ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE
International Journal of Advanced Research in Computer and Communication Engineering
ICITCSA 2017
Pioneer College of Arts and Science, Coimbatore
Vol. 6, Special Issue 1, January 2017

A New QPSO Based Network Intrusion Detection System using Feature Selection

V. Attchara[1], K. Sujitha[2], S. Sayina[3]

Assistant Professor, Dept. of Computer Applications, Pioneer College of Arts and Science, Coimbatore, India

PG Scholar, Dept. of Computer Science, Pioneer College of Arts and Science, Coimbatore, India [2,3]

Abstract: As the Internet services spread all over the world, many kinds and a large number of security threats are increasing. Therefore, intrusion detection systems, which can effectively detect intrusion accesses, have attracted attention. This paper proposes a novel approach for feature selection based on Genetic Quantum Particle Swarm Optimization (GQPSO) attribute reduction in network intrusion detection which aiming to problem of classification algorithm with low detection speed and low detection rate in high dimensional network data intrusion detection. In the approach, selection and variation of genetic algorithm with QPSO algorithm are combined to form GQPSO algorithm; normalized mutual information between attributes defined as GQPSO algorithm fitness function to guide it's reduction of attributes to realize optimal selection of network data feature subset. KDD99 data-set are used to experiment. The experimental result shows that the approach is more effective than QPSO and PSO algorithms in discarding independent and redundancy attributes. As a result, intrusion detection rate and speed of classification algorithm are greatly heightened by using the method.

Keywords: Genetic Quantum Particle Swarm Optimization(GOPSO); Normalized Mutual Information; Attribute reduction; Intrusion Detection; Feature Selection.

I. INTRODUCTION

Many kinds of systems over the Internet such as online shopping, Internet banking, trading stocks and foreign exchange, and online auction have been developed. However, due to the open society of the Internet, the security of our computer systems and data is always at risk. The extensive growth of the Internet has prompted network intrusion detection to become a critical component of infrastructure protection mechanisms. Network intrusion detection can be defined as identifying a set of malicious actions that threaten the integrity, confidentiality, and availability of a network resource [1], [2].

Intrusion detection is traditionally divided into two categories, i.e., misuse detection and anomaly detection. Misuse detection mainly searches for specific patterns or sequences of programs and user behaviors that match well-known intrusion scenarios. While, anomaly detection develops models of normal network behaviors, and new intrusions are detected by evaluating significant deviations from the normal behavior. The advantage of anomaly detection is that it may detect novel intrusions that have not been observed yet. While accuracy is the essential requirement of an intrusion-detection system (IDS), its extensibility and adaptability are also critical in to- day's network computing environment [3]. Currently, building an effective IDS is an enormous knowledge engineering task. System builders rely on their intuition and experience to select the statistical measures for anomaly detection. Experts first analyze and categorize attack scenarios and system vulnerabilities, and hand-code the corresponding rules and patterns for misuse detection. Because of the manual and ad hoc nature of the development process, such IDS has limited extensibility and adaptability. And also In network intrusion Detection, independent and redundancy attributes leads to low detecting rate and speed of classification algorithms. Therefore, how to reduce network attributes to raise performance of classification algorithms by applying optimal algorithm has become a research branch of intrusion Detection. Srinoy, S.[1] and Tiejiu etal[2] applied PSO algorithm is applied in network intrusion detection feature selection combined immunity thought with PSO to reduce attribute set of network data, the method keeps particle varieties to a certain degree and raise convergence accuracy WANG Shi-yi [4] presented an method for feature selection in network intrusion based GQPSO.

This new approach for network intrusion detection feature selection based on GQPSO attribute reduction is presented in this paper is more effective in discarding independent and redundancy attributes and greatly raises intrusion detection rate and speed of classification algorithm.

II FEATURE SELECTION BASED ON GQPSO ATTRIBUTE REDUCTION

A. Feature Selection

Feature selection can be considered an important asset in building classification models as some data may hinder the classification process in a complex domain. Moreover elimination of useless features enhances the accuracy of detection while speeding up the computation. Thus, feature selection improves the overall performance of the detection mechanism. A few data mining techniques have used feature selection techniques. The simplest approach

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

consists of removing one feature at a time and testing the performance of a classification algorithm against the removed features. This approach was used by Mukkamala and Sung [9] and was tested with two different classification algorithms: Support Vector Machines (SVMs) and Artificial Neural Networks (NNs). Another more efficient approach to feature selection is proposed by Chebrolu, Abraham, and Thomas [10]. The authors proposed two different approaches: Bayesian networks and Classification and Regression Trees (CARTs) From our experiments done with feature selection, we have observed that feature selection contributed to improve overall accuracy, reduced the number of false positives, and improved the detection of instances with low frequency in the training data.

B. Quantum-Behaved Particle Swarm Optimization

Particle Swarm Optimization (PSO) algorithm [9, 10], originally introduced by Kennedy and Eberhart in 1995 [5], is an evolutionary computation technique motivated by the simulation of social behavior. A PSO system, in which individuals (particles) representing the candidate solutions to the problem at hand fly through the n dimensional space to find out the optima or sub-optima, got more and more attention according to its explicit mechanism and simple calculation, with the position vector and velocity vector of particle being represented as

Xi(t) = (Xi1(t), Xi2(t),……. Xin(t) and

Vi(t) = (Vi1(t), Vi2(t),……. Vin(t) respectively.

For the particle i and iterative j generation is calculated as follows

v(t+1) = ω. vi(t)+ rand1().c1.(pbesti - xi(t)) + rand2().c2.(gbest - xi(t))…….(1)

xi(t+1) = xi(t) + vi(t+1)…… ……(2)

where i=1,2,….,N; t=1,2,….,L As shown in above equations, where V_i(t) and V_i(t+1) denote current velocity and modified velocity of particle i, which are restricted in the interval [−Vmax ,Vmax] ; rand1() and rand 2() are random functions whose values are between 0 and 1; c1 and c2 are called the acceleration coefficients for each term; $pbest_i$ is the best previous position (the position giving the best fitness value) of particle i known as the personal best position (pbest); gbest is the position of best particle among all the particles in the population and is known as the global best position (gbest); x_i (t) and x_i (t+1) denote the current position and modified position of particle i ; j denotes current iteration number; ω is the inertia weight which is introduced by Shi and Eberhart in order to accelerate the convergence speed of the algorithm [6].

C. Selection and Variation of Genetic Algorithm

Aiming to low convergence speed in later stage of iteration and falling into local optimum of QPSO algorithm, selection and variation of genetic algorithm is introduced into QPSO. It's basic thought is that fitness value of particle i (FV_i) in population is compared with average fitness value (AFV) of all of particles. If FVi is more than AFV , particle i is preserved, otherwise, selection and variation is done in every bit of particle i according to random possibility P_m.

D. Structure of Fitness Value Functions

Fitness value function is essential to performance of intelligent optimization algorithm; therefore, normalized mutual information based on joint entropy in rough theory is used for estimate standard. It's thinking is to select an attribute set in which correlation degree between condition attributes and category attributes is higher and correlation degree category attributes is lower. If X and Y and serve as two attributes, correlation degree between them is measured by the following formula.

SU(X,Y) = 2 * I (X;Y) / H(X) + H(Y) ……..(1)

In formula(1),

I (X;Y) = H(X)+H(Y) – H(X,Y) is mutual information between X and Y; H(X) is entropy function and is defined as

H(X) = -∑ P(ai) log2 p(ai)……(2)

H(X,Y) is joint entropy of X and Y and is defined as

H(X,Y) = -∑∑ P(ai, bj,) log2 p(ai, bj)…(3)

Therefore, fitness value function of GQPSO attribute reduction algorithm is defined as

∑SU (Xj,c) / SQRT(∑∑ SU(Xi, Xj)……(4)

E. Encoding Method for Particles

Attribute reduction is to discard independent and redundancy attributes in attribute subset under the condition of keeping classification ability of original dataset unchanged. Therefore, every attribute is defined as a discrete binary variable and M attributes consist of discrete binary space of M dimension. In every particle, if bit i is 1, this means that attribute i is selected, otherwise, it is not selected bit i is 0.

For example, particle j =1010100 means that attributes 1, 3, 5 are selected and attributes 2,4,6,7 are not selected .as a result, optimal attribute subset is {1, 3, 5}.

F. Description of Algorithm

Input: high dimensional network data, maximum iterative times T

Output: optimal attribute subset

1) Particle population, pbest , gbest are initialized ;
2) Fitness value of every particle is calculated according to formula (4);
3) Fitness value of every particle is compared with pbest , if the former is superior to the later, the former serves as pbest ;
4) Pbest of every particle is compared with gbest , if the former is superior to the later, the former serves as gbest ;
5) Renewal of position of particle population is done according to formulas (5), (6), (7);
6) Selection and variation is done;

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

7) If iterative times is more thanT , operation goes to 8), otherwise, it goes to Step2;
8) Optimal position of population is converted to reduced attribute subset.

III RESULTS AND DISCUSSIONS

A. Experimental Dataset

The KDD-Cup99 data set from UCI repository [8] is widely used as the benchmark data for IDS evaluation. The KDD-99 data consists of several components, can be seen in TABLE I.

TABLE I . BASIC CHARACTERISTICS OF THE KDD DATA SET

Data set lable	Total	Normal (%)	DOS (%)	Probe (%)	U2R (%)	R2L (%)
10% KDD	494,020	19.79	79.2	0.8	0.01	0.2
test KDD	311,029	19.58	73.9	1.3	0.02	5.2
Whole KDD	4,898,430	19.8	79.3	0.84	0.001	0.02

As in the case of the International Knowledge Discovery and Data Mining Tools Competition, only the '10% KDD' data is employed for the purposes of training. This contains 22 attack types and is essentially a more concise version of the 'Whole KDD' data set. In our experiments, we apply its 10% training data consisting of 494 021 connection records for training. Each connection record represents a sequence of packet transmission starting and ending at a time period, and can be classified as normal traffic, or one of 22 different classes of attacks.

The test data set has not the same probability distribution as the training data set. There are 4 new U2R attack types in the test data set that are not present in the training data set. These new attacks correspond to 92.90 %(189/228) of the U2R class in the test data set. On the other hand, there are 7 new R2L attack types corresponding to 63% (10196/16189) of the R2L class in the data set. In addition there are only 104(out of 1126) connection records present in the training data set corresponding to the known R2L attacks present simultaneously in the two data sets. However there are 4 new DoS attack types in the test data set corresponding to 2.85% (6555/229853) of the DoS class in the test data set, and 2 new Probing attacks corresponding to 42.94% (1789/4166) of the Probing class in the test data set.

IV. CONCLUSION

Aiming to problem of classification algorithm with low detection speed and low detection rate in high dimensional network data intrusion detection, a novel method for network intrusion detection feature selection based on GQPSO attribute reduction is proposed in the paper. The method realizes optimal selection of network intrusion features by discarding independent and redundancy attributes. Experimental results show that classification detecting rate and detecting speed of GQPSO algorithm is higher than those of PSO and QPSO algorithms.

REFERENCES

[1] Zadeh, L.A., 1994. "Fuzzy logic, neural networks, and soft computing. Commun," ACM 37, 1994, 77–84.
[2] Cho, S.-B., "Incorporating soft computing techniques into a probabilistic intrusion detection system," IEEE Trans. Systems Man Cybernet. vol. 2, pp. 154, 2002.
[3] Kosko B, "Neural network and fuzzy systems," New Jersey Prenticl Inc., 1992.
[4] Sun, J., Feng, B., Xu, W.B. "Particle Swarm Optimization with Particle Having Quantum Behavior," in: Proceedings of 2004 Congress on Evolutionary Computation, Piscataway, NJ, pp. 325-331, 2004.
[5] Kennedy, J., Eberhart, R.C. "Particle Swarm Optimization," in: Proceedings of IEEE International Conference on Neural Network, IV. Piscataway, NJ, pp.1942-1948, 1995.
[6] Shi, Y.H., Eberhart, R., "A Modified Particle Swarm Optimization," in: Proc. IEEE International Conference on Evolutionary Computation, Anchorage, Alaska, pp. 69-73, 1999.
[7] Clerc, M., Kennedy, J., "The Particle Swarm: Explosion, Stability and Convergence in a Multi-Dimensional Complex Space," IEEE Transaction on Evolutionary Computation, vol. 6, pp. 58-73, 2002.
[8] UCI Machine Learning Repository (Online), Available: http://www.ics.uci.edu/~mlearn/MLRepository.html.
[9] Systems with Applications, vol. 2, PART 1, pp.2097-2106, March 2009.
[10] Kuo I-Hong, Horng Shi-Jinn Kao, Tzong-Wann, Lin Tsung-Lieh, Lee Cheng-Ling, Terano Takao, Pan Yi, "An efficient flow-shop scheduling algorithm based on a hybrid particle swarm optimization model," Expert Systems with Applications, vol. 3, PART 2, pp.7027-7032, April 2009.
[11] Srinoy. S, Intrusion Detection Model Based On Particle Swarm Optimization and Support Vector Machine, Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Security and Defense Applications, Honolulu, Hawaii, USA, 2007.
[12] Zhou Tiejun, Li Yang, and Li Jia, Research on intrusion detection of SVM based on PSO, Proceedings of the 2009 International Conference on Machine Learning and Cybernetics, Baoding, China, 2009.
[13] Tian. Wenjie, and Liu Jicheng, Network intrusion detection analysis with neural network and particle swarm optimization algorithm, Proceedings of 2010 Chinese Control and Decision Conference, CCDC 2010, Xuzhou, China, 2010.
[14] Bahrololum. M, Salahi. E, and Khaleghi. M, Machine learning techniques for feature reduction in intrusion detection systems:A comparison, Proceedings of ICCIT 2009 - 4th International Conference on Computer Sciences and Convergence Information Technology, Seoul, Korea, Republic of, 2009. Computer, 2010, 27(1). (in Chinese)
[15] KDD99Cupdataset[DB/OL].[2010-07-07]. http://kdd.ics.uci.edu/database es/kddcup99/kddcup99.Html.
[16] Cortes C, and Vapnik V, Support vector networks, J. Machine Learning, 1995, 20(3).