The Application of Network Anomaly Mining Technology in Campus Network Information Security

Wang Bin
College of Information Science & Electronic Technology
Jiamusi university
Jiamusi,154007,China;
E-mail:jmsuwang@163.com

Zhi-chao Zhao*(Author for correspondence)
College of Information Science & Electronic Technology
Jiamusi university
Jiamusi,154007,China;
E-mail: 9_19zhou@sina.com

Yong-cheng Jiang
College of Mechanical Engineering,
Jiamusi University
Jiamusi,154007,China;
E-mail: jiangyongcheng@126.com

***Abstract*—Web Service architecture gradually matures in the related applications of campus network, but campus network anomaly mining technology still needs further development in its confidentiality, integrity, and non-repudiation problems. In SOA, campus network anomaly mining technology still needs to strengthen safety and reliability. Based on suffix tree technology, this paper proposed the application of campus network information security technology based on network anomaly mining technology, analyzed the campus network security features in detail, and designed an anomaly mining algorithm on suffix tree campus network security. Computer simulation results show that the proposed method can rapidly mine the abnormalities of the network and ensure the security of campus network.**

Keywords- Web log; Campus network; information security; data mining algorithm

I. INTRODUCTION

Today, the Internet got rapid development at home and in the world scope. With the development of social economy, the Internet becomes closer with people's life. It has already begun to infiltrate every detail of the people work and life, and human become attached to it. However, with the increasing degree of information level, the computer network resource sharing is further strengthened, and subsequent information security issues have become increasingly prominent, becoming the urgent problem [1 ~ 3]. Among them, the campus network security has become a lot of challenges to many colleges. Domestic campus network security issues are becoming increasingly prominent in the campus network. Web Service architecture gradually matures in the related applications of campus network, but campus network anomaly mining technology still needs further development in its confidentiality, integrity, and non-repudiation problems. Campus network anomaly mining technology still needs to strengthen safety and reliability. In the SOA architecture, the reliability and safety of data transmission arouse attention. This paper proposed the application of campus network information security technology based on Network anomaly mining technology, analyzed the campus network security features in detail, and designed an anomaly mining algorithm on suffix tree campus network security. Computer simulation results show that the proposed method can rapidly mine the abnormalities of the network and ensure the security of campus network [4 ~ 5].

II. CAMPUS WEB SECURITY SERVICE MODEL

A. Web Service architecture

In campus network distributed environment, Web Service technology completely realized the platform independence and cross application function from the application level. Web Service can adopt any kind of service component technology to fulfill the interaction between service and application via WSDL, using the UDDI protocols and release services to exchange data by XML through the SOAP protocol to achieve the interactions between service and application. The campus Web Service architecture based on the interaction of three characters of Web Service: a Service provider, Service demander, Service registry. Service providers realize service, define service description, and apply for registration service. Service registry fulfills the service registration and complete service release. On campus, service demanders find service, bind service, call service and interact with service in service registry. Figure 1 shows the operation of the process:

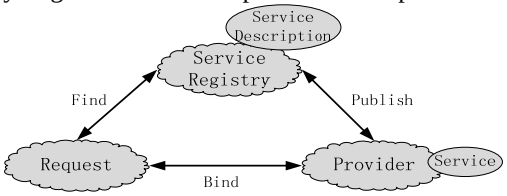

Figure 1. Web Service role and operation

B. Web Service system safety analysis

In the relationship between both ends of campus network transmission equipment, network transmission can be divided into Point-to-Point transmission and End-to-End transmission. SSL/TLS, VPNS and IPSec guarantee Web Service system information security based on the network point to point information security mechanism. However, point to point transmission has a defect that whether or when the receiving end will be able to receive data is not sure after sending out data. End-to-end transmission has advantages that the sending end knows whether the receiving equipment has received data or not and transmission delay is decreased

through the intermediate node equipment without information store and forward after the link establishment phase is completed.

In campus Web Service system, the data exchange between service provider and consumer service must be reliable. Only reliable data exchange can guarantee service provider accurately receive consumer service request, service consumer accurately receive service provider request feedback, and the whole business process of Web Service system may run smoothly. End-to-end transmission is reliable transmission with connection, suitable for Web Service system's requirements for data exchange reliability. The end-to-end data transmission of Web Service system is shown in Figure 2.

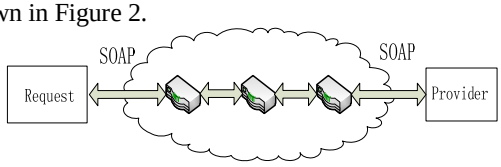

Figure 2. Data transmission of Web Service system

In Web Service information security mechanism, Web Service system adopts SOAP message interaction based on XML format. In SOAP message transmission process, Service system must consider the following safety problems:

(1) Whether the data transmission between Server and Client has been stolen or tampered with;

(2) How to confirm data comes from trusted party.

From the basic process analysis of campus Web Service work, Web Service security problem can be divided into SOAP data security problem (confidentiality, integrity) and how to confirm the identity of Service party (non-repudiation). Therefore, the key to solve Web Service security problem is how to deal with the existing Web Service confidentiality, integrity, and non-repudiation problems based on the information security mechanism as is shown in Figure 3:

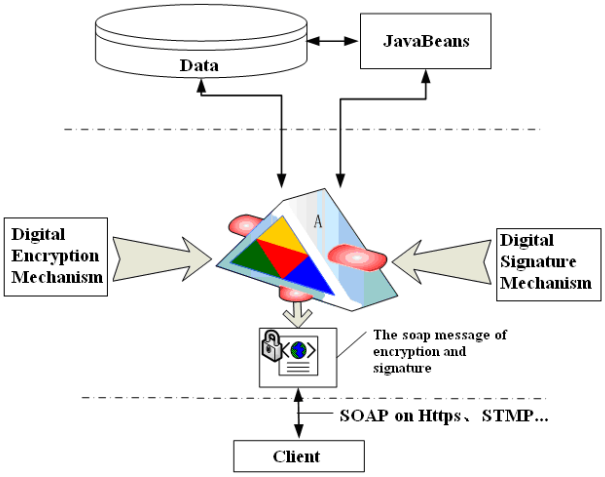

Figure 3. Web Service security mechanism based on information

In Figure 3, A is a service interface of Web Service. The client makes remote calls of the service interface A, then A can invoke related business components or operate database and use data encryption mechanism and digital signature mechanism to ensure the security of result data reflected to the client. In Web Service system, Web Service safety problems can be effectively solved with end-to-end communication mode based on end-to-end information security mechanism.

III. SUFFIX TREE MINING ALGORITHM

This research collected the third quarter of quarterly report and business data of the case enterprise from 2010 to 2012, and apply panel data model to the analysis of the collected data.

A. Related concepts of Suffix tree

(1)Phrase

Phrase is an order sequence with one or more words, which may be of arbitrary length without crossing the phrase boundary. Phrase boundary is inserted into phrases while document parser identifying special grammar markings, which can be punctuation mark (full stop '. ',comma', ',semicolon'; ',question mark'? ', etc.) or HTML tags (such as < p >, < br >, < li >, < td >, etc.). The beginning and the ending of the document are also regarded as phrase boundary.

(2) Phrases string

Phrase string is a phrase shared by at least two documents and all the documents containing the phrase. One biggest phrase string must satisfy that phrases in the phrase string can't be expanded with words of any language types without reducing the number of documents.

(3) Suffix tree

A suffix tree is a data structure that supports effective string matching and query. A suffix tree T of string S with m words is a directed tree contains a root node with just m leaves, and the leaves are endowed with labels from 1 to m. Each internal node, in addition to the root node, has at least two child nodes with each side identified with one non - null substring of S. The identifications of two arbitrary sides from the same node won't begin with the same words. The key features of suffix tree are: To any leaf I, all identifications concatenated from the root node to the sides of the leaves experienced spell the S's suffix from position I, namely S [I,..., m]. Tree node identification is defined as the concatenation of all identifications from the root to the sides. Figure 4 shows suffix tree of the string "I know you know I know". Internal nodes are presented as circles, leaves as rectangles. There are six leaves in this case, labeled from 1 to 6.

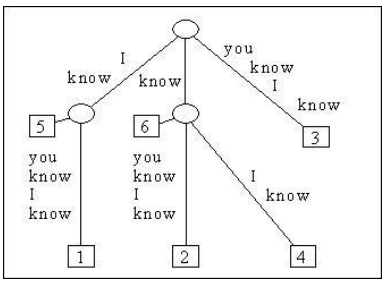

Figure 4. A suffix tree case

Similarly, suffix tree composed of some strings is called an extended suffix tree, in which, n string is Sn, the length of the string mn. The extended suffix tree T composed of these strings is a directed tree with a root node that has Smn leaves. Each leaf is identified with a two figures coordinate tuple (k, l), k ranges from 1 to n, and l ranges from 1 to mk. Each internal node, in addition to the root node, has at least two child nodes with each side identified with one non - null substring of S. The identifications of two arbitrary sides from the same node won't begin with the same words. To any leaf (I, j), all identifications concatenated from the root node to the sides of the leaves experienced spell suffix Si from position j, namely Si [j.. mi].

B. Suffix tree algorithm analysis and construction

1. Suffix tree takes document as a string with some phrases, not as a word set.

As a novel, incremental linear time operation method, suffix tree algorithm generates very compact data structure and saves a lot of storage space. This algorithm is very suitable in solving basic string problems, such as finding the maximal repeating substrings, similar string matching, string comparison, text compression and English document clustering, and solves them quickly.

2. The construction of the mining algorithm

To construct a suffix tree of string S with length m, firstly the suffix S [1.. m] is added to the tree as a side, then the suffix S [I.. m] to the growing tree, in which I grows from 2 to m. This algorithm is shown in detail as follows:

(1) let Ni be intermediate tree, which encode all suffixes from 1 to i.

(2) Tree N1 is composed of one side between the leaves from root to the leaf marked 1. The side is identified with string S.

(3) Tree Ni+1 is constructed by tree Ni, the process is as follows:

i. From the root node of Ni, operation rule find the longest path from the root, and the identification in the path should match the prefix of suffix S [i + 1.. m]. The path is found through successfully comparison and matching suffix S [i + 1.. m] and words from the root in the path until can't match any more.

ii. When there are no further matching, operation rules either reaches a node, called w, or the middle of a side. If the operation rule is in the middle of the side called (u, v), then it divides (u, v) into two sides by inserting a new node w.

iii The new node w is added to the end of the last match (the identification of that side should match suffix S [i + 1.. m])

iv In two cases (with or without an original node), operation rules have created a new side (w,i + 1). This side extends to a new leaf marked as i + 1 from w, and the new edge is identified by the unmatched part in suffix S [I + 1.. m].

IV. THE ANALYSIS OF EXPERIMENTAL RESULTS

The performance test goal of campus web anomaly intrusion detection system is to test the system's data processing and analysis skills in high-speed network environment. According to these requirements, the proposed methods would perform the following two tests:

(1) In data acquisition, test the capture ability of data packet. Especially compare the capture ability of high speed campus network and the traditional intrusion detection system.

(2) The packet processing capacity. To process the data package captured is another important indicator, and at the same time, data processing ability is the bottleneck of traditional system. The aim is to compare various performance indexes of the proposed algorithm and the traditional matching way in data processing.

A. Test environment

Figure 5 is a gigabit stand-alone campus network intrusion anomaly detection monitoring test environment, the models are PC machine with Windows XP operating system, P4 dual-core 1.8 G processor, 1 G DDR memory, 160 gb hard disk, gigabit Ethernet card.

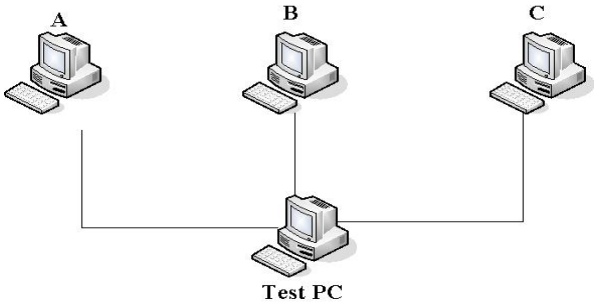

Figure 5. Test environment

Data packet collection test

Test network card will be sent directly to the user layer in Ethernet frame without any processing. The capture package performance in different IP packet size as is shown in Table I:

TABLE I. THE CAPTURE RESULTS OF ABNORMAL IP PACKET WITH PROPOSED METHOD

IP Package (byte)	Sended speed (pps)	Flow rate (dps)	Received speed (dps)	Received Flow rate	Received rate
64	585968	200M	585968	200M	100%
	645875	280M	645875	280M	100%
	786550	350M	663061	295M	84%
256	244141	500M	244141	500M	100%
	400000	600M	324000	486M	81%
1024	97656	800M	97656	800M	100%
	110000	900M	100100	819M	91%

Capture packet characteristics testing with traditional capture bag library Libpcap without any processing, as in table II:

TABLE II. LIBPCAP PACKET CAPTURE RESULTS WITH THE TRADITIONAL ALGORITHM

IP Package (byte)	Sended speed (pps)	Flow rate (dps)	Received speed (dps)	Received Flow rate	Received rate
64	104491	53M	101538	53M	100%
	190800	104M	190017	97M	100%
	315119	171M	191495	98M	57.3%

From test data, it can be seen that this kind of trap package platform has improved performance than traditional capture packet platform, high-speed capture packet rate enables it to adapt to the big flow network environment.

B. Pattern matching algorithm performance test

(1) Test project

Two groups of test are applied to test the influence of proposed algorithm and BM algorithm to system. The rules of the test are 1000 articles, in which 70 won't detect the packet while the rest 930 detect the packet with 200 amplitude increases through 300M network traffic data. Test the detection time and memory usage in both rules.

(2) Test results and analysis

Time and memory usage tested by 1000 rules as shown in Figure 6.

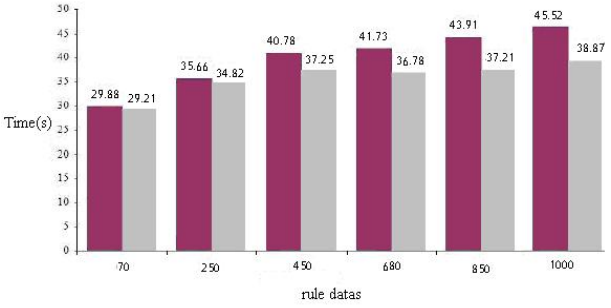

Figure 6. Performance comparisons of BM algorithm and the new algorithm

The Figure 6 test results show that the speed of the new algorithm is 1.5 times that of the BM algorithm, time is basically kept in a certain scope. Therefore, when the rule base is bigger, this algorithm almost constantly detects time with more tests. In this paper, the system performance is obviously improved as the number of rules in rule base increase. But time reduction must bring space consumption, so the space consumption of the new algorithm is much bigger than BM algorithm. With the development of storage technology, the price of memory shows a more and more low trend and memory spending has become a secondary factor.

V. CONCLUSION

This paper proposed the application of campus network information security technology based on Network anomaly mining technology, analyzed the campus network security features in detail, and designed an anomaly mining algorithm on suffix tree campus network security. Computer simulation results show that the proposed method can rapidly mine the abnormalities of the network and ensure the security of campus network.

ACKNOWLEDGMENT

This research was supported by the National Nature Foundation of China (Grant No. 61002004), the National Science Foundation of Heilongjiang Province (Grant No. E200908) , the Natural Science Foundation of Heilongjiang Province Education Office(Grant No. 12511551), the National Science Foundation of Jiamusi University(Grant No. L2012-074）, all support is gratefully acknowledged.

REFERENCES

[1] Hugo Haas.W3C:"Web Services Glossary". http://www.w3.org/TR/ws- Gloss, 2004,2.

[2] Heather Kreger.Web Services Conceptual Architecture(WSCA 1.0)[M].2001,5.

[3] MicrosoftTechNet:"Web Service Security". http:// www. microsoft.com/china/technet/security/guidance/secmod10.mspx#E2D. 2004.4.

[4] Sune Jakobsson."Security mechanism for web services". http:// www.eurescom.eu/message/messageJun2003/Security_mechanism_for_Web_Services.asp.

[5] Kent Ka lok Tong.Developing Web Services With Apache Axis2[M].ISBN:978-99937-929-1-8.2008,3：150-198.