t=-\dfrac{1}{2}\left[\dfrac{(rR_1)^2+R_2^2}{(r\sigma)^2}\right]+iw_0R_1 \end{cases} \quad (1)$$

Thereinto, $R_1=x\cos\theta+y\sin\theta$, $R_2=-x\sin\theta+y\cos\theta$, w_0 is the radius of the frequency on the arc length per unit, θ is the direction of the radian, and σ is the standard deviation of the elliptical gauss envelope in the direction of x axis. Within the spatial scale, the Gabor filter is centered on $x=0, y=0$. It defines the aspect ratio of the elliptical gauss envelope of Gabor filter as:

$$r=\sigma_y/\sigma_x \quad (2)$$

Thereinto, σ_x and σ_y are the standard deviations of the elliptical gauss envelope in the direction of x axis and y axis respectively.

As mentioned earlier, in order to realize the effect of the layered feature extraction algorithm, the data feature shall be equipped with the following characteristics: the features with different lelvels could

represent different information, the date with higher layer could carry more global information, while the date with lower layer carries more local information. In consideration of this point, the Gabor filter bank could be established through recursive fashion. Firstly, define that a sampling feature point is equipped with the highest level, and then such sampling point is decomposed into several sampling feature points with relatively low level. In order to realize extract the global information in data points with relatively high level and extract local information in data points with relatively low level, the standard deviation σ^{ls} msut be strictly restricted as per the level:

$$\sigma^{ls} = \frac{k}{\sqrt{\ln 2}} \tag{3}$$

Thereinto, ls is the subscript of the sampling point s in the tier of l, $l = 1, \cdots, N_l$ is the quantity of the tiers, and $s = 1, \cdots, N_{ls}$ is the size of the tier of l. In order to better extract the local information, the value of k is:

$$k = \frac{mean(d1, d2, d3, d4)}{2} \tag{4}$$

Thereinto, $d1, d2, d3, d4$ are respectively the distances between the sampling points with the four adjacent sampling points of $n1, n2, n3, n4$ (Fig.1).

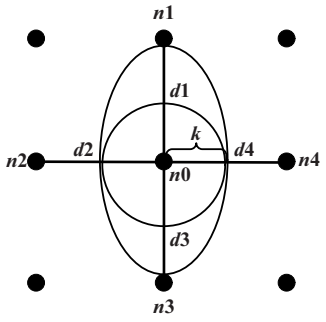

Figure 1. Semi-gauss Envelope

When the standard deviation of the node s in the level of l and the aspect ratio r of Gabor filter are confirmed, the Gabor filter bank $\mathbf{GB}_j^{ls}$ can be defined as:

$$\begin{cases} \mathbf{GB}_j^{ls} = \left\{ G_{j1}^{ls}, \cdots, G_{jN_\omega}^{ls} \right\} \\ G_{ji}^{ls}(x,y) = G\left(x^{ls} - x, y^{ls} - y, \omega_i, \sigma^{ls}, r, \theta_j \right) \\ \omega_i \in \mathbf{\Omega} = \left\{ \omega_1, \cdots, \omega_{N_\omega} \right\}, i = 1, \cdots, N_\omega \\ \theta_j \in \mathbf{\Theta} = \left\{ \theta_1, \cdots, \theta_{N_\theta} \right\}, j = 1, \cdots, N_\theta \end{cases} \tag{5}$$

Thereinto, $\mathbf{\Omega}$ is the spatial frequency set, $\mathbf{\Theta}$ is the collection of direction angle, G_{ji}^{ls} is the center of the Gabor filer in the node of $\left(x^{ls}, y^{ls} \right)$ in formula (1). Thus, for each sampling point and direction, the Gabor filter bank of $\mathbf{GB}_j^{ls}$ is the Gabor filter set with different frequencies in the spatial frequency set of $\mathbf{\Omega}$.

2.2. Criterion of the Optimum Gabor Filter

Definition 1: (Within Class Scatter Matrix) Based on certain criterion, an optimum Gabor filter can be selected in the Gabor filter bank of $\mathbf{GB}_j^{ls}$, and for the issue in c type, assume n d dimensional instances of $\boldsymbol{\chi}$:

$$\boldsymbol{\chi} = \left\{ \mathbf{x}_1, \cdots, \mathbf{x}_n \right\} \tag{6}$$

Then the Within Class Scatter Matrix of S_w can be defined as:

$$\begin{cases} \mathbf{S}_w = \sum_{i=1}^{c} \mathbf{S}_i \\ \mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in \boldsymbol{\chi}_i} \mathbf{x} \\ \mathbf{S}_i = \sum_{\mathbf{x} \in \boldsymbol{\chi}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \end{cases} \tag{7}$$

Definition 2: (Between Class Scatter Matrix) After the total average vector of $\mathbf{m}$ is defined, the between class scatter matrix of $\mathbf{S}_B$ can be defined as:

$$\begin{cases} \mathbf{m} = \frac{1}{n} \sum_{\mathbf{x} \in \boldsymbol{\chi}_i} \mathbf{x} = \frac{1}{n} \sum_{i=1}^{c} n_i \mathbf{m}_i \\ \mathbf{S}_B = \sum_{i=1}^{c} n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T \end{cases} \tag{8}$$

Definition 3: (Criterion) Based on the above mentioned scatter matrixes, the criterion can be defined as:

$$DM = Discriminant\ Measure = \frac{|\mathbf{S}_B|}{|\mathbf{S}_w|} \tag{9}$$

Make the (h_d, I_d) as the presort training data, thereinto, $h_d \in \mathbf{C}$, and $I_d \in \mathbf{I}$:

$$\mathbf{C} = \left\{ c_i : i = 1, \cdots, N_c \right\} \tag{10}$$

There into, N_c is the quantity of the classification, $\mathbf{I}$ is the training dataset, and the selection method of the optimal Gabor filter is:

$$\begin{cases} OG_j^{ls} = \arg\max_{G_{ji}^{ls}} \{i\} (DM_i) \\ \boldsymbol{\chi}_i = \left\{ (h_1, g_1), \cdots, (h_{N_I}, g_{N_I}) \right\} \\ g_d = \sum_{\{x\}} \sum_{\{y\}} I_d(x,y) G_{ji}^{ls}(x,y) \end{cases} \tag{11}$$

There into, $G_{ji}^{ls} \in \mathbf{GB}_j^{ls}$, g_d is the response of the Gabor filter, and N_I is the quantity of the training dataset. After the OG_j^{ls} of each sampling point is ac-

quired, the Gabor feature of the sampling point can be defined as:

$$\begin{cases} \mathbf{a}^{ls} = \left[a_1^{ls}, \cdots, a_{N_\theta}^{ls} \right]^T \\ a_j^{ls} = \sum_{\{x\}} \sum_{\{y\}} I(x,y) OG_j^{ls}(x,y) \end{cases} \quad (12)$$

The Gabor feature of the sampling point of s in the level of l becomes the vector quantity of one N_θ dimension, and the vector element is the response data of the optimal Gabor filter for the training data of $\mathbf{I}$ in different directions. Then the Gabor feature of the sampling point in the level of l can be defined as:

$$\begin{cases} \mathbf{a} = \left\{ \mathbf{a}^1, \cdots, \mathbf{a}^{N_L} \right\} \\ \mathbf{a}^l = \left\{ \mathbf{a}^{l1}, \cdots, \mathbf{a}^{lN_{LS}} \right\} \end{cases} \quad (13)$$

3. Multilayer Bayesian Network

For the finite set of random variables $\mathbf{U} = \{A_1, \cdots, A_n\}$, common Bayesian network is generally defined as $\langle DAG, \mathbf{CP} \rangle$, thereinto, $DAG = (\mathbf{V}, \mathbf{E})$ is the acyclic directed connection diagram of Bayesian network, $\mathbf{V} = \{A_1, \cdots, A_n\}$ is the collection of nodes, $\mathbf{E} = \left\{ (A_i, A_j) : A_i, A_j \in \mathbf{V}, i \neq j \right\}$ is the directed connecting line from node to node, and (A_i, A_j) represents the directed connection from A_i to A_j and it means the direct influence on A_j generated by A_i. $\mathbf{CP}$ is the conditional probability distribution set of nodes, and the conditional probability distribution of the node of A_i, thereinto, the ΠA_i is the father node of the node of A_i in the network structure of DAG, then the joint probability function of $P(\mathbf{U})$ can be defined as:

$$P(A_1, \cdots, A_n) = \prod_{i=1}^{n} P(A_i | \Pi A_i) \quad (14)$$

For common uncoursed Naïve Bayesian classifiers, as the father nodes of the all nodes of A_i are assumed to be one class node of C, then though the formula (14), the joint probability function of the Naïve Bayesian network is:

$$P(A_1, \cdots, A_N, C) = \prod_{i=1}^{N} P(A_i | C) P(C) \quad (15)$$

The multilayer Bayesian network structure is as shown in Fig.2, and it is often defined as $DAG_H = \langle \mathbf{V}_H, \mathrm{E}_H \rangle$. Make the A^{ls} as the node of the level in $\mathbf{V}_H$ as well as the random variable in $\mathbf{U}_H$. According to Section 2.2, the node of A^{ls} is equipped with the Gabor feature of $\mathbf{a}^{ls}$, and the node collection of $\mathbf{V}_H$ can be defined as:

$$\begin{cases} \mathbf{V}_H = \mathbf{A}^1 \cup \cdots \cup \mathbf{A}^{N_L} \\ \mathbf{A}^l = \left\{ \mathbf{A}^{l1}, \cdots, \mathbf{A}^{lN_{LS}}, \right\} \end{cases} \quad (16)$$

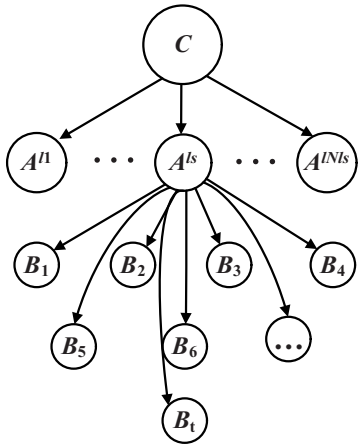

Figure 2. Multilayer Bayesian Network

Thereinto, $\mathbf{A}^l$ is the node collection of the level of l, make $\mathbf{\Phi}^{ls}$ as the point set of sub-sampling for the node of A^{ls}, thus it can be defined as:

$$\begin{cases} \mathbf{\Phi}^{ls} = \left\{ B_1^{ls}, \cdots, B_t^{ls} \right\} \\ B_i^{ls} \in \mathbf{A}^{l+1} \end{cases} \quad (17)$$

Thereinto, $l = 1, \cdots, N_L - 1$, and the directed line set of $\mathbf{E}_H$ can be defined as:

$$\begin{cases} \mathbf{E}_H = \mathbf{E}^1 \cup \cdots \cup \mathbf{E}^{N_L - 1} \\ \mathbf{E}^l = \mathbf{E}^{l1} \cup \cdots \cup \mathbf{E}^{lN_{LS}} \\ \mathbf{E}^{ls} = \left\{ \left(A^{ls}, B_1^{ls} \right), \cdots, \left(A^{ls}, B_t^{ls} \right) \right\} \end{cases} \quad (18)$$

Thereinto, $\left(A^{ls}, B_i^{ls} \right)$ is the directed line from the father node of A^{ls} to the child node of B_i^{ls}, and $B_i^{ls} \in \mathbf{\Phi}^{ls}$.

For the classification problem, the multilayer Bayesian network structure obtained must be properly revised, make it include the node of C, $DAG_H \to DAG'_H$, then the multilayer structure of $DAG'_H = \langle \mathbf{V}'_H, \mathrm{E}'_H \rangle$ including the node of C can be defined as:

$$\begin{cases} \mathbf{V}'_H = \mathbf{U}'_H = \mathbf{V}_H \cup \{C\} \\ \mathbf{E}'_H = \mathbf{E}_H \cup \mathbf{E}_C \\ \mathbf{E}_C = \left\{ \left(C, A^{ls} \right), A^{ls} \in \mathbf{V}_H \right\} \end{cases} \quad (19)$$

Thereinto, $\mathbf{E}_C$ is the directed line set from the node of C to the node collection of $\mathbf{V}_H$, all nodes in DAG'_H except the class node of C take the class node of C as the father node (advance node).

For the DAG'_H network with integral hierarchical structure, the conditional probability distribution set of CP must` be defined. The multilayer Bayesian network structure includes the continuous and discrete random variables, the variables in $\mathbf{U}_H$ is continuous, while the class variable of C is discrete. Thus, for the

above continuous Gabor feature variables, its conditional probability distribution of $P\left(A^{ls}\left|\Pi'_{ls}\right.\right)$ can be defined as one multivariate Gaussian function, thereinto, the Π'_{ls} is all the father node sets of the node of A^{ls} in DAG'_H. For the discrete classification variable of C, there is no father node, and its conditional probability distribution of $P(C)$ can be defined as one polynomial distribution. Thus the joint probability distribution function of $\mathbf{U}'_H$ can be defined as formula (14).

4. Intrusion Detection Based on the Improved Bayes

Similar with neural network, SVM and other learning algorithms, the improvement of Bayesian intrusion algorithm needs sufficient training data to train the established multilayer Bayesian model, then make use of the trained model to classify the actual intrusion data, distinguish the normal data from the abnormal data and realize the forewarning function for the abnormal attack data. Meanwhile, the Bayesian model is just suitable for the behavior of uncertain probability of the intrusion detection. Thus the establishment of intrusion detection based on the improved Bayes is feasible. The framework of the intrusion detection algorithm based on the improved Bayes is as shown in Fig. 3.

It can be seem from Fig. 3 that the multilayer Bayesian intrusion detection flow mainly includes

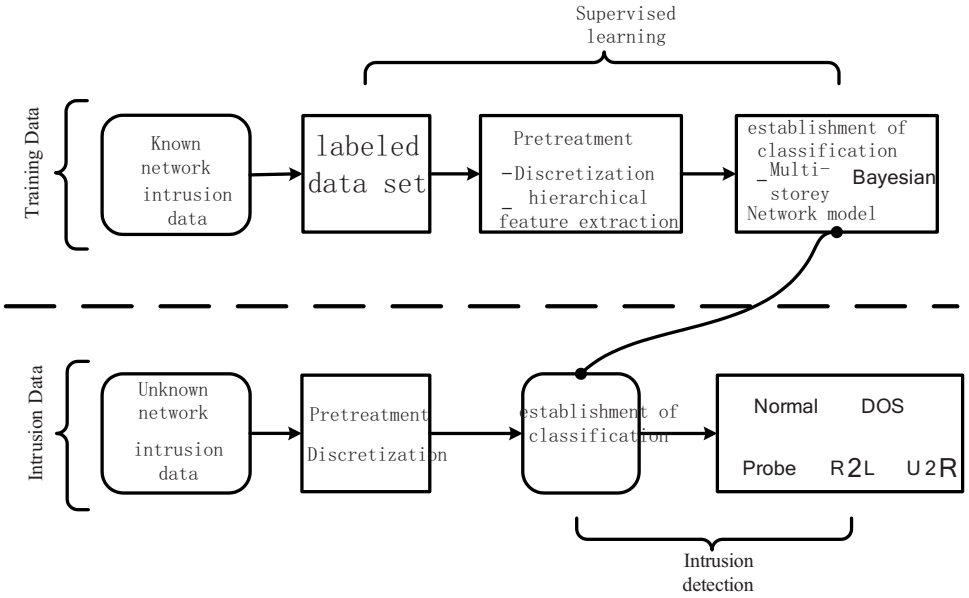

Figure 3. Multilayer Bayesian Intrusion Detection Flow

two steps: the first step is the establishment of the classifier. The training data set is established through the mapping relation between the existing tagged network intrusion data sample of T_k and the known class sample of C_k:

$$Y_{TC}=\left(\exists T_k,C_k\left|\forall T_k\in C_k\right.\right) \tag{20}$$

Make the discretization pretreatment and multilayer feature extraction for the training set with such marks, then according to the pretreatment data, and combined with the posterior probability of $P\left(T_k\left|C_k\right.\right)$, the multilayer Bayesian network classifier of $U_k=\left\langle DAG'_{Hk},\mathbf{CP}\right\rangle$ is established. The second is the detection classification of the intrusion data. Make the discretization pretreatment for the unkown intrusion data of T_a, and make the classification detection on the network intrusion data according to the established Bayesian classifier of $\left\{\left(U_a=\left\langle DAG'_{Ha},\mathbf{CP}\right\rangle\right)\subseteq\left(U_k=\left\langle DAG'_{Hk},DAG'_{Ha},\mathbf{CP}\right\rangle\right)\right\}$. The class of the network intrusion data is obtained through the mapping function of $f_{HGHB}:T_a\to C_k$, and the sample data set is renewed. The multilayer Bayesian intrusion detection algorithm is as shown in Table 1.

In the algorithm as shown in Table 1, χ is the input data (network intrusion data set), $\mathfrak{A}_\chi$ is the observed value for n individuals, C represents the information decay estimate criterion when establishing the HGHB algorithm, t is used for restraining the information dilution threshold, and PartitioningAlg is the algorithm used to change the intrusion data set into the non-overlapping clusters. The a, b and cardmax are used to estimate the parameters for the base of latent variables.

5. Simulation Experiment and Analysis

In research literatures concerning the intrusion detection algorithm, related personnel often adopt the KDD CUP 99 intrusion detection database as the as-

Table 1. Algorithm Flow Title

Multilayer Bayesian Network Classification Algorithm
INPUT：： $\chi=\left(\mathbf{X}_1;\cdots;\mathbf{X}_p\right)$、 $\mathfrak{A}_\chi$、C、t、PartitioningAlg、 a, b, cardmax； OUTPUT： 1: nbw← p/s /* comput the numb of conti-wind */ 2: $DAG\leftarrow\varnothing;\theta\leftarrow\varnothing;L\leftarrow\varnothing;D_L\leftarrow\varnothing$ 3: for $i=1$ to nbw 4: $\xi_i\leftarrow\left\{\mathbf{X}_{(i-1)\cdot s-1};\cdots;\mathbf{X}_{i\cdot s}\right\};\mathfrak{A}\left[\xi_i\right]\leftarrow\mathfrak{A}\left[(i-1)\cdot s-1:i\cdot s\right]$ 5: $\left\{\cup_{j\in\xi_i}DAG_{univ_j},\cup_{j\in\xi_i}\theta_{univ_j}\right\}\leftarrow learnmodel\left(\xi_i\right)$ 6: $DAG_i\leftarrow\cup_{j\in\xi_i}DAG_{univ_j};\theta_i\leftarrow\cup_{j\in\xi_i}\theta_{univ_j}$ 7: $step\leftarrow1$ 8: while true 9: $\left\{\ell_1,\cdots,\ell_{\#\boldsymbol{C}}\right\}\leftarrow partition\left(\xi_i,\mathfrak{A}\left[\xi_i\right],PartitioningAlg\right)$ 10: if all clusters ℓ_q are singletons 11: break 12: otherwise 13: $\left\{\ell_1,\cdots,\ell_{j\#C_2}\right\}\leftarrow greaterthan\,one\left\{\ell_1,\cdots,\ell_{\#\boldsymbol{C}}\right\}$ 14: end if 15: $nbValidCluster\leftarrow0$ 16: for k=1 to #C2 17: $card_{LV}\leftarrow\min\left(size\,of\left(\ell_{jk}\right)+b;card_{max}\right)$ $\left\{DAG_{jk},\theta_{jk},L_{jk},\mathfrak{A}\left[L_{jk}\right]\right\}\leftarrow learnmodel\left(\ell_{jk},\mathfrak{A}\left[L_{jk}\right],card_{LV}\right)$ 18: /* validation of current cluster - see Subsection Controlling information decay */ 19: if $\left(C\left(DAG_{jk},\mathfrak{A}\left[\ell L_{jk}\right]\cup\mathfrak{A}\left[L_{jk}\right]\right)\geq t\right)$ 20: incr(nbValidClusters) 21: $DAG_i\leftarrow merge\_struct\left(DAG_i,DAG_{jk}\right)$ 22: $\theta_i\leftarrow merge\_parameter\left(\theta_i,\theta_{jk}\right)$ 23: $L\leftarrow L\cup L_{jk};\mathfrak{A}_L\leftarrow\mathfrak{A}_L\cup\mathfrak{A}\left[L_{jk}\right]$ 24: $\mathfrak{A}\left[\xi_i\right]\leftarrow\left(\mathfrak{A}\left[\xi_i\right]\setminus\mathfrak{A}\left[\ell L_{jk}\right]\right)\cup\mathfrak{A}\left[L_{jk}\right]$ 25: $\xi_i\leftarrow\left(\xi_i\setminus\ell L_{jk}\right)\cup L_{jk}$ 25: end if 26: end for 27: if (nbValidClusters = 0) 28: break 29: otherwise 30: incr(step) 31: end if 32: end while 33: $DAG\leftarrow DAG\cup DAG_i;\theta\leftarrow\theta\times\theta_i$ 34: end for

sumed object of attack. There are many literatures in this way. In order to make the transverse result comparison better with the existing research result of the intrusion detection algorithm, it also uses the ways of the above mentioned literatures for reference in this paper and applies KDD CUP 99 as the experimental subject. There exist many different types of network intrusion data in this intrusion detection database, thereinto the most representative ones are four types of attack data such as DOS, Probe, U2R and U2L, and their corresponding attributive features are as shown in Table 2.

When making the compassion and appraisal for the effect of the intrusion detection algorithm, the most used evaluation indexes in related literatures are mainly the detection rate and fallout ratio. The so-called detection rate refers to the ratio of the quantity of the intrusion data and the sum of the intrusion

data detected by the algorithm; the fallout ratio refers to the ratio of the quantity of the normal data misinformed as intrusion and the sum of the normal data. For the test data set selected in the above Table 2, select the classification results of the intrusion detections of HGHBID algorithm, INBID algorithm and ACO-SVMID algorithm and make the comparison as shown in Table 3.

Table 2. Selected Quantity of the Test Data

Attack Type	Training Sample	Test Sample	10%KDD Training Set Distribution
Normal	97277	60592	19.69%
DOS	391458	237594	73.91%
Probe	4107	4166	0.83%
U2R	55	73	0.01%
U2L	1126	8606	0.23%
Total	494020	311028	100%

Table 3. Comparison of Classification Results (%)

Intrusion Type	HGHBID		INBID		ACO-SVMID	
	Detection Rate	Fallout Ratio	Detection Rate	Fallout Ratio	Detection Rate	Fallout Ratio
U2R	96.21	7.09	91.32	9.86	90.85	8.96
DOS	98.35	4.16	91.47	5.67	93.82	6.18
Probe	98.77	3.04	93.28	4.67	96.39	3.61
U2L	89.18	20.55	74.59	25.72	71.91	28.10

The higher the detection rate, the comprehensive the detected intrusion data will be. The low fallout rate means the error ratio of the intrusion data detected by the algorithm is low. Of course the higher the detection rate and the lower the fallout ratio, it represents the performance of the algorithm is better. The detection rate is as shown in Formula (21) and the fallout ratio is as shown in Formula (22):

$$\text{Detection rate} = \frac{\text{Detected amount of the intrusion data}}{\text{Actual Sum of the intrusion data}}$$

$$\text{Fallout ratio} = \frac{\text{Detected amount of the false data}}{\text{Detected amount of the intrusion data}}$$

The simulation correlation data in Table 3 are the comparison experimental data of the detection algorithms such as HGHBID, INBID an in KDD CUP 99 database on the basis of the Formula (21) and Formula (22). It can be seem from Table 3 that in the classification recognition of the four type of network intrusion data of DOS, Probe, U2R and U2L, HGHBID algorithm is superior than INBID and ACO-SVMID in comparison in two evaluation indexes of detection rate and fallout ratio. Thus, the comprehensive defense effect of HGHBID algorithm is obviously better than INBID and ACO-SVMID. For the fallout ratio, INBID detection algorithm is better than ACO-SVMID algorithm in the detection effect in two types of intrusion data of U2L and DOS. Thus, the performances of INBID intrusion detection algorithm and ACO-SVMID intrusion detection algorithm have distinctive features respectively. We can get the inclusion through the data comparison that HGHBID intrusion detection algorithm is available and valid and it could effectively improve the recognition rate of the intrusion detection.

In order to further verify the performance advantage of HGHBID intrusion detection algorithm in the intrusion data recognition, the intrusion detection algorithms in two other literatures are selected to make the experimental comparison, PIDE and RIDS respectively. KDD CUP 99 database is still selected for the experimental database, the amount of the intrusion data selected in the simulation experiment is as shown in Table 2, and the simulation comparison result is as shown in Fig. 4 and Fig. 5.

It gives the simulation comparison result of the three types of algorithms of HGHBID, PIDE and RIDS in the detection rate in Fig. 4. It can be seen from Fig. 4 that the detection accuracy rate of HGHBID in-

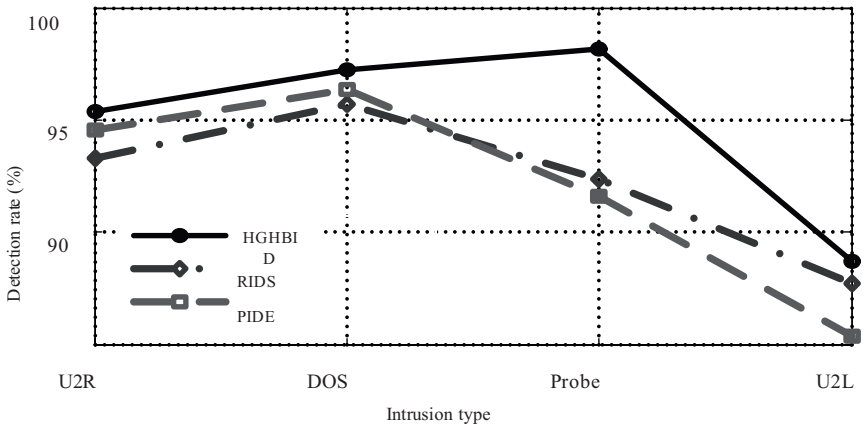

Figure 4. Detection Rate Comparison of Three Types of Algorithms

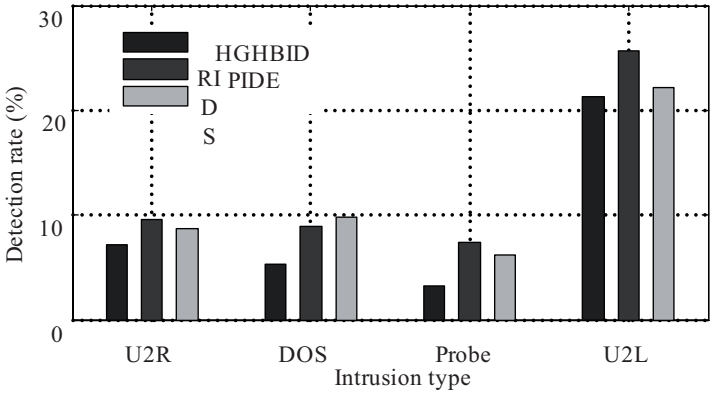

Figure 5. Fallout Ratio Comparison of Three Types of Algorithms

trusion detection algorithm on the four types of network intrusion data of DOS , Probe , U2R and U2L is superior than the algorithms of PIDE and RIDS . The comparison results of the other two algorithms in comparison are that in the two types of network intrusion methods of DOS and U2R , the detection rate of the intrusion data for the PIDE intrusion detection algorithm is superior than RIDS intrusion detection algorithm, while in the two types of network intrusion methods of U2L and Probe , the detection rate of the intrusion data for the RIDS intrusion detection algorithm is superior than PIDE intrusion detection algorithm. It gives the simulation comparison result of the three types of algorithms of HGHBID , PIDE and RIDS in the fallout ratio in Fig. 4. It can be seen from Fig. 5 that the detection error rate of the intrusion data of HGHBID algorithm in four attack methods are lower than that of the other two comparison algorithms, there into, in the three attack methods of U2R , Probe and U2L , the fallout ratio of the intrusion data of PIDE algorithm is higher than that of RIDS algorithm. In conclusion, the experimental result reveals that HGHBID algorithm is available and effective for intrusion data detection.

Conclusion

In this paper, it mainly creatively proposes a kind of Bayesian intrusion detection algorithm based on Gabor multilayer feature extraction through solving the existing problem of high fallout ratio and omission ratio in intrusion detection system. It designs multilayer Bayesian intrusion detection model and utilizes Gabor multilayer feature extraction algorithm to make the layered extraction for the data feature. The simulation experimental result in the standard intrusion detection database reveals that Bayesian intrusion detection algorithm based on Gabor multilayer feature extraction could effectively improve the precision ratio and recall ratio of the intrusion detection and satisfy the requirements in practical application.

Acknowledgements

The research is supported by .Hunan Province Natural Science Foundation Project No. 14JJ2124, and the scientific research project of Education Department of Hunan Province No. 14C0792.

References

1. A. Hamieh, J. Ben-othman, L. Mokdad. Detection of radio interference attacks in VANET. 2009 IEEE Global Telecommunications Conference, 2009, 6(9): 45-56
2. A. L Toledo, X. Wang. Robust detection of MAC layer denialof-service attacks in CSMA/CA wireless networks. IEEE Trans. Inf. Forensics and Security, 2008, 3(3): 347-358.
3. F. Borgonovo, A. Capone, M. Cesana, and L. Fratta. ADHOC MAC: New MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services. Wireless Netw, 2004 vol.10, pp. 359–366.
4. H. Hartenstein, K. P. Laberteaux. A Tutorial Survey on Vehicular Ad Hoc Networks. IEEE Communications Magazine, 2008, vol.46, no.6, pp. 164–171.
5. H. Omar, W. Zhuang, L. Li. VeMAC: A TDMA-based MAC protocol for reliable broadcast in VANETs. to be published.
6. Haiping Huang, Hao Chen, Ruchuan Wang, Qian Mao, Renyuan Cheng. (t, n) Secret Sharing Scheme Based on Cylinder Model in Wireless Sensor Networks. Journal of Networks, 2012, vol. 7, no.7, pp. 1009-1016.
7. J. Isaac, S. Zeadally, J. Camara. Security attacks and solutions for vehicular ad hoc networks. Communications, IET, 2010, vol. 4, no. 7, pp. 894–903.
8. Jiang, Dingde, Zhengzheng Xu, Peng Zhang, Ting Zhu. A transform domain-based anomaly detection approach to network-wide traffic.

Journal of Network and Computer Applications, 2014, 40: 292-306.
9. K. Pelechrinis, M. Iliofotou, and S. V. Krishnamurthy. Denial of service attacks in wireless networks: the case of jammers. IEEE Commun, Surveys and Tutorials, 2011, 13(2): 245-257.
10. Lv, Zhihan, Tianyun Su. 3D seabed modeling and visualization on ubiquitous context. In SIGGRAPH Asia, 2014, p. 33. ACM.
11. Lv, Zhihan, Liangbing Feng, Shengzhong Feng, and Haibo Li. Extending Touch-less Interaction on Vision Based Wearable Device. Virtual Reality (VR), IEEE. 2015.
12. M. Burmester, E. Magkos, V. Chrissikopoulos, Strengthening Privacy Protection in VANETs. in IEEE Int. Conf. Networking and Communications, 2008 (WIMOB '08), pp. 508–513, 2008
13. M. Hassan, H. Vu, and T. Sakurai.Performance analysis of the IEEE 802.11 MAC protocol for DSRC safety applications. IEEE Trans. Veh. Technol., Oct. 2011, vol.60, no.8, pp. 3882–3896.
14. Telecommunications and Information Exchange between Systems–Local and Metropolitan Area Networks–Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments, IEEE Standard for Information Technology, 2010.
15. United States Department of Transportation. Intelligent transportation systems. [Online]. Available: http://www.its.dot.gov/index.htm
16. W. Xu, W. Trappe, Y Zhang, and T. Wood. The feasibility of launching and detecting jamming attacks in wireless networks. 2005 ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2005, 6(8):34-42.
17. W. Xu, W. Trappe, Y. Zhang, and T. Wood. The feasibility of launching and detecting jamming attacks in wireless networks. MobiHoc 05, Urbana-Champaign, Illinois, USA May 25-27, 2005, 34(2):46-57.
18. Yishuang Geng, Jie He, Kaveh Pahlavan, Modeling the Effect of Human Body on TOA Based Indoor Human Tracking, International Journal of Wireless Information Networks (IJWIN) Dec. 2013, 20(4), 306-317,.
19. Yishuang Geng, Yadong Wan, Jie He, Kaveh Pahlavan, An Empirical Channel Model for the Effect of Human Body on Ray Tracing, 2013 IEEE 24nd International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), London, Britain Sep. 2013 pp. 47-52.