Network Intrusion Detection Method Based on Improved Particle Swarm Optimization Algorithm in Cloud Computing Platform

Yi Wenzhou

School of Information Engineering, Guangdong Vocational and Technical College of Engineering, Guangzhou, Guangdong, 510520, China

email: security7506@163.com

Keywords: Cloud Computing, Intrusion Detection System, Mr GA-BP Mean Method, Hadoop

Abstract: The existing intrusion detection algorithm, based on the concept of parallel genetic algorithm and BP neural algorithm, is based on the cloud computing platform of network intrusion detection algorithm Mr real-time large-scale intrusion detection data for detection. Compared with the existing intrusion detection system, the efficiency and accuracy are greatly improved. Through two groups of comparative experiments, the optimization algorithm Mr GA - BP average algorithm is realized, which proves the performance advantage of intrusion detection system.

1. Introduction

The existing intrusion detection system does not meet the requirements of today's large-scale data, and there are limitations and bottlenecks in detection accuracy, efficiency and accuracy [1]. In the future, intrusion detection will have efficient detection efficiency and improve service quality on the basis of improving efficiency in the short term. In order to collect the efficiency of node processing activities used by the cloud environment of intrusion detection environment activities, and greatly improve, timely warning can upload the source code of the data collected by the cloud for intrusion detection analysis. The traditional intrusion detection algorithm (BP neural algorithm) can adopt the idea of parallelization. In the cloud environment, the distributed processing of data sources is deployed. If the distributed intrusion detection is implemented, the above problems can be solved. It can not only expand the modern application field of cloud computing, but also improve the performance of existing intrusion detection algorithm to another level, and upgrade the development of intrusion detection to a new level.

2. Establishment of Network Intrusion Detection Algorithm

2.1 MR GA-BP

The network topology BP neural network of the algorithm is determined by the network layer, the number of nodes, the activation function, the initial weighting factor, the learning algorithm and the number of system errors. These decisions need to follow the following principles [2]. According to the past experience, the three layers of BP neural network are considered as the priority, that is, the output layer, the input layer and the hidden layer. The number of nodes in each layer is related to many factors, such as the characteristics of sample data, the form of transfer function, the number of input / output nodes. The initial weight coefficient determines the initial weight, which is randomly generated in a specific category. In general, the initial weight is assigned between 0 and 1. End condition. Determining the number of individuals in the population: when the number of individuals in the population is small, the speed of GA algorithm can be improved, but the diversity of individuals in the population is greatly reduced. The efficiency of the algorithm is reduced. According to the previous research, generally 20 ~ 500. Coding method: real number coding; fitness function: use BP neural algorithm to obtain each sample, and set the inverse proportion of the sum of squared errors of all samples after execution as fitness function, selection operator: betting method; crossover operator: real number coding mutation operator: This article takes the value of

  DOI: 10.25236/ictmic.2020.074

0.05. The system uses genetic algorithm to optimize the weight of BP neural network and control the number of evolutions. When the number of evolutions reaches the maximum, it ends.

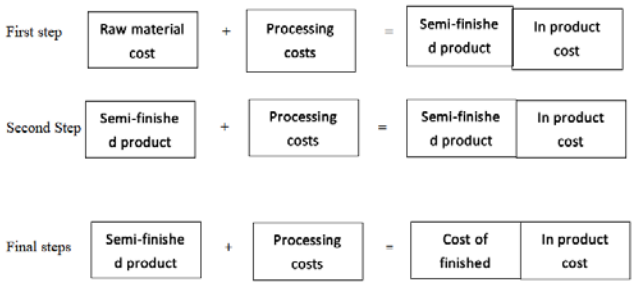

Figure 1 Cost calculation procedure of step-by-step method

2.2 The Idea of Parallelization

When dealing with a large number of data, in order to deal with serial data, using the traditional idea, if the processing time and efficiency spend more time, so the initial data modularization, these data modules of the machine need to be divided for parallel processing[3]. Because the processing between them is not related, the processing efficiency is greatly improved. Each computing node has a complete network. Moreover, the initial state of the network is the same. Reflect parallelization in training. Each node acquires some sample data to train BP neural network. The computer nodes satisfy some convergence conditions, and are induced. After collection, judge whether to perform the next repeated processing.

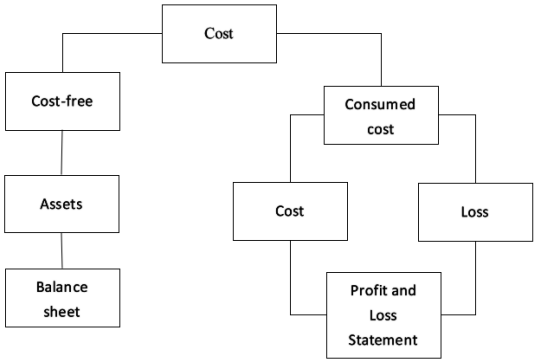

Figure 2 Chart of cost and cost

2.3 Research on Security of Virtualized Network Environment in Computing Technology

Virtualization security innovation is the first step. For the traditional physical environment, its security mainly depends on the traditional intrusion detection methods[4]. However, due to the security differences between cloud computing and traditional systems, traditional intrusion detection methods are difficult to effectively apply to cloud computing virtualization environment. The security of cloud computing is closely related to the security of virtualization technology. Because all the services provided by cloud computing come from the virtual machine of the server. Only by protecting users' virtual machines can we overcome the security problems of cloud computing. Therefore, it is very important to provide intrusion detection service for cloud computing security. In order to ensure the safe operation of virtual devices, if the traditional intrusion detection system is directly applied to the virtual environment, the corresponding intrusion detection system must be installed on all virtual devices of the same cloud server. Typically, each server has multiple virtual machines. The above solutions lead to a great waste of computer resources of cloud service providers and greatly reduce the overall performance of the cloud platform. In addition, cloud users with different security levels need to set different security policies for each virtual machine. This makes people doubt the difficulty of security management of cloud platform. At present, the feasible solution is to deploy intrusion detection system in privileged virtual machine. This virtual machine is responsible for detecting the intrusion of each unprivileged

virtual machine on the same server. The security of this deployment method is higher than that of all user virtual machines [5]. Because the virtual machine is detected in the virtual machine manager, it is difficult for users to find the virtual machine. Although the creation and elimination of virtual machines are dynamic, the one-way visibility of virtual machines in the virtual machine manager can be used to obtain the deployment and status information of users' virtual machines. In this way, intrusion sensing virtual machine can dynamically detect user's intrusion behavior in virtual machine.

Table 1 Characteristic of basic method of product cost calculating

Cost calculation method	Cost calculation object	Cost calculation period	The calculation of product cost at the end of the term	Production characteristics
Variety method	Product Varieties	Monthly calculation, consistent with the accounting reporting period	Generally, no calculation is needed in single-step production, but in multi-step production, it is necessary to calculate.	Large-scale single-step or multi-step production
Batch method	Product Batch	Irregular calculation, consistent with production cycle	Generally, no calculation is required.	One-piece, small-batch, one-step or multi-step production
Step-by-step method	Product variety and its steps	Monthly calculation, consistent with the accounting reporting period	Need to calculate	Mass multistep generation

3. Mr GA-BP Algorithm Description

BP neural network has strong adaptability to learning and mapping [6]. In practical application, the combination with genetic algorithm can avoid the shortcomings of BP. Based on the shortcomings of current methods, MapReduce, GR BP average method is suggested. Using the distributed computer system, based on the sample output corresponding to the weight of each sample, rather than output the change of each weight. In order to form a matrix and add the corresponding changes of weights of all samples, the average changes of each weight are obtained. After learning, change a weight, change a new weight, as the initial weight of the next learning sent to each node. The maximum number of learning stages reaches the stop, or the learning error of each sample reaches the appropriate range.

4. The Main Viewpoints of Mr GA - BP Average Algorithm are as Follows.

In order to use BP genetic algorithm to optimize the weight of BP neural network globally, a better search space is found in the solution space [7]. Then GA algorithm is used to optimize and weight GA repeatedly. Find the most appropriate solution and weight for the train in this small solution space. In order to perform the above algorithm uniformly for intrusion detection, intrusion detection is organically combined. The weight randomly generated by HDFS in the stored scatter table. In order to optimize the initial weight of map neural network, optimize the smaller solution space and improve the convergence rate, MapReduce GA is used. In order to train the correct times of maprespesebp neural network training, please start maprespesebp neural network training to make the total error of samples reach the allowable range or reach the preset repetition frequency. After training, the experimental results were compared statistically.

5. System Design and Implementation

5.1 Overall Structure of the System

The overall solution of IDS is mainly for current quality data [8]. In the past, the intrusion

detection system can't pass a large number of data quickly and instantaneously. Because of a large number of data, the weight adjustment process needs a huge program operation, and the most important thing is that the measurement rate is also very low. The shortcomings of the previous intrusion detection are solved by using the average algorithm of Mr GA BP proposed here. The cloud based intrusion detection system is divided into intrusion detection data source monitoring, intrusion detection data source preprocessing, data source storage and intrusion detection modules.

5.2 Data Source Acquisition

In the data source acquisition phase, the components commonly used to acquire data sources are transceivers, agents, and adapters. The data source is mainly from host, network and log.

5.3 Data Source Preprocessing

Data preprocessing is very important to the whole process, because it needs to provide data sources for subsequent intrusion detection and analysis [9]. Using high quality data sources can improve the efficiency of the whole process. This study completed the follow-up processing of BP neural network based on Hadoop platform, and because of the need for specific data format in the training process, so in the data preprocessing stage, it needs to be further processed and converted into BP. At this stage, the preprocessed data source is directly stored in the Hadoop distributed file system. Therefore, data source preprocessing first removes redundant fields and redundant formats from the source data.

6. Convergence Speed and Efficiency Test of the Algorithm

Comparison items: MapReduce BP, Mr GA - BP algorithm, Mr - BP. See the method proposed in this experiment. Data source: complete data set, 708.2 MB, KDD cup 99 data set [10]. The experimental results show that MrgA BP average method can improve the training speed more than MrgA BP. This study suggests that mapressebp neural network training will be completed on Hadoop platform. GA can shorten the training time and intrusion detection time of BP neural network.

7. Conclusion

Big data security has always been the focus of attention. With the growth of business, a single node computing platform can not deal with the security problem of adding a large amount of data. In order to improve the calculation efficiency and detection accuracy, the distributed computing platform must be used. This article provides a solution. The core of intrusion detection algorithm is to use Mr GA average algorithm and the idea of parallelization. The optimal weight is obtained by genetic algorithm. Then start the neural network training. The process adopts the Hadoop framework of distributed computing platform, and the genetic algorithm and neural network algorithm are implemented in the cloud computing platform. At the same time, in order to improve the efficiency and accuracy of intrusion detection, the algorithm is improved.

References

[1] Gunasekaran Manogaran, Daphne Lopez, Naveen Chilamkurti. In-Mapper combiner based Map-Reduce algorithm for big data processing of IoT based climate data. Future Generation Computer Systems, vol. 86, 2018.

[2] Y. Zhu, L. Li, Y. Song,. Storage and Parallel Processing of Big Data of Power Equipment Condition Monitoring on ODPS Platform. Transactions of China Electrotechnical Society, vol. 32, no. 9, pp. 199-210, 2017.

[3] José Francisco Colom, David Gil, Higinio Mora,. Scheduling framework for distributed intrusion detection systems over heterogeneous network architectures. Journal of Network &

Computer Applications, vol. 108, 2018.

[4] Song Deng, Aihua ZHOU, Dong YUE,. Distributed intrusion detection based on hybrid gene expression programming and cloud computing in cyber physical power system. Iet Control Theory & Applications, vol. 11, no. 11, , pp. 1822-1829, 2017.

[5] Khaled Aldebei, Xiangjian He, Wenjing Jia,. SUDMAD: Sequential and unsupervised decomposition of a multi‐author document based on a hidden markov model. Journal of the Association for Information Science & Technology, vol. 69, no. 2, 2018.

[6] Mohamed Idhammad, Afdel Karim, Mustapha Belouch. Distributed Intrusion Detection System for Cloud Environments based on Data Mining techniques. Procedia Computer Science, vol. 127, pp. 35-41, 2018.

[7] Hodo, Elike, Bellekens, Xavier, Hamilton, Andrew,. Threat analysis of IoT networks Using Artificial Neural Network Intrusion Detection System. Tetrahedron Letters, vol. 42, no. 39, pp. 6865-6867, 2017.

[8] Syed Ali Raza Shah, Biju Issac, Seibu Mary Jacob. Intelligent Intrusion Detection System Through Combined and Optimized Machine Learning. International Journal of Computational Intelligence & Applications, no. 4, pp. 1850007, 2018.

[9] Wenjuan Li, Weizhi Meng, Lam-For Kwok,. Developing advanced fingerprint attacks on challenge-based collaborative intrusion detection networks. Cluster Computing, vol. 21, no. 3, pp. 1-12, 2018.

[10] Xingshuo An, Jingtao Su, Xing Lü,. Hypergraph clustering model-based association analysis of DDOS attacks in fog computing intrusion detection system. EURASIP Journal on Wireless Communications and Networking, vol. 2018, no. 1, 2018.