A Feature Selection Algorithm to Intrusion Detection Based on Cloud model and Multi-Objective Particle Swarm Optimization

Liu-Hong Zhou, Yan-Hua Liu, Guo-Long Chen
College of Mathematics and Computer Science Fuzhou University Fuzhou, China
Email: zoe_zlh@foxmail.com, lyhwa@foxmail.com,cgl@fzu.edu.cn

***Abstract*—there exist many problems in intrusion detection system such as large number of data volume and features, data redundancy and so on, which seriously affected the efficiency of the assessment. In this paper, we propose an approach called EFSA-CP to intrusion detection based on Cloud model and improved multi-objective Particle Swarm Optimization. The algorithm evaluates the characteristics of the attribute weights by the Cloud model and generates the optimal feature subsets which achieve the best trade-off between detection rate and rate of false alarm by MOPSO, which solves the problem of feature redundancy and helps improve the speed of the evaluation. Experimental results show that EFSA-CP can solve the feature selection problem of intrusion detection effectively. It can also achieve balanced detection performance on different types of attacks, with better convergence at the same time.**

Keywords-component; Cloud model; feature selection; intrusion detection; multi-objective particle swarm optimization

I. INTRODUCTION

Intrusion Detection System (IDS) is an important part in network security architecture. It determines the security state of network through the collection and analysis of the various threats, and logs them in a form of detailed record, which is an important basis to audit and evaluate the safety of network. However, due to various factors, the rate of false alarm is too high in current IDS, leading to a very large amount of IDS log data, with lots of redundant data. Therefore, it becomes an essential processing step to reduce the data size. Feature selection is widely used to achieve the reduction of massive data through reducing data dimension and speeding up the data classification.

In this paper, we present an approach called EFSA-CP to intrusion detection based on Cloud model and multi-objective Particle Swarm Optimization, MOPSO. Firstly, Cloud model generates different clouds and evaluates the characteristics of the attribute weights. After that MOPSO algorithm generates the optimal feature subsets which achieve the best trade-off between detection rate and rate of false alarm. Experimental results show that the model can not only solve the feature selection problem of intrusion detection effectively, but also achieve balanced detection performance on different types of attacks, with better convergence at the same time.

II. RELATED WORK

Some research work has been carried out on intrusion detection feature selection algorithm. The diversity of different learning algorithms was utilized. *Chebrolu*, *et al.* [1] investigated the performance of two feature selection algorithms involving Bayesian Networks (BN) and Classification and Regression Trees (CART) respectively. Both of the approaches considered the diversity of different learning algorithms for intrusion detection. So, the detection performance was promoted accordingly. *Optiz* [2] presented a genetic algorithm (GA) approach for searching for an appropriate set of feature subsets for ensembles. Using neural networks as the classifier, results showed better than the ensemble approaches of Bagging and Boosting. *Tsymbal*, *et al*. [3] presented an algorithm for building ensembles of simple Bayesian classifiers by using different feature subsets generated with the random subspace method.

Cloud model [4] was proposed by Academician *Li Deyi*, an uncertainty conversion model between the concept of qualitative and quantitative numerical by using qualitative description. *Du*, *et al.* [5] proposed a classification method based on cloud model. It effectively divided the domain of quantitative attributes into qualitative multiple cloud-based concepts according to the actual distribution of the data. *Zhu*, *et al.* [6] gave an approach with combination of cloud theory and support vector machines to solve classification of high-*dimensional* weight, using the cloud model to build attributes model of the training set.

The rest of the paper is organized as follows. In section 3, we briefly introduce the concept of cloud and some important definitions in Cloud model. Then comes the simple introduction of improved MO-PSO for feature selection. After that we describe our EFSA-CP algorithm. In Section 4, we present the experimental results and analysis of using EFSA-CP algorithm in intrusion detection. The paper concludes with Section 5.

III. THEORIES RELATED AND EFSA-CP ALGORITHM

A. The concept of Cloud

Definition 1(Cloud and Cloud droplet).Let U be a numerical representation of the quantitative domain, C is a qualitative concept on U, if the quantitative value $x \in U$ is a random realization of C, $\mu(x)$ the certainty of x to C, $\mu(x) \in [0,1]$, is a random number having a stable tendency

$$\mu : U \to [0,1] \quad \forall x \in U \quad x \to \mu(x) \tag{1}$$

So, the distribution of x on U is known as cloud, denoted by $C(X)$. Every x in Cloud is called as cloud droplet. If the

978-0-7695-4500-4/11 $26.00 © 2011 IEEE
DOI 10.1109/ISCID.2011.147

domain, the concept of corresponding, is an *n-dimensional* space, then it can be extended to the *n-dimensional*.

Cloud with 3 digital features of the expectation *Ex* (Expected Value), entropy *En* (Entropy), hyper entropy *He* (Hyper Entropy) to overall characterize a concept, denoted by *C (Ex, En, He)*. *Ex* is the point value which can best represent the qualitative concept in the number domain space. *En* is the measure of randomness of qualitative concept, which reflects the range of values accepted in the domain. *He* is the measure of uncertainty of *En*. It reflects dispersion degree and thickness of clouds.

Definition 2(*N-dimensional* backward cloud generator, *N*-CG^{-1}).It means to convert a certain amount of precise data to the digital features (Ex_1, En_1, He_1, Ex_2, En_2, He_2,…Ex_n, En_n, He_n), representation of a qualitative concept. For the *n-dimensional* sample $x(x_1,x_2,...x_{i...}, x_m)$, Sample point $x_i(x_{i1}, x_{i2}...x_{in})$, where $i=1,2,...,m$, satisfy the following conditions:

①Calculate *Ex* (Ex_1, Ex_2...Ex_n), expectation of the sample *x*, as

$$Ex(Ex_1,Ex_2\ldots Ex_n)=\left[\overline{X_1},\overline{X_2},...,\overline{X_n}\right]=\left[\frac{1}{m}\sum_{j=1}^{m}x_{j1},\frac{1}{m}\sum_{j=1}^{m}x_{j2},...,\frac{1}{m}\sum_{j=1}^{m}x_{jn}\right] \quad (2)$$

②Calculate *En* (En_1, En_2...En_n), standard deviation of the sample *x*, as

$$En(En_1,En_2\ldots En_n)=[\ \sqrt{\frac{\pi}{2}}\times\frac{1}{m}\sum_{j=1}^{m}\left|x_{j1}-Ex_1\right|,\sqrt{\frac{\pi}{2}}\times\frac{1}{m}\sum_{j=1}^{m}\left|x_{j2}-Ex_2\right|,..., \quad (3)$$

$$\sqrt{\frac{\pi}{2}}\times\frac{1}{m}\sum_{j=1}^{m}\left|x_{jn}-Ex_n\right|\ ]$$

③Calculate *He* (He_1, He_2...He_n), as

$$He(He_1,He_2\ldots He_n)=[\ \sqrt{S_1^2-En_1^2},\sqrt{S_2^2-En_2^2},.....,\sqrt{S_n^2-En_n^2}\ ] \quad (4)$$

$$S(S_1,S_2,...,S_n)=[\ \frac{1}{m-1}\sum_{j=1}^{m}(x_{j1}-\overline{X_1}),\frac{1}{m-1}\sum_{j=1}^{m}(x_{j2}-\overline{X_2}), \quad (5)$$

$$...,\frac{1}{m-1}\sum_{j=1}^{m}(x_{jn}-\overline{X_n})\ ]$$

B. The object membership degree of cloud

Definition 3(Object membership degree).*U* is a discourse domain in *n-dimensional* space, denoted by $U=\{u_1, u_2...u_m\}$, where u_i represents the object in *i-th* class, and u_{ij} represents the value of *j-th* dimension of u_i in feature space. X_{ij} expresses the degree of membership corresponding to *j-th* dimensional feature of u_i within the range of cloud, calculated as

$$X_{ij}=e^{\frac{-(u_{ij}-Ex_{ij})^2}{2En_{ij}^2}} \quad (6)$$

$$Ex_{ij}=\frac{1}{n}\sum_{j=1}^{n}u_{ij} \quad (7)$$

$$En_{ij}=\max(\frac{Ex_{ij}-u_{min_{ij}}}{3},\frac{u_{max_{ij}}-Ex_{ij}}{3}) \quad (8)$$

Ex_{ij} is the gravity of u_{ij}; En_{ij} describes the range of value accepted by cloud; u_{min} and u_{max} respectively represent the minimum point and maximum point of the *i-th* dimensional feature.

Let X_i be the degree of membership of u_i to cloud, calculated as

$$X_i=\frac{1}{n}\sum_{j=1}^{n}X_{ij}=\frac{1}{n}(e^{\frac{(u_{i1}-Ex_{i1})^2}{2En_{i1}^2}}+e^{\frac{(u_{i2}-Ex_{i2})^2}{2En_{i2}^2}}+\cdots+e^{\frac{(u_{in}-Ex_{in})^2}{2En_{in}^2}}) \quad (9)$$

Definition 4(Attribute distance). In *n-dimensional* disco-urse domain *U,* exist $u_i,u_j\in U$, let $d(u_{ik},u_{jk})$ expresses the distance between u_i *and* u_j in *k-th* dimensional feature space[7].

$$d(u_{ik},u_{jk})=\begin{cases}0, & if\ \ u_{ik}\subseteq u_{jk} or\ \ u_{jk}\supseteq u_{ik}\\ \dfrac{Ex_{ik}-Ex_{jk}}{3(En_{ik}-En_{jk})}, & if\ \ u_{ik}\not\subset u_{jk} or\ \ u_{jk}\not\subset u_{ik}\end{cases} \quad (10)$$

When $d(u_{ik},u_{jk})\geqq 1$, it shows that u_i disjoints from u_j and *k-th* dimensional attribute can be distinguished; When $d(u_{ik},u_{jk})=1$, the *k-th* dimensional attribute can not completely distingui-sh u_i from u_j.

Let d_k expresses the distance of the *k-th* dimensional feature, calculated as

$$d_k=\frac{1}{m}\sum_{i=1,j=1,i\neq j}^{m}d(u_{ik},u_{jk}) \quad (11)$$

Definition 5(Attribute weights). In the discourse domain *U*, let Aw_k be the weight of the *k-th* dimensional feature, calculated as

$$Aw_k=\frac{d_k}{\sum_{i=1}^{n}d_i} \quad (12)$$

Obviously, if d_k is smaller, the weight Aw_k will be smaller, which means that the *k-th* dimensional feature is more difficult to distinguish from others.

C. Multi-Objective Particle Swarm Optimization

Multi-objective Particle Swarm Optimization (MOPSO) is a population based paradigm to solve multi-objective optimization problems. MOPSO adopts the particle swarm optimization paradigm which in turn mimics behavior of the flocking birds. The MOPSO, similar to PSO, is based on a simple flight of the particle:

$$v_{ij}^t=w^t v_{ij}^{t-1}+c_1 rand_1(p_{ij}^{t-1}-x_{ij}^{t-1})+c_2 rand_2(g_j^{t-1}-x_{ij}^{t-1}) \quad (13)$$

$$x_{ij}^t=x_{ij}^{t-1}+v_{ij}^t \quad (14)$$

Where $\begin{cases}v_{id}=v_{\max}, & if\ \ v_{id}>v_{\max}\\ v_{id}=-v_{\max}, & if\ \ v_{id}<-v_{\max}\end{cases}$

Let the search space is set *d-dimension* and the scale of population is *n*; $\mathbf{x}_i^t=(x_{i1}^t,x_{i2}^t,...,x_{id}^t), i=1,2,...,n$ is the position of the *i-th* particle at time *t*; $\mathbf{v}_i^t=(v_{i1}^t,v_{i2}^t,...,v_{id}^t)$ is the velocity of the *i-th* particle at time *t*. Where w^t is inertia weight, c_1 and c_2 are the constant values that are called personal and global acceleration which give different important weight to personal or global term of (13) ; $rand_1$ and $rand_2$ are uniform random numbers from (0, 1) to give stochastic characteristics to the flight of particles; p_{ij}^{t-1} *i*s the personal best position of the *i-th* particle at time t-1, and g_j^{t-1} is the global best position at time *t-1*.

Definition 6(Population diversity).$S_{ij}(t)$ is the position vector similarity between particles x_i and x_j in *t-th* iteration.. Let *k* be the dimension, *n* the total dimension, then

$$S_{ij}(t)=\frac{1}{n}\sum_{k=1}^{n}(i_k==j_k\ ?\quad 1:0) \quad (15)$$

D (t) expresses the population diversity in *t-th* iteration, calculated as

$$D(t)=1-\frac{1}{n(n-1)/2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}S_{ij} \tag{16}$$

D. Encoding mode of characteristic particle

Feature selection is substantially to choose a subset of *n'* attributes from a sample of $n(n' \leqq n)$,. Therefore, a property of the sample can be defined as one-*dimensional* discrete binary variables of each particle, and *n* attributes constitute a discrete binary *n-dimensional* space of particle. For each particle, if the *i-th* bit is 1, it means that the *i-th* attribute is selected; the other hand means not being selected. So each particle represents a different subset of attributes. For example, particle *x* = {011000} represents the second and third properties are selected, the corresponding subset of attributes is {1, 3}.

E. Fitness function Design

In MOPSO, fitness function is used to evaluate mass of particles. The greater value of fitness function is, the better mass is. In this paper, the aims of our feature selection are

- The number of final attribute subset is as small as possible.
- The overall weight of attributes in subset is as large as possible, that means the distinction between the classified objects has better effect.
- The detection rate and rate of false alarm are as small as possible.

Therefore, we define that

$$f(x)_1=\frac{n-n(x)}{n_{\max}-n_{\min}} \tag{17}$$

$$f_2=\sum_{i=1}^{n}(H_i\times A_{wi}) \tag{18}$$

$$f_3=\frac{1}{n}\sum_{i=0}^{m-1}T_{ii}+1-\frac{1}{n}\left(\sum_{j=1}^{m-1}T_{0j}+\sum_{j=1}^{m-1}T_{j0}\right) \tag{19}$$

where $H_i=\begin{cases}1, & the\ attribute\ is\ chosen\\ 0, & the\ attribute\ isn't\ chosen\end{cases}\ (i=1,2,\ldots n)$,

$T_{ij}=\begin{cases}1, & i\ \text{detected as}\ j\\ 0, & i\ not\ \text{detected as}\ j\end{cases}\ (i,j=0,1,2,\ldots m-1)$

In (17), *n(x)* is the number of selected feature in the subset of particle *x*; *n* is the total number of attributes; n_{max} and n_{min} is maximum and minimum number of selected feature in the current particle population. In (18), A_{wi} is the weight of the *i-th* feature, whose calculating method is in (12). The first part of (19) is the detection rate; the third part of formula is the sum of rate of false alarm and false negative rate.

Therefore, the fitness function is defined as

$$f(x)=f(x)_1+f_2+f_3 \tag{20}$$

We can see that, *f(x)* will be larger when *n(x)* is smaller, the overall weight larger, detection rate much higher, false positive and negative rate smaller. That is, the chosen feature subset is up to our aims.

F. EFSA-CP Algorithm.

Based on the above analysis, we propose our EFSA-CP algorithm that is represented in the next subsection in detail.

Algorithm Representation

The main procedure of EFSA-CP algorithm is described as follows.

Input: Data set Data=*U*/*ind*(*D*)={u_1,u_2,u_3…u_m},*U* is discourse domain, *D* is decision attribute set，*n* is the number of attributes

Output: The feature subset.

Procedure:

1. Generate several clouds with *Ex*, *En* and *He.*
2. Generate a set of attribute weights, according to (14-16)
3. Generate initial population with scale, evolution generation and initial particle location
4. Update position of particles and get their own fitness according to (20)
5. Compare every particle's current fitness with their best personal fitness, if better, then update their personal best fitness
6. Compare every particle's current fitness with the best population fitness, if better, then update the population best fitness
7. If population diversity is lower than a certain threshold, mutation operation will be performed.

When it comes to termination condition, end; else go to 4.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset

Experiments have been carried out on data from KDD99 database, which is widely used to analyze IDS log data. Each sample in database contains 42 attributes, first 41 conditions attributes, where exists 9 discrete types, and 32 continuous type. In order to verify the validity of EFSA-CP algorithm, we randomly collect 10% of the dataset in KDD which contains 492 000 records. We choose a dataset comprises of 5000.

B. Experiment and analysis

(1)We use EFSA-CP algorithm to search for feature subsets of 6 kinds of dataset, the results shown in Table I. In the left column of table, ALL includes Normal class, ATTACK represents type of attack (including DOS, PROBE, R2L, U2R), the other as a separate type of attack. Right column is the subsets selected for each type of attack. For example, for U2R attack selected feature subset is 1, 2, 23, 31, where figures indicate the ID of the 41 features (initially 0). As we see from the table, properties of service and srv_count are the common properties of selected subsets, besides the number of feature subset is 5 or so, and the maximum number of selected features is for ALL.

Next, we build intrusion detection model respectively for the data without feature selecting and selected feature subsets. The results are shown in Table II from which we can see that, the model of EFSA-CP id better on detection accuracy rate. Table III shows detection rates of different algorithm. We can notice that EFSA-CP algorithm have much higher detection rate generally on 4 kinds of attacks.

TABLE I. ATTRIBUTE SUBSETS SELECTED FOR DIFFERENT KINDS

Type	Attribute Subsets selected
ALL	1,2,11,23,28,31{protocol_type,service,logged_

	in,srv_count,same_srv_rate ,dst_host_count }
ATTACK	1,2,23,28,31{protocol_type,service,,srv_count, same_srv_rate ,dst_host_count }
DOS	1,2,23,28,31{protocol_type,service,,srv_count, same_srv_rate ,dst_host_count }
PROBE	2,13,23,28,31{service, root_shell,srv_count, same_srv_rate ,dst_host_count }
R2L	2,17,23,28{service, num_file_creations,srv_ count,same_srv_rate }
U2R	1,2,23,31{protocol_type,service,,srv_count, dst_host_count}

TABLE II. DETECTION ACCURACY COMPARISON BEFORE AND AFTER SELECTING ATTRIBUTES

Type	Detection accuracy (%)	
	Before selecting	After selecting
NORMAL	87.8	94.05
DOS	84.9	91.85
U2R	85.4	94.0
R2L	87.5	90.86
PROBE	88.0	100

TABLE III. COMPARISON OF DETECTION RATES CORRESPONDING TO DIFFERENT ALGORITHMS

Type of Attack	Wenke Lee's(%)	BP network(%)	EFSA-CP(%)
DOS	79.9	92.71	91.85
U2R	75.0	48.0	94.0
R2L	60.0	95.73	90.86
PROBE	90.0	97.47	100

(2) Figure 1 shows the trends of fitness convergence on 6 kinds of dataset by EFSA-CP. It can be seen that fitness converges to the best fitness value when the algorithm runs at iteration of 200 or so. The fastest convergence is R2L while Attack the most slowly but the maximum fitness value. In general, the convergence of the algorithm is good. Figure 2 shows the population diversity curve of the above 6 datasets for selecting features. It can be seen obviously that the diversity remain at a high level above 90%, to better avoid the local convergence.

From the results of experiments (2) and (3), for those containing only attack types and those also including Normal type, EFSA-CP algorithm gets much bigger feature subset, smaller fitness value and faster convergence speed from the former dataset than that of the latter. Therefore, the algorithm processes much better on the data containing only attack types.

In conclusion, under the premise of effective classification, EFSA-CP algorithm can extract a smaller number of feature subset, which contains those important features reflecting the security of the system. Therefore, it improves the speed of detection and safety analysis.

V. CONCLUSIONS

For feature selection problems on IDS log data, this paper presents a feature selecting algorithm EFSA-CP, based on cloud model and multi-objective particle swarm optimization. The approach can get a more streamlined and accurate feature subset from high dimensional data. Experiments show that the algorithm can achieve the best trade-off between detection rate and rate of false alarm. It can also achieve balanced detection performance on different types of attacks, with better convergence at the same time.

ACKNOWLEDGMENT

This research has been supported by the National Natural Science Foundation of China under Grant No.60673161, the Technology Innovation Platform Project of Fujian Province under Grant No.2009J1007, and the Key Project of Fujian Provincial Department of Science & Technology under grant No.2007H0023.

REFERENCES

[1]Chebrolu S.et al. Feature deduction and ensemble design of intrusion detection systems. Computer & Security, 2004, 24(4): 295-307.

[2]Opitz DW. Feature selection for ensembles. In: Proc. of the 16th National Conf.on Artificial Intelligence(AAAI).Orlando:AAAIPress,1999.379-384.

[3]Tsymbal A.et al. Ensemble feature selection with the simple Bayesian classification. Information Fusion,2003,4(2):87-100.

[4]LI Deyi, DU yi. Uncertainty Artificial Intelligence. Beijing. National Defence Industry Press,2005

[5]DU Yi，LI Deyi. Concept Partition Based on Cloud and Its Application to Mining Association Rules. Journal of Software,2001,12(2):196-202

[6]ZHU Jie.et al.New cloud classifier based on SVM weight vector. Application Research of Computers.2009,6(6):2098-2100

[7]ZHANG Guo-ying.et al. Dimensional Cloud Model and Its Application in Multiple Attribute Evaluation. Transactions of Beijing Institute of Technology, 2004,24(12):1065-1069

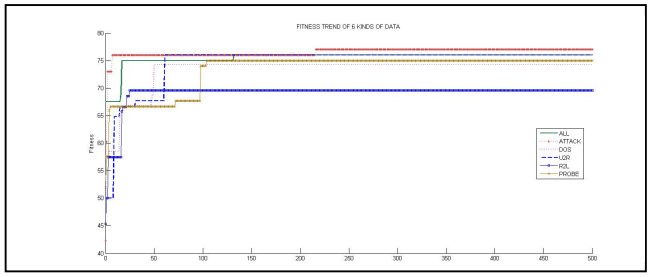

Figure 1. Fitness convergence trend of EFSA-CP

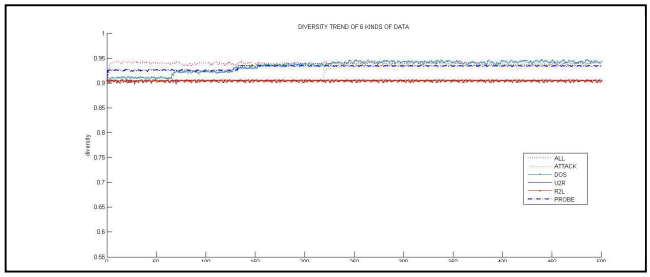

Figure 2. Population Diversity of EFSA-CP