Research of Self-Set for Cloud Security Immune System Based on Bloom Filter

Huo Lin
Guangxi University
Nanning, China
nnxhy@163.com

Zhou Jingxiong
Guangxi University
Nanning, China
jaysonzhou18@qq.com

Xiao Yuchuan
Guangxi University
Nanning, China
xiaoyuchuancq@163.com

***Abstract*—In order to improve the performance of Cloud Security Immune System(CSIS), the representation and storage of the feature data of self-set is studied. Firstly, it proposes a Bloom Filter Self-Set storage model (BFSS), which can reduce the storage space of self-set effectively and save query time. Secondly, in order to support the delete operation of the feature data, it put forwards a Counting Bloom Filter Self-set storage model (CTBFSS). Finally, to reduce the storage space of self-set and reduce querying time further, it raises Compressed Bloom Filter Self-Set storage model (CPBFSS).Experiments demonstrate that BFSS, CTBFSS and CPBFSS can reduce the time cost between customer and service effectively on the premise of higher recognition rate, speed up the search procedure and improve the overall performance of the system.**

Keywords—Cloud Security; Human Immune; Bloom Filter

I. INTRODUCTION

Since the concept for cloud computing has been put forward in 2007, just a few years' time, from theoretical research to platform development and practical application has seen a considerable progress. However, cloud computing not only brings convenience to people, but also brings a great deal of data security risks. It is confronted with the problem that how to realize the user data security in the environment with massive nodes and massive information processing demand.

In order to protect the safety of data operation under the cloud environment better, Literature[1] proposed a CSIS based on human immune theory, in order to improve its performance, this paper studies the representation and storage of the self-set feature data for the system, introducing the excellent data structure named Bloom Filter on the CSIS to improve storage the self-set feature data, and the improved self-set storage model can effectively reduce the time cost of the customer service and improve the overall performance of the CSIS with the high recognition rate.

II. INTRODUCTION OF CSIS

Cloud computing security problems and problems encountered for HIS has a lot of similarity. First of all, both of them have similar structural components, the human immune system is a complex giant system composed by the mass of cells, cloud computing is a high-complexity network composed by the mass of nodes; Secondly, the function for the two systems is similarity, both of them have the functions for self-stabilizing, monitoring, learning, clear alien invasion, to ensure security and stability of internal environment; Finally, both of them have similar running manner, HIS runs with the operation of manner with fluid circulation, using B cells and T cells to recognize and kill pathogens, cloud computing is in a continuous manner to monitor the accessed nodes, after found the abnormal situation immediately deal with it and clear exception. Therefore, drawing the operation mechanism of HIS, it has a reliable theoretical support for building CSIS to solve the problem of data security in the cloud. Research of this paper is based on the CSIS, using Bloom Filter technology to research on improvement its storage of self-set feature data.

III. ANALYSIS FOR BLOOM FILTER TECHNICAL

Bloom Filter (BF)[2] is a binary string vector data structure proposed by Howard Bloom in 1970, it has a very good space and time efficiency, its function is to be used to detect an element weather is a member of the set. As a simple information representation scheme, Bloom Filter can well meet the demand for resource efficient interaction and research on the high-speed network, particularly suitable for distributed systems in data storage and efficient query.

A. Algorithm principle

The standard Bloom Filter is described as follows: data set $S=\{s_1,s_2, \cdots,s_i, \cdots,s_n\}(1 \leqslant i \leqslant n)$, there are n elements, through the hash functions $H_1,H_2, \cdots,H_j, \cdots,H_k(1 \leqslant j \leqslant k)$ which range from 0 to m-1 mapped to the bit sequence of length m BF vector, hash functions are independent. Principle as shown in Fig.1.

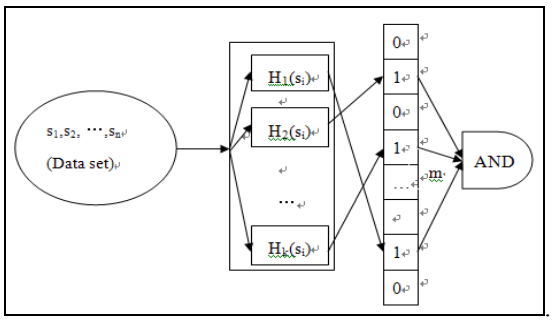

Fig.1 Standard Bloom Filter

Standard Bloom Filter can implement the insertion and query operation of the elements, there are false positive questions during the query, and the schematic diagram of the element insertion and query operation is shown in Fig.2.

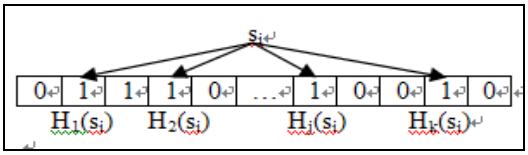

Fig.2 Insertion and Query of element Bloom Filter

When inserting operation, call k hash functions to hash the elements separately, and then set the corresponding BF bit string values to 1.When querying operation, also firstly called k hash functions hash the elements separately, and then query the BF bit string whether the value of the operation result corresponding bit is 1, as long as there value 0, the element is not set in.

Literature[3] demonstrated when $k = \lfloor (\ln 2)(m/n) \rfloor$, the error rate of the query is minimum. For example, when m=1280 bits, n=100, K takes 9 can make the false positive rate get to the minimum 0.0021. The impact of such a low rate of miscarriage of justice in the network monitoring system will be very small.

B. *Units Improvement research and typical application*

Since such an excellent data structure raised, improvement research based on it has been carried, a lot of research have been emerged. There are more typical of the extended algorithm: Counting bloom filter[4] solved the question which elements cannot be deleted; Compressed Bloom filter[5] using arithmetic coding technique and the ultimate entropy to compress BF vector as a message transfer in distributed system, it can obtain higher compression rate and the error rate will be reduced; others improvement research are: Spectral Bloom filter query algorithm[6], Split type Bloom filter query algorithm[7], and so on. A lot of improvement research shows that: Bloom Filter is a very potential technology, and its application is very wide..

Bloom Filter is mainly played two roles in the following practical application:

(1) Due to the Bloom Filter has the characteristics of saving space, which can be as a compression of the small size of the data collection used to replace the original large-scale data, complete the judgement weather the element is in the data set, to express large data sets and improve the searching efficiency. This kind of application mode mainly concentrates in the aspects such as database operation, Dictionary inquiry and file operation, resource route, packet route, network intrusion detection and so on.

(2) The large scale raw data set is abstracted as abstract information with Bloom Filter algorithm, and then the summary information is transmitted in other distributed nodes such as database. The most typical of this application is the transfer of the contents of the file directory information to each host in the distributed system. Will file directory information in form of a bloom filter, storage to the DRAM in node, because of the concise Bloom Filter storage, which can not only speed up file search process, and can reduce the interaction of bandwidth consumption.

IV. RESEARCH OF SELF-SET BASED ON BLOOM FILTER

"Self-set" feature data is a very important part for CSIS, it is an important judgement whether to activate the immune mechanism of cloud security or not in the running process of the monitoring node.

A. *Self-set definition*

CSIS to run the HIS as the theoretical basis, referring HIS concept of "self" and "non-self", "self-set" and "non-self set" is defined as follows on the CSIS:

Definition 1: "Self-set" is refers to the cloud server nodes and under normal operating circumstances, CSIS and the server node extracted feature data sets that normally feature data.

Definition 2:"Non-self set" is refers to the cloud computing nodes or servers suffered an invasive or internal fault, in the system operation under abnormal conditions, the CSIS on node and server to extract the feature data set, namely the anomaly characteristics of the data set.

The representation and structure of self-set feature data is an important factor for the performance of the immune system, and the relationship between self-set and CSIS is shown in Fig.3.

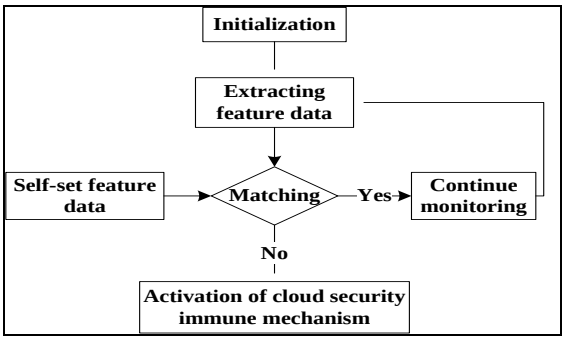

Fig.3 Relationship between "self-set" and cloud security immune mechanism

Among them, the characteristic data is by server nodes or to extract some of the field is composed of binary strings, composed of destination IP address and a service ID, with good structure and performance[8], which represents as shown in Fig.4.

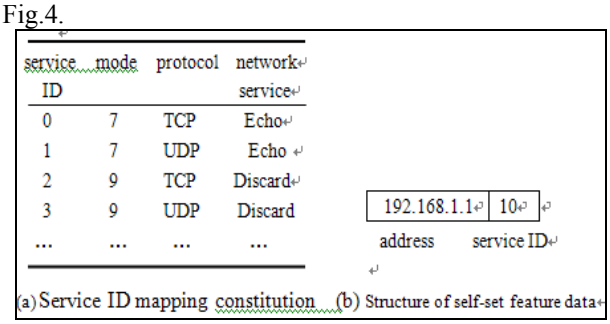

Fig.4 Represent of self-set feature data

Data structure using self-set feature of this structure represents a relatively simple and easy, but it is relatively large. Therefore, the server can be designed large-capacity memory to store them, because only once on the server side storage, so even designed storage capacity larger database is also possible, in the nodes only need to store and run capacity very small BF string.

B. Self-set construction algorithm

Positive selection algorithm is well suited for the generation of self-set in AIS. Self-set in CSIS is used to construct the algorithm, and the algorithm flow is shown in Fig.5:

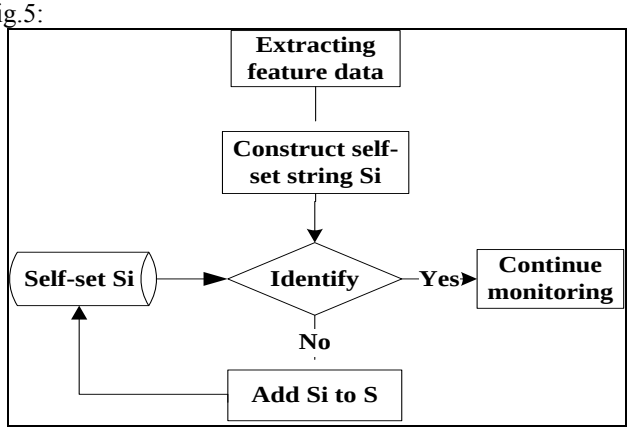

Fig.5 Process of self-set construction algorithm

After the completion of training, a self-set feature data string S which describes the normal network behavior of the node is obtained, and the network behavior of the nodes is monitored by the collection S. For network node receives and sends the service behavior, the query on a set S is corresponding to judge the behavior of network services belong to the normal network service request, or belong to the abnormal network service request.

C. Improvement of self-set structure algorithm based on Bloom Filter

A very important research content in CSIS is how to determine whether the node operation is abnormal or not. Its server and client ideas and Bloom Filter distributed in a typical application idea is the same, all the nodes in order to reduce storage overhead and reducing bandwidth overhead transfer interaction, saving search time, speed up the search process, saving resources to improve system performance. Therefore, we proposed that the Bloom Filter data structure is applied in CSIS, constructs a kind of heavy server and light client self-set storage pattern.

The construction process as follows, the cloud computing system the normal operation of system server and an access node of data feature extraction, using PSA for constructing and training self-set feature data set $S=\{s_1,s_2,s_3,\ldots,s_n\}$, the S stored in the server side of the database; Then, the BF vector string of length m is constructed by counting Bloom Filter algorithm; Finally, the compression BF vector string is compressed by the compressed Filter Bloom algorithm, and the BF vector is pushed to the client node.

Because access to the cloud node running environment is very complex and constantly changing, which run the same way and HIS are continuous, thus self-set feature data collection is dynamic, it should be able to insert data, query and delete operations. Standard Bloom Filter algorithm supports the insertion and query elements, but does not provide the element deletion. Therefore, we use counting Bloom Filter type to improve it, the improved vectors generated BF series can support query elements, insert and delete operations.

The Counting Bloom Filter to generate the BF vector on every value is zero or not. Therefore, we are able to compress the BF vector and send to the client node as an compressed message, arrives at the node of the DRAM again decompress operation, after to run as a process, for searching by CSIS, due to the relatively small BF vector and the occupation of the memory resources very little influence on the normal operation of the nodes will be very small. Literature[3] demonstrated that there is 69.3% compression rate by Compressed Bloom Filter, and can significantly reduce the data transfer between server and client in CSIS, improve the performance of the system.

D. Experimental results

Experimental environment for the AMD2.09GHz processor, 2G memory, XP Windows operating system, cloud computing simulation software platform CloudSim3.0, the program to Java programming language, Eclipse4.4 as the development tool. The experimental data is tested with the clipku08 data set. The data sets are trained to get the self-set data sets.

TABLE1 Test data set

Data Set	DS1	DS2	DS3	DS4	DS5	DS16
Number	35	83	136	177	223	261
Size (MB)	4.5	11.8	18.7	23.6	30.2	35.1

Test in four kinds of modes: improvement of Self-Set storage mode (SS), using Bloom Filter representation of the Self-Set storage mode (BFSS), the Counting Bloom Filter improved increases the query operation of Self-Set storage mode (CTBFSS), and the Compressed Bloom Filter of the performance was further improved after the Self-Set storage model (CPBFSS). Figure 6 compares the time consumption of the four modes.

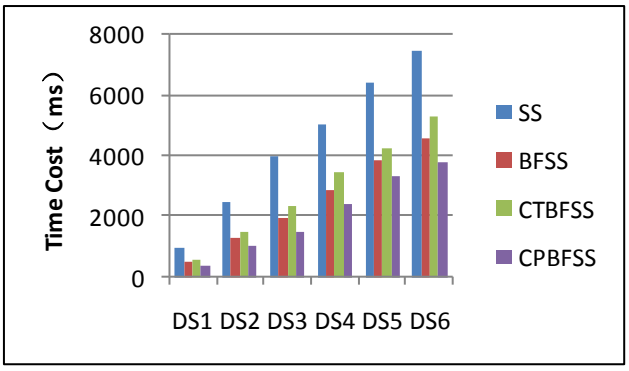

Fig.6 Comparison of consumption time efficiency

Because BFSS characteristic data in determining whether self-set, just running in the nodes BF string lookup, therefore its query time has been greatly reduced compared with SS. CTBFSS usage increases due to the counter BF bit string capacity, so its query times larger than the BFSS. CPBFSS is first to CTBFSS the BF bit string is compressed and then transmitted to the node to node and then decompress compressed because the bit string vector is relatively small, it is less than the transit time CTBFSS, but due to the increased compression and decompression compression of time, thus consuming time than whole BFSS and CTBFSS have reduced but the effect is not very obvious.

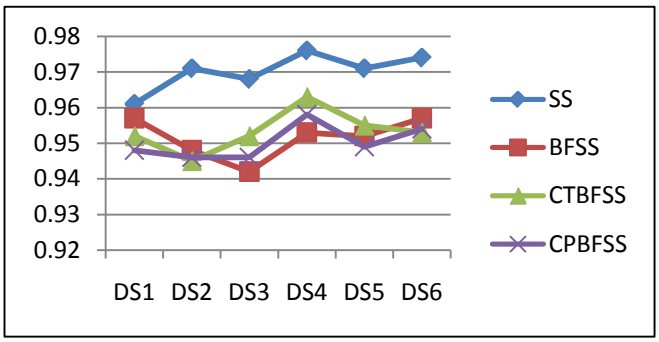

Fig.7 Comparison of recognition rate for the test results

Since the introduction Bloom Filter will bring minimal false positive rate, thus the recognition rate for the entire query is reduced for CSIS tests, but the recognition rate improved algorithm, or about 95%, is relatively high, compared to consumption reduce the amount of time, it is worth.

V. CONCLUSION

This paper firstly introduces the concept of self-set in human immune, then the concept, principles, research related improvements, and typical applications of Bloom Filter technology are studied and discussed, finally focus on the definition of self-set, construction algorithm in CSIS and the applications for Bloom Filter. Experimental results show that the Bloom Filter and related algorithms improved after the introduction of CSIS, the recognition rate of self-set feature data is reduced slightly, but it can significantly reduce the time spent at query, therefore it can improve overall performance of the CSIS.

REFERENCE

[1] Huo Lin; Zhou Jingxiong; Liu Xiangwei. Research on Security Mechanism of Cloud Security Immune System[C]. Sanya:The 2nd Congress on Computer Science and Application(CCSA2014)&Industrial Engineering，Machine Design and Automation(IEMDA2014), 2014:83-89.

[2] Howard Bloom.Space/time tradeoffs in hash coding with allowable errors[C].Communications of the ACM, 1970,13(7):422-426.

[3] Xie Kun; Wenjigang; Zhang defang, etl. Bloom Filter query algorithm[J]. Journal of Software ,2009,20(1):96108

[4] Flavio B, Michael M, Rina P, Sushil S, George V. An improved construction for counting Bloom filters[C]. In: Proc. of the 14th Conf.on Annual European Symp., Vol.14. Zurich: Springer-Verlag, 2006:684−695

[5] Mitzenmacher M. Compressed Bloom filters[C]. IEEE-ACM Trans. on Networking,2002,10(5):604−612.

[6] Saar C, Yossi M. Spectral Bloom filters. In: Proc. of the 2003 ACM SIGMOD Int'l Conf. on Management of Data. San Diego:ACM, 2003. 241−252

[7] Xiao MZ, Dai YF, Li XM. Split Bloom filter[J]. Acta Electronica Sinica, 2004,32(2):241−245

[8] Hong zheng; Wu lifa. Worm detection system based on positive selection[J]. Journal of Software, 2010，21(4):817-826.