Design and Implementation of Anomaly Detection System for Cloud Platform Based on Multiple Attribute Information

Yu Yong-Wu
College of Computer Science, Neijiang Normal University
Neijiang, 641000, China
yyw7467@163.com

ABSTRACT: *With the rapid development of computer, people are increasingly demanding the quality of the mainstream cloud platform which is main forms of computer industry. In cloud monitoring system, picking up information, monitoring data anomalies, abnormal transmission and alarm should be vigilant. By studying the theory of cloud platform monitoring technology, this paper uses FASTMASOD algorithm to design the node machine, and uses the host computer to test the anomaly detection system of cloud platform in the context of multi-attribute information. Finally, the ROC curve is used to verify the effect of the algorithm in the anomaly detection system. The research of this paper not only lays a theoretical foundation for the anomaly detection of multi-attribute information, but also has very important significance to the research of cloud platform.*

Keyword: Cloud Platform, FASTMASOD Algorithm, Anomaly Detection System, ROC Curve, Multi-Attribute Information

Received: 29 April 2017, Revised 2 June 2017, Accepted 5 June 2017

1. Introduction

With the rapid development of computer technology, the superiority and reliability of virtual technology function has become the focus of academic circles, and the development of virtual technology has become a catalyst for the development of cloud computing technology. Based on a variety of computing technologies, cloud computing came into being and has been rapidly developing. In the cloud computing platform, the back end is composed of three clouds including storage, desktop and application, and data, operating systems and applications are generally deployed in the data center. Based on this structure, people only need to pay attention to the data, and could realize real-time data access, data processing and data sharing. Therefore, the essence of cloud computing is to integrate the global data and information [1]. With the further research and application of cloud technology, the security of cloud platform has become a key factor considered by users. Platform monitoring and anomaly detection are the priority among priorities to ensure a timely and accurate cloud computing services, and are also the premise of load management, and whether cloud platform virtual machine could run normally or not is determined by the use of each physical node resource. The utilization ratio of CPU, I/O, and relevant information of memory and anomaly detection of

useful information are important issues in this research [2]. According to these problems, this paper designed and implemented a multiple attribute anomaly detection system under the background of cloud platform, and the system includes: data collection, data processing, anomaly detection and abnormal transmission, failure warning. These parts are linked together and could fully meet the new requirements of cloud monitoring and have a good effect on abnormal detection [3].

2. Related Research based on Cloud Monitoring Technology

2.1 The Concept and Characteristics of *"Cloud Computing"*

As a kind of new and high technology, cloud computing does not has a perfect definition now. And it is usually interpreted as: as a model, cloud computing stores software and hardware resources in shared and distributed way, and could share all the information resources in the resource pool in real-time, and resources could also supply and release rapidly. Characteristics of cloud computing are flexible self-service, network access anywhere, measurement payment, resource sharing; its architecture includes core services, management services and access interface provided for users, as shown in the figure below. The figure shows the three service models composed of infrastructure, platform, software [4].

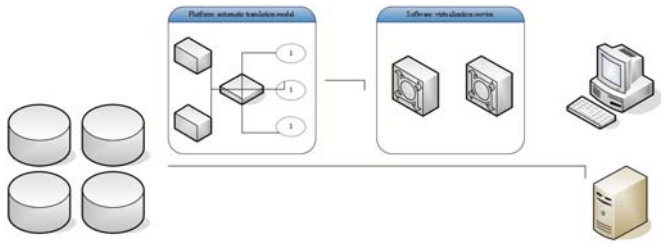

Figure 1. Cloud Computing Architecture

2.2 Anomaly Detection of Multiple Attribute Information

A large number of data have the characteristics of multiple attributes, so in the process of anomaly detection, it has the characteristics of large scale, self-organization, dynamic and reliability. Therefore, the anomaly information detection rate is facing newer and higher requirements. At present, the methods of multi attribute information anomaly detection include spatial temporal correlation data compression, time dependent characteristics of node data, compression algorithm of wavelet transform and distributed compression algorithm. These algorithms have changed the traditional computing way, through the computer or the workstation operation, it alleviates the node limited storage spaceÿalleviates the hindrance of network consumption. Among them, FASTMASOD algorithm is one of the reasonable and effective solutions to the anomaly detection [5].

3. Design of Cloud Platform Anomaly Detection System based on Multi Attribute Information

3.1 System Structure

The architecture consists of three parts: global monitoring, local monitoring and monitoring agent. Each level consists of a combination of different numbers of nodes. The node at the top level can send information to the next level, the next node as the third node can be accepted at the same time. The information can isolate faults and reduce network load in the transmission process of cascade structure. The system architecture diagram of this paper is as follows:

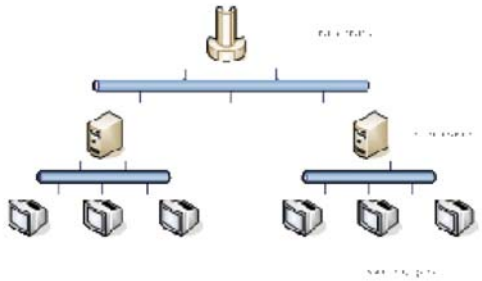

Figure 2. Anomaly Detection System Architecture of Xen Cloud Platform Based on Multi-attribute Information

3.2 Algorithm Preparation

Before the research, this paper reasonably plans the multi attribute information on the cloud platform, and put all the records on a dimension. The foundation of this dimension is to realize spatial index. The foundation of this dimension is to realize spatial index. In this paper, the research and discussion of multi-dimensional spatial objects with multi attribute information are improved by building spatial index structure. At present, the index structure has grid file, *K-D* tree, *B* tree and *R* tree, etc. This paper uses *R* tree with multi-dimensional attributes.

R tree index structure is composed of the main leaf points and non-leaf points. The boundary rectangle of *R* tree structure is achieved by overlapping values of all leaves. Each node's maximum index option *A* and the minimum index option a meet 2< = *a* < =*A*. a indicates the space availability of disk storage. *A* indicates that each node cannot exceed the disk page. The application of *R* tree structure in computer is as follows:

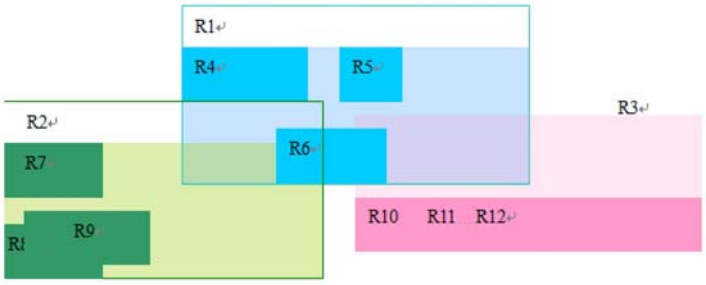

Figure 3. *R* tree structure

R tree structure has the main structure node *R*1, *R*2 and *R*3, after each main node, there are child nodes. Among them, *R*4, *R*5 and R6 are the child nodes of *R*1. *R*7, *R*8 and *R*9 are child nodes of *R*2. *R*10, *R*11 and *R*12 are child nodes of *R*3. There is a cross between nodes, that is, multiple information can be combined together. Considering the height of the R tree balance, it can not only quickly find the object operation, but also achieve forced insertion and deletion, and its automatic recovery function is very powerful.

3.3 Detection Algorithm

In this paper, FASTMASOD algorithm is used to detect abnormal data of cloud platform. FASTMASOD algorithm is a method that it excludes some objects which are not likely to appear in the outlier space, and then detects the spatial outliers [6]. Because of this characteristic, FASTMASOD algorithm can save running time and improve detection efficiency.

3.3.1 Prerequisite

If there is a distance l, and the distance l comparison function is *f*, the expected value is *q*; there is a neighborhood area comparison function:

$$f(1)=\frac{(f(1)-q)^{T}(f(1)-q)}{q^{T}q}$$

In the upper form, *f* (l) indicates coefficient value that the spatial data point l is similar to the central function around the space. Among them, when the *f*(l) value is smaller than the threshold value, there is data point l that belongs to the spatial region. Through the detection of these redundant data object points, it can meet the needs of spatial outlier detection and improve the operating efficiency of the system.

3.3.2 Workflow

Before anomaly data detection, we need to transform the attributes of the data into numerical attributes to lay the foundation for the computation of the data. At the same time, dividing data points into spatial and non-spatial attributes. After that, the framework of spatial tree is established to realize spatial index and determine the interval of data points attribution. Next, remove the spatial data, which is professional pruning, streamlined structure. Finally, the minimum unbiased estimate q of the comparison

function is calculated, and the similarity coefficient of the comparison function is obtained.

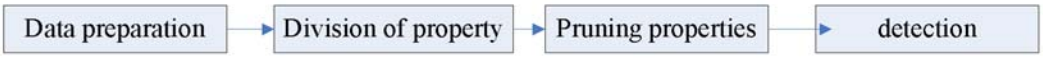

Figure 4. Anomaly detection algorithm process

Detection algorithm consists of four parts: prepare data, attribute partition, type pruning, and anomaly detection of cloud platform with multi attribute information. First, input space data set X, $X = \{1_1, 1_2, 1_3, ...,1n\}$, the corresponding attribute values for the data are $S = \{\eta_1, \eta_2, \eta_3,....., \eta_n\}$.The output data is detection results.

3.4 Design and implementation of node machine

1. Function Design

The core of the whole anomaly detection system is the node machines, and the function mainly consists of four parts: information collection, Information arrangement, data storage and deletion, anomaly detection and early warning. Each node has established UDP in order to communicate with the host, and the function flow chart is shown in Figure 5. First node machine goes on line and the system starts and sends command to master control machine. After receiving the command, the main control machine send a message to node machine, entering the background.

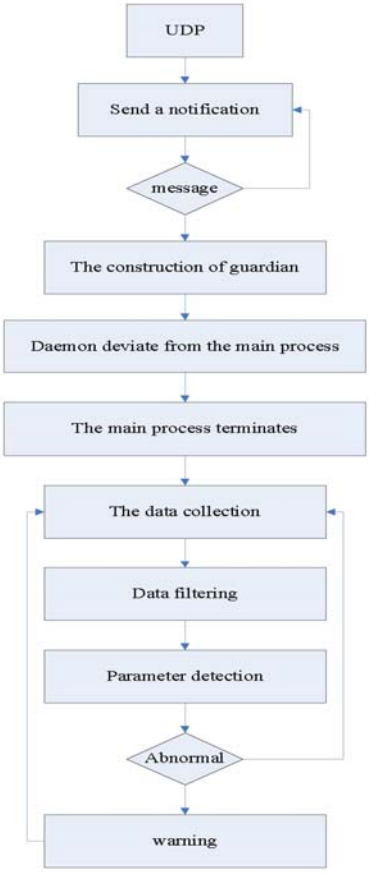

Figure 5. Centos data node function frame

2. Design of Background Guard Process

In this paper 5 virtual machines (1 MCU, 4 node machines) are used to form a guard process [7]. Through the background guard process finish the functions of master control machines and node machines and the node machine could finish the functions including data collection, analysis, prediction and sending abnormal information and the master machine completes the functions

including starring of node machines, alarming, storing information. The flow chart is shown in Figure 6.

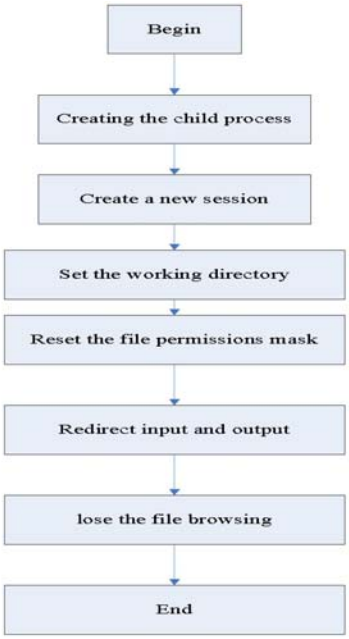

Figure 6. Create a daemon process flow chart

3. Design of UDP Communication Module

This paper uses the UDP protocol to finish the communication. In the detection system, node machine starts and the master machine would get a message and send messages to the node machines to realize the monitoring, but when more than one node machines send the on-line message, it could not be guaranteed that the master machine could receive all the information, so improvements are made based on the traditional UDP. A confirm retransmission would be done after the node machine sends on-line message in order to reduce the loss. Detail communication processes are shown in figure Figure 7.

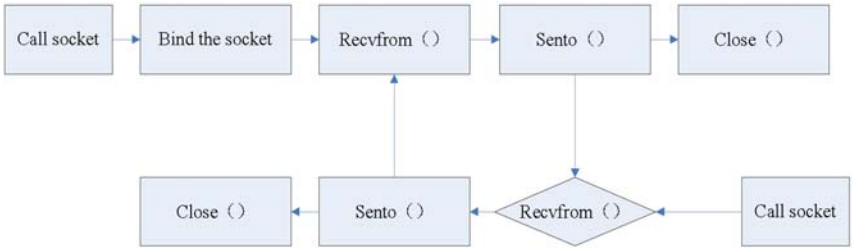

Figure 7. UDP Communication Process

4. Design of Data Persistence

In this paper, Mysql database is used for real-time data storage. Compared with other databases, Mysql database is relatively simple, easy to operate, and could use the tool of Nevicat for Mysql to operate directly on the Mysql database, and the data is presented in the form of visual presentation, so it is able to ensure the data persistence. But in the process of saving data to the database, because the Mysql database also has capacity constraints, a large amount of data will cause the database to collapse, so this paper designs the data deletion module and sets data retention period. Taking into account the update of the data, the retention period would be set as a week, after a week the data in the database would be deleted, that is to say, which is always stored in the database is the latest data within a week.

3.5 Design and Implementation of Master Control Machine

The execution framework of the master control machine is shown in Figure 8, the content of this part is the man-machine interface. Sending and receiving of each command is finished by the master machine, and the processes are the same with node machine, so not tired in words here [8]. There are three main modules: sending the message of starting (master control machine waits for the mesage from node machine and after receiving the message sends the message information. At the moment, node machine begins to collect data and detect abnormal information), abnormal alarm (the main function is to receive the abnormal information from node machine and when there is abnormal information, sound the alarm, otherwise go into dormancy), information storage and release (the module is storing abnormal information, through the interface to monitor information real-time, or to obtain monitoring content of virtual machine through the command input).

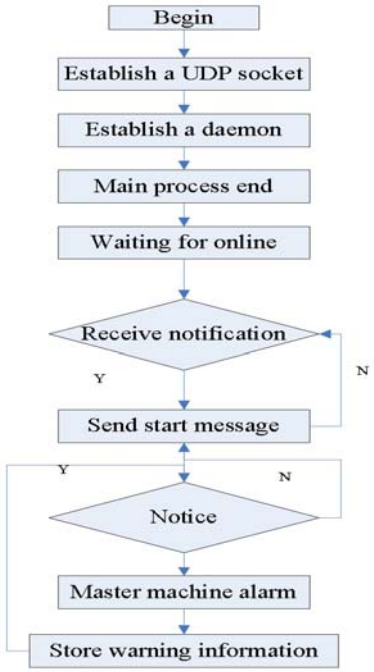

Figure 8. Centos Name Node Function Frame

4. Implementation and Verification of the System

4.1 Implementation of Test System

4.1.1 Construction of Test Platform

Considering a variety of factors, the testing platform in the paper is carried out on the Hadoop platform in the Xen virtual machine, so the platform construction is divided into two parts, Xen and Hadoop.

Xen is a virtual machine with a less performance consumption and its installation is finished based on the Linux system; Hadoop could realize the login between virtual machines without password through the SSH configuration, and the configuration information is for the purpose of master control machine to view information of node machine, especially in the case of abnormal warning.

4.1.2 Data Collection

In this paper, the Sysstat is employed to collect data. Sysstat is a very powerful tool for collection and could completely meet the data requirements in this paper, and the data collected is stored in the text file. Because the Linux system is different from the Windows system, so it would occur that memory would increas sharply along with program on and would reduce drastically along with program off. Linux will allocate the rest of the memory to cached, but cached is different from free. When Linux is in the operation, the free memory decreases, and the system automatically turns free into cached.

4.1.3 Operation Analysis

In this paper, the experiment is operated based on the Hadoop platform of Xen virtual machine, and by 4 node machines and 1

master control machine to test alarm sensitivity of the system.

First integrate the detection algorithm into the detection system, and then run the gurad process of node machine and master control machine. Experiment results are as follows.

Name	IP	Time	Because
CentosDataNode3	192.168.1.105	18:06:45	Worry
CentosDataNode1	192.168.1.115	18:06:51	Worry
CentosDataNode1	192.168.1.101	19:24:02	Worry
CentosDataNode2	192.168.1.111	20:24:03	Worry

Table 1. Failure Message

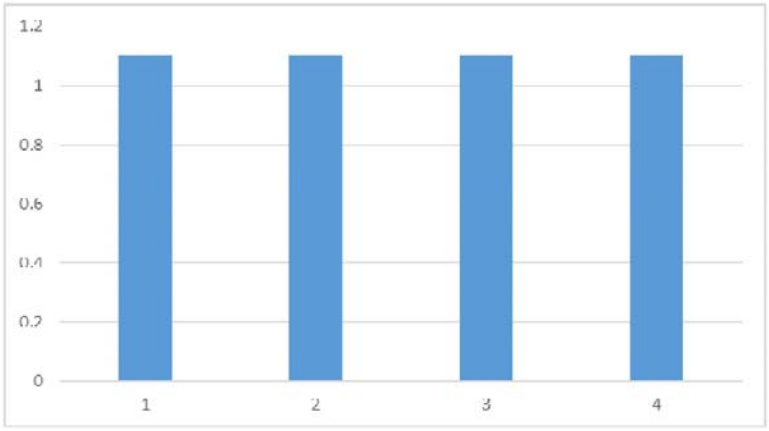

Figure 9. Alarm Sensitivity Test Results

From the above figure we can see that the alarm time delays, and the main reason is that seeding message needs time, leading to a slight lag of the alarm point, which is inevitable error. Because of the limited experimental conditions, experiments are conducted on four node machines. With the increase of node machines, the delay would increase, but compared with the advance of anomaly detection time, it will not cause any impact on the system.

4.2 Evaluation of Test Systems

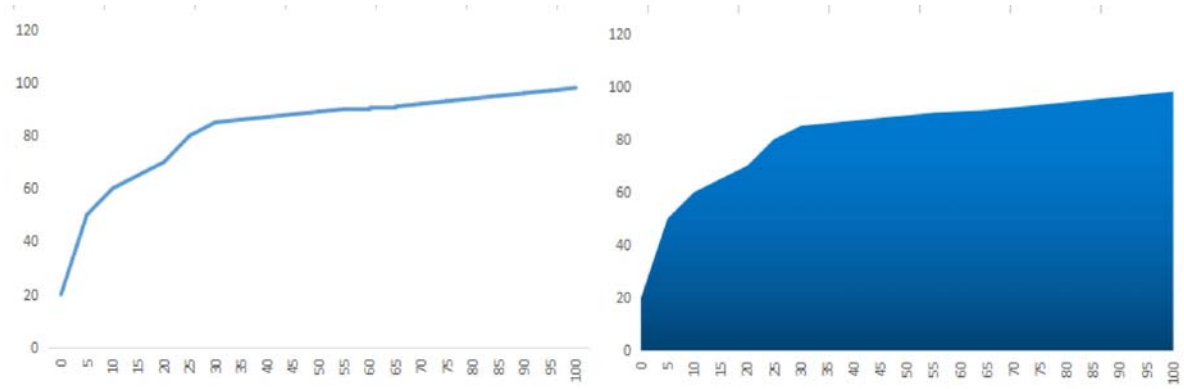

Figure 10. Alarm Sensitivity Test Results

System evaluation criteria are mainly composed of two parts, one is the higher system anomaly data detection rate, the other is the lower detection error value [9]. Among them, the detection rate is the proportion of abnormal data that occupy all abnormal data in spatial data. False alarm rate is the ratio of normal data occupying all normal data. In this paper, a simple ROC curve is used to detect the anomaly detection system of cloud platform in the context of multi-attribute information:

The area of the ROC curve represents the degree of optimization of the algorithm. From the above figure, the area of the graph has reached more than 80%, FASTMASOD algorithm has a good effect in the context of multi-attribute information cloud platform anomaly detection system detection [10].

5. Conclusion

Along with the wide application of "cloud computing", "*cloud security*" has become the focus of attention, so the importance of cloud platform monitoring is self-evident. Through various researches on cloud platform monitoring, the anomaly detection system of cloud platform under multiple attribute background is proposed in the paper. It reduces multi-attribute in order to process data. Judge the abnormal information by using the FASTMASOD algorithm. In the limited experiment condition, with the increase of nodes, the delay is increased. But compared with the time of anomaly detection, it does not have any effect on the systems. Finally, through the ROC curve verification, the algorithm has a good effect above 80%, and can detect the cloud platform anomaly detection system in the context of multi-attribute information effectively. The system has also laid a theoretical foundation in the application of related industries.

Development Suggestions

In this paper, the detection of cloud platform anomaly detection system in the context of multi-attribute information can detect the abnormal data. But the time is still in the stage of rapid development, and information technology is constantly updated and improved during the period. The application of anomaly detection algorithm is becoming more and more common, and the further development is the need of development. The algorithm in this paper realizes the detection of multi attribute information anomaly data, and avoids the problem of low accuracy and poor value selection. But in future research, the author should further improve the accuracy of the experiment and select the appropriate threshold. Faced with such challenges, the recommendations are as follows: 1. According to the actual situation, use reasonable threshold. Threshold provisions not only related to the level of local technology, but also closely related to the specific circumstances of the relevant industry. By combining the relevant literature and research on the industry, do research preparation on the basis of existing standards. 2. Jiahe GIS technology takes full account of the temporal granularity existing in spatial state, based on the research of spatial attributes, combined with time attribute research. 3. Expanding research needs of data sets. Data growth showed a straight upward trend, which requires the realization of large capacity data, large specifications and large components in the next study.

References

[1] Leung. H., Chen, S. (2014). Anomaly detection for cloud monitoring.

[2] Hongyan, Y. U., Cen, K., Yang, T. (2015). Design and implementation of abnormal behavior detection system in cloud computing, *Journal of Computer Applications*.

[3] Huang, Q., Lee, P. P. C. (2014). LD-Sketch: A distributed sketching design for accurate and scalable anomaly detection. *In:* network data streams, *In:* Proceedings - IEEE INFOCOM, 420-1428.

[4] Zhang, X. Z., Huang, C. Y., Zhang, Z. B. (2014). Design and Implementation of Intrusion Detection System Experiment Based on Snort. *Research & Exploration in Laboratory*.

[5] Jehangiri, A.I. (2015). Distributed Anomaly Detection and Prevention for Virtual Platforms.

[6] Memon, V. I. M. (2014). A Design and Implementation of New Hybrid System for Anomaly Intrusion Detection System to Improve Efficiency, *International Journal of Engineering Research & Applications*.

[7] Arthur, M. P, Kannan, K. (2015). Cross-layer based multiclass intrusion detection system for secure multicast communication of MANET in military networks, *Wireless Networks*, 22 (3) 1035-1059

[8] Moss, D. J. M., Zhang, Z., Fraser, N. J. (2015). An FPGA-based spectral anomaly detection system, *In:* International Conference on Field-Programmable Technology. IEEE.

[9] Silvestre, G., Sauvanaud, C., Kaaniche, M. (2014). An Anomaly Detection Approach for Scale-Out Storage Systems, *In:* IEEE International Symposium on Computer Architecture & High Performance Computing. IEEE, 294-301.

[10] Inc. G. (2016). Systems and methods for anomaly detection and guided analysis using structural time-series models.