Network Intrusion Detection System Model Based on Data Mining

Yanjie Zhao
School of Computer Engineering
Weifang University
Weifang, China
yanjie.zhao@163.com

***Abstract*—The paper's object is to develop a network intrusion detection model based on data mining technology, which can detect known intrusion effectively and has a good capacity to recognize unknown data schema which can't be detected effectively in traditional IDS. The paper mainly does the following work: by analyzing the intrusion deeply, extract the properties which can reflect intrusion characteristics effectively; combine misuse detection, anomaly detection and human intervention, establish rule library based on C.45 decision tree algorithm and use the optimal pattern matching so as to improve detection rate; the hosts are clustered to be IP group based on visit number by k-means clustering algorithm, the audit data are divided into parts under the IP group's direction, and the classifiers are built up by divided audit data respectively, then the detected Data apply different rules according to their own IP group, thereby reduce false positives. The experiments proved that the method is effective to detect intrusion such as scanning and Deny of Service.**

Keywords—intrusion detection system; data mining; network security

I. Introduction

Currently, the application of data mining in intrusion detection system had become a hotspot. In this research, the most influential researchers are the Wenke Lee Study Group of Columbia University. Most subsequent researchers followed the Wenke Lee and Portnoy, and on the basis, improved or combined data mining with other intelligent technologies (such as genetic algorithms, fuzzy technology) [1-4]. But the research in the field is still in the stage of theory-exploring, is not mature and complete, and there are mainly two shortcomings as follows:

(1) Mining rules come from data, but the data were mainly produced automatically and the data combined with the characteristics of intrusion are too little, therefore, rule-making is blind。

(2) Although the rule acquisition is automatic, but the methods of analysis is simple. The misuse detection based on data mining still does not automatically recognize the new intrusion, and the anomaly detection based on data mining is still unable to overcome the abuse that normal behavior is mistaken for intrusion.

And because of the reason above, currently, the detection accuracy of the intrusion detection system based on data mining is relatively low, and unable to meet requirements of practical application.

So basing on the prior knowledge of information security, we extract the special attributes which reflect the network communications rule effectively, then, apply appropriate data mining algorithms , combine the misuse detection with anomaly detection together, and propose a data mining-based network intrusion detection method to improve the previous method.

II. Method Description

No a detection method can detect all of intrusion behaviors, a perfect IDS should be a unity of multiple detection means. The research goal of this paper is finding a detection method based on data mining which can identify both the known intrusion and the unknown intrusion, thereby attain satisfied detection accuracy to two kinds of attack means: scanning and Deny of Service (DoS)[5-7].

Based on the above ideas, we design a data mining-based network intrusion detection model. The model builds classifier based on historical network data, and recognizes intrusion in the new network audit data with the aid of the classifier.

This paper argues that the behavior between different systems is quite different. For example, the audit data of workstations and network servers are significantly different in the data patterns. As server, the host's typical characteristics are that the traffic is great and continuous, connection frequency is high and generally it does not send connection requests on its own initiative, mostly as the destination host of the TCP connection; And as workstations, the host acts as the initiator of TCP connections mostly, its main aim is to obtain data. If the workstation suddenly appears a large number of high frequency active connections in unit time, the behavior is very likely abnormal. To the server, the large number of active connections can be understood as normal if the connection number do not exceed the upper limit, that is, to the server and the workstation, the judgment of normal and abnormal behavior is not same.

Because the service and traffic of hosts are different, in the ideal situation, each host should have its own behavior patterns database, but this would increase the data processing expenses. Based on the above ideas, the model analyzes visits per unit

978-1-5090-2239-7/16/$31.00 copyright 2016 IEEE
SNPD 2016, May 30-June 1, 2016, Shanghai, China

time of the hosts in the network, and then clusters the hosts to be IP group based on k-means clustering algorithm[8]. The hosts in the same group should have more similar data traffic. Each group set up their own behavior pattern library, thus avoiding the great individual differences. IP group established process is shown in Fig. 1.

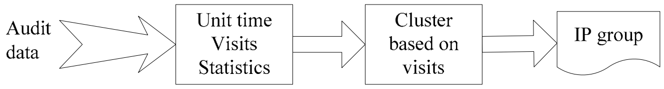


Fig. 1. IP group established process

IP group established is used to guide the network data distribution, and the audit data distributed establish their own classifier based on the C.45 decision tree algorithms[9], define different rules thresholds, and establish behavior pattern library for their own. The data flow is shown in Fig. 2.

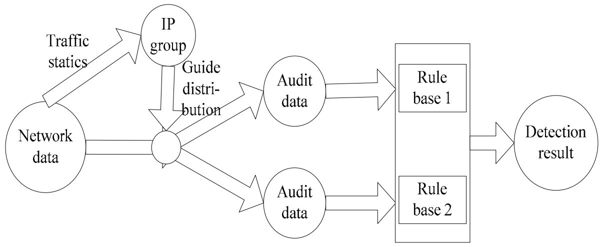


Fig. 2. IP groups guide data distribution

The model combines two kinds of analysis methods: the misuse detection and anomaly detection. The normal behavior and abnormal behavior are established pattern base separately. The test data were matched with the abnormal pattern and normal pattern respectively. If the similarity with normal pattern is greater than or equal to normal pattern threshold$\delta 1$, the data are regarded as normal behavior, and if the similarity with the abnormal pattern$\geq$ abnormal pattern threshold$\delta 2$, the data are regarded as intrusion. In reality, network data patterns are infinite, and the initial training data which can be collected are limited. In test data, there must be some pattern types which system can not recognize, that is, the similarity of the pattern to the existing normal and abnormal patterns is lower than both thresholds. The new data pattern will be submitted to the administrator, the administrator determined its type based on experience and append the type identification to it, then sent it to the audit record database. After the rules are amended based on the updated database, the system will have the ability to identify the new data pattern.

III. MODEL FRAMEWORK

The model architecture meets CIDF specifications[10], system structure as shown in Fig. 3.

The model has six main modules:

A. Data capture

Data are captured by the NAM Module of network core switch. Set the appropriate capture parameters, capture network data flow through the specified network, and save the captured data into database.

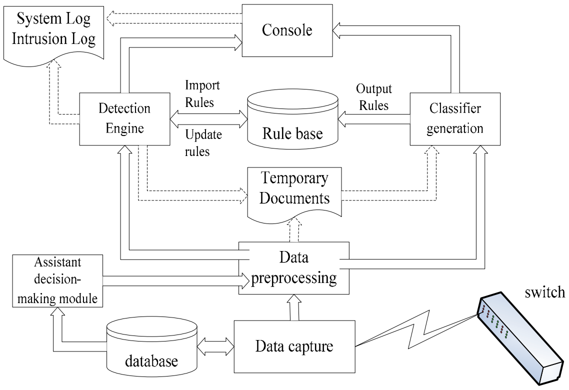


Fig. 3. System structure

B. Data preprocessing

The captured network data is binary. First of all, the module resolves network data, extracts the information we are interested in and process large numbers of data packets into TCP connection records which reflect session information. Then the system will complete the works such as data filtering, information statistic based on the window and format conversion, ultimately, generate the property list which can reflect network communications rule.

C. Assistant decision-making

This module regard the network data as data source, analyze host visits and cluster the hosts to be IP group based on visits, here, the visits refers to the host's unit time average connected frequency. Then network audit data are divided based on IP group, the purpose is that the classifiers are built up by divided audit data respectively, then the detected Data apply different rules and define different threshold so as to reduce the impact of host individual differences to classifier.

D. Classifier generation

The audit records with teacher signal will be imported into the module, and generate a decision tree based on the C4.5 algorithm, and then rules are derived from decision tree and put into rule base so as to be called by detection engine. The module has two data source, one from the data preprocessing module, the other from the audit records which can not be identified by detection engine and then marked with teacher signal by the administrator with experience.

E. Detection Engine

Detection engine is the important component of the system, which judges the audit record normal or not according to detection rules. In order to obtain high detection rate, low false positive rate and the ability to identify new patterns, three methods such as misuse detection, anomaly detection and manual identification are applied in the module. Details see Section VI.

F. Console

This module is a interact interface of administrator and system, which controls all aspects of the system progresses, we can also consult the documentation for the system through the console.

Here, the key modules of the model will be introduced detailed.

IV. DATA PREPROCESSING

Network intrusion detection system is mainly targeted at network data, and the quality of feature extraction determines system work efficiency [21]. Under normal circumstances, large amounts of data flow through the network, so capturing long data segment will put pressure on storage space. The model analyzes the packet header information, and specifies the capture length 64Byte.The network data captured by system are binary data, and can not be directly used, it is necessary to extract the information we are interested in, and process it to obtain the final characteristic properties which can reflect network communication rule. In this process, we mainly complete the following steps:

Step 1 Transform raw binary network packets into the ASCII code audit records. Each packet has a timestamp, and audit records are sorted by timestamps. Audit record format shown in table I.

TABLE I. PACKET FORMAT.

format	**No., timestamp, s-ip, d-ip,s-port,d-port,protocol type, packet length, serial number , ack number,URG,ACK,RST,SYN,FIN**
Value	201, Mar21 18:15:43.048120, 202.194.64.135, 202.194.67.9, 2926, 8000, tcp, 0, 0, 0, 0, 0, 0, 1, 0

Where: s denotes source, d denotes destination. For example, s-d denotes from source host to destination host,s-ip denotes the IP address of source host。

TABLE II. SESSION CONNECTION FORMAT

format	**No., timestamp, s-ip, d-ip,s-port,d-port, Duration, connection_ FIN_id, s-d traffic, s-d_URG_id, s-d_resend_id,d-s_resend_id**
Value	18, Mar21 18:15:43.048120, 202.194.64.135, 202.194.67.9, 2926, 8000, 0:00:01.191029, 1, 716, 0,0 ,0

Step 2 By analyzing the network protocol, transform the above format data into TCP connection records (as shown in table II), each record reflects a complete session:

Step 3 Associate session records based on time window, and calculate statistics. At this point the connection record already contains part of the inherent characteristics which embody an internal connection. But the signatures of attacks such as scanning and denial of service are often reflected in the relationship between records. Therefore, we associate and count the connection record within the "w" time window, and get the following additional features:

a) sip_diff_dip number: count the number of different destination IP connected by the host in the "w"window according to the source IP of the connection record.

b) sip_diff_dport number: count the number of different destination port connected by the host in the "w"window According to the source IP of the connection record.

c) dport_same_sip number: count the number of destination port connected by the same host in the "w" window According to the destination port of the connection record.

d) dport_diff_sip number: count the number of destination port connected by different host in the "w" window According to the destination port of the connection record.

e) The ratio of sip_diff_dip number to all connections in the "w" window.

f) The ratio of sip_diff_dport number to all connections in the "w" window.

g) The ratio of dport_same_sip number to all connections in the "w" window.

h) The ratio of dport_diff_sip number to all connections in the "w" window.

Statistical properties list shown in table III.

TABLE III. STATISTICAL PROPERTY LIST

format	**No., timestamp, s-ip, d-ip,s-port,d-port,Duration,connection_ FIN_id, s-d traffic, s-d_URG_id,s-d_resend_id,d-s_resend_id, sip_diff_dip_num,sip_diff_dport_num, dport_same_sip_num, dport_diff_sip_num, sip_diff_dip_num_rate, sip_diff_dport _num_rate, dport _same_sip _num_rate, dport_diff_sip_num _rate**
Value	18, 181543.04, 20219464135.00, 202194679.00, 2926.00, 8000.00, 1.19, 1.00, 716.00 , .00 , .00, .00 , .00, .00, .00, 4 .00, .00, .00, .00, .19

Step 4 The audit data are divided into parts under the IP group's direction and build the rule library respectively.

Intrusion can occur at any time, any port could become invasion target, and any host can become invaders, are also likely to become victims. As reference attributes, the connection initiated time, IP address and port number play an important role to generate additional property, but these values themselves does not judge whether the invasion happens. So the final attribute list which reflects the network communication rules as shown in table IV:

TABLE IV. AUDIT RECORD FORMAT

format	**No., Duration,connection_FIN_id,s-d traffic,s-d_URG_id,s-d_resend_id,d-s_resend_id, sip_diff_dip_num,sip_diff_ dport_num,dport_same_sip_num,dport_diff_sip_num,sip_diff_dip_num_rate,sip_diff_dport_num_rate, dport _same_sip _num_rate,dport_diff_sip_num_rate**
Value	18, 1.19, 1.00,716.00, .00, .00, .00, .00, .00,.00, 4 .00, .00, .00, .00, .19

V. CLASSIFIER GENERATION

The modules do two works. One part is to generate the original detection rules, the other part is to update rule during system operation. The module structure is shown in Fig. 4:

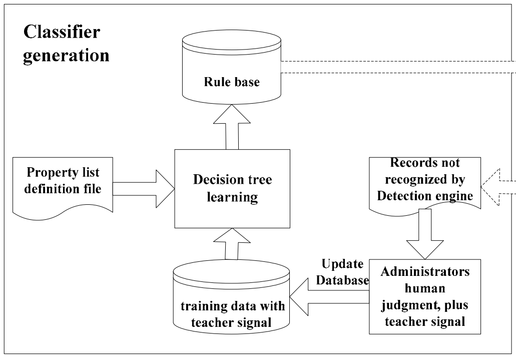

Classifier generation
Rule base
Property list definition file
Decision tree learning
Records not recognized by Detection engine
Detection enqine
Update Database
training data with teacher signal
Administrators human judgment, plus teacher signal

Fig. 4. Classifier generation module structure diagram

Generating original detection rules is the foundation for system normal operation. Decision tree learning is an important process to achieve detection rules from the training data.

First, mark the teacher signal to the original audit records as training samples: teacher signal “1” indicates intrusion data, “0” indicates normal data. Then, build tree using the training sample based on C4.5 algorithm, and prune to form the decision tree shown in Fig. 5:

```
    Decision Tree:
    sip_diff_dport <= 8 :
    |  dport_diff_sip_ratio <= 0.88 :
    |  |  dport_same_sip_ratio <= 0.03 :
    |  |  |  sip_diff_dport_ratio <= 0.13 :
    |  |  |  |  s->durgernt > 0 : 4.00 (37.0)
    |  |  |  |  s->durgernt <= 0 :
    |  |  |  |  |  sip_diff_dip_ratio <= 0.08 :
    |  |  |  |  |  |  dport_diff_sip_ratio <= 0.74 :
    |  |  |  |  |  |  |  s->dreset <= 1 :
    |  |  |  |  |  |  |  |  dport_same_sip_ratio <= 0 :
    |  |  |  |  |  |  |  |  |  s->dreset <= 0 :
    |  |  |  |  |  |  |  |  |  |  s->ddata <= 16 :
    |  |  |  |  |  |  |  |  |  |  |  tcpbz = .00:[S1]
    |  |  |  |  |  |  |  |  |  |  |  tcpbz = 1.00:[S2]
    |  |  |  |  |  |  |  |  |  |  |  tcpbz = 2.00:
    |  |  |  |  |  |  |  |  |  |  |  |  s->dreset > 0 : .00 (33.0/2.0)
    |  |  |  |  |  |  |  |  |  |  |  |  s->dreset <= 0 :[S3]
    |  |  |  |  |  |  |  |  |  |  s->ddata > 16 :[S4]
    |  |  |  |  |  |  |  |  |  s->dreset > 0 :
    |  |  |  |  |  |  |  |  |  |  dport_diff_sip_ratio <= 0.04 : .00
(32.0)
    ...
```

Fig. 5. Example of decision tree

In the figure above, [Si] represents decision subtree; (A/B) expresses that A records participate in the current judgment, In which B records do not support the judgment.

Then the rules derived from the tree and put into rule base, rule expressions as shown in table V.

TABLE V. EXAMPLES OF DECISION RULES.

Rule expression	Interpretation
tcpid = 2.00, s->ddata > 4, sip_diff_dip <=33, dport_diff_sip_ratio> 0.88 →class 3.00 [99.0%]	if the connection end id is reset,the traffic from source host to destination host is greater than 4Bype, the number of destination host connected by the same source host is less than or equal to 33 in 2 seconds,the num of different source host connecting the destination port is greater than 88% of all connection,the connection is weak password scanning, and credibility is 99.0%

Another task of this module is to update rules when system operates. The data patterns reflected by original audit records are limited, new data pattern may appear in the network every day. These patterns may be new invasion, or may be the normal mode which not included in the training data, they can not be recognized by the detection engines. Here human intervention should be used. The relevant data pattern would be sent to the console by detection engine, and the administrator judge based on experience, plus a flag to the data pattern, and then adds the audit records with flag into the database. When rule base updates, the system would has the ability to recognize the pattern.

VI. DETECTION ENGINE

This module is used to perform intrusion detection, which receives the data processed by data capture unit, calls the rules generated by classifier generation unit ,determines each data pattern normal or abnormal. The module structure as shown in Fig.6:

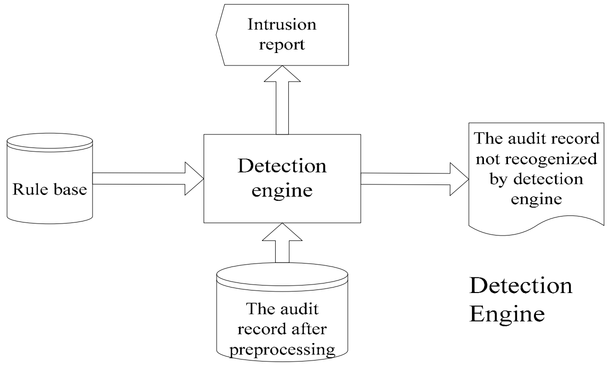
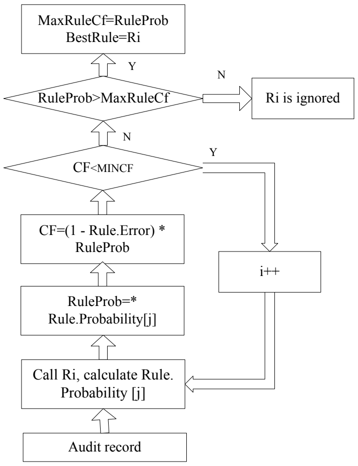

Intrusion report
Rule base
Detection engine
The audit record not recogenized by detection engine
The audit record after preprocessing
Detection Engine

Fig. 6. Detection engine structure

Detection engine determine audit records based on the detection rules. Each audit record represents a data pattern. Detection results should be based on the highest degree of rule matching. The process of selecting the optimal rule as shown in Fig.7:

After calculation, the optimal matching rule is produced. If the rule’s inference class is normal, confidence level MaxRuleCf will be compared with the normal behavior threshold $\delta 1$. If MaxRuleCf$\geq \delta 1$, the record is determined normal and will be ignored. If the rule’s inference class is abnormal, MaxRuleCf will be compared with the abnormal behavior threshold $\delta 2$ .If MaxRuleCf$\geq \delta 2$, the record is determined abnormal, then report the intrusion information such as type, source and destination to the console, and record the appropriate log; If MaxRuleCf< $\delta 2$ and MaxRuleCf< $\delta 1$, namely discrimination confidence level is below the set threshold, view the record as an unknown pattern and put all of its information (including the original record, post-processing statistical information) into the "ambiguous data" list, so that system administrator do a judgment based on experience.


MaxRuleCf=RuleProb
BestRule=Ri
Y
N
RuleProb>MaxRuleCf
Ri is ignored
N
Y
CF<MINCF
CF=(1 - Rule.Error) * RuleProb
i++
RuleProb=* Rule.Probability[j]
Call Ri, calculate Rule. Probability [j]
Audit record

Ri:
The current record
MINCF:
Minimum confidence threshold set by the system
Rule.Error:
Rules estimation error rate
Rule.Probability[j]:
The matching possibility of the current rule's i-th condition to the given pattern
RuleProb:
Cumulative value of rule's condition branch matching degree
CF:
Match confidence of the current rules for the data pattern
MaxRuleCf:
credibility of optimal matching rule
BestRule:
Optimal matching rules

Fig. 7. Select The Best Matching Rules

VII. Performance Test and Analysis

A. Test data

Prototype system implements by the C++ language, and verification of this model relies on a particular campus network, and experimental data derive from VLAN5 and VLAN6 of the campus network. VLAN5 provide essential services for campus network, and its data traffic is great; VLAN6 mainly includes ordinary workstation, data traffic is relatively small. In this study, we choose some hosts among the two VLAN as experimental data source. The data traffic generated by two VLAN has big difference, which can meet the experiment requirements. VLAN5 and VLAN6 act as data capture objects, in two weeks time, we acquire representative data as normal data samples at different times every day.

Campus network has stable network data traffic, which provides an ideal source of normal experimental data. However, the network intrusion data in the actual network is very little, even if the intrusion occurs in the network, it is difficult to determine exactly what data constitute a threat to the network. In order to capture intrusion data, we made attack simulation in VLAN5 and VLAN6.

The types of simulated attack include two major categories: scanning and denial of service, which contains 10 small categories. Each attack category is given an identification number, as shown in table VI:

In the end, 30000 pieces of representative connection data are selected from the captured data as the experimental data, including the normal data flowing through the VLAN and the intrusion data generated under the simulated attack environment. 25000 records are selected as training data to establish classification, and other data are used for system testing.

TABLE VI. Attack types.

Attack type		mark	Attack type		mark
scanning	TCP connect scanning	1	Deny of Service	OOB	7
	Loopholes Scanning	2		Syn flood	8
	Weak password scanning	3		Land attack	9
	OS scanning	4			
	Tcpsyn scanning	5		DDoS attack	10
	Ack scanning	6			

B. Known intrusion types detection

The purpose of this study is to test the system detection ability for known intrusion types. In the experiment, test data includes 3882 records which come from part of the normal data and eight kinds of known intrusion data. Final test results shown in table VII:

TABLE VII. Detection results of known intrusion types.

Data types	**Normal data**	**Tcpport canning**	**loopholes canning**	**OS Scanning**
Record Number	1400	400	306	144
detection rate(%)	98.84	96.46	99.34	87.41
Data types	**Tcpsyn Scanning**	**Ack Scanning**	**Landattack**	**Syn flood**
Record Number	400	50	8	577
detection rate(%)	99.25	0	50.00	99.65
Data types	**DDoS**	**All data**		
Record Number	600	3882		
detection rate(%)	95.83	96.50		

Among them, almost all of Ack scanning are identified as Tcpsyn scanning. As can be seen from the above data, Ack scanning and the Land Attack reduces the overall detection rate.

After careful observation to the original audit records generated by Ack scanning and Tcpsyn scanning, we found that although two scanning principle is different, but the features shown by audit records in the system are identical. In other words, the selected attributes characteristics can not distinguish them. Because the two scanning have the same attack purpose, we can group them into a category, then the Ack scanning should be considered to identify effectively. After re-calculating, the total detection rate is 97.78%.

In this system, the judge rules derived from the training data to LAND attack is S-> dreset> 3700 → class 5.00, which means that If the packet retransmission times from source IP to destination IP are greater than 3700, it is the land attack. The rules do not reflect the fundamental characteristics of the land attack. So, we should apply the feature matching method: if source address and destination address are same, it is the land attack. The latter rule can more accurately detect out the Land Attack.

C. *Unknown intrusion types detection*

The purpose of this study is to test system capability of detecting unknown type data which include the normal data schema and the intrusion types do not appear in the training data. The system is difficult to guarantee the collected data patterns are complete when it set up, which requires the system possessing good recognition ability to the unknown data patterns. In the experiment, the test data include part of normal data and two kinds of intrusion data which do not appear in the training data, a total of 1,000 records, among them, there are 500 records with attack mark. Final test results as shown in table VIII:

TABLE VIII. DETECTION RESULTS OF UNKNOWN INTRUSION.

Data type	**OOB attack**	**Weak password scanning**	**All intrusion data**
Record num	300	200	500
Recognition rate(%)	85.00	61.50	75.60

In this experiment, there are 85% of OOB attack data are identified as unknown pattern data, 14% of OOB attack data are mistakenly believed to be the normal data; 61.5% of weak password scanning data are identified as unknown pattern data, and another 32% of the data to be mistaken for loopholes scanning, 6.5% of the data are mistaken for normal patterns.

Here, the recognition rate reflects the system's ability to discover unknown patterns of data, and can not determine which type it should be. When the system detects an unknown type of data, the related audit records and statistical data will be submitted. The administrator defines the type mark based on his experience, then the original records with the teachers signal are appended into the audit records database, until the rule base update, the system will have the ability to distinguish the types of data. Repeat the previous test, the test results obtained as shown in table IX:

TABLE IX. THE TEST RESULTS OF RULE UPDATE ABILITY.

Data type	**OOB attack**	**Weak password scanning**	**All intrusion data**
Record num	255	123	496
Recognition rate(%)	97.64	84.55	93.38

Here, there are about 10% of the weak password scanning records were mistaken for loopholes scanning.

D. *Effect of data distribution to system performance test*

Because the traffic of each host is different, the network data patterns of the hosts should be different. Based on the above considerations, we propose network data distribution based on traffic clustering, precisely define the rules and threshold so as to achieve higher accuracy .The purpose of this study is to verify the correctness of ideas.

The former two experiments are based on data distribution. In this experiment, all audit records are regarded as training data to establish classification. Recognition rate as shown in table X:

TABLE X. THE TEST RESULTS ABOUT DATA DISTRIBUTION.

Data type	**data distribution**	**data without distribution**
Record num	3882	3882
Recognition rate (%)	96.50	92.11

The results show a reasonable data distribution can improve the detection rate. The test results without data distribution show that some regular data is mistaken for the distributed denial of service in the case of large traffic.

VIII. CONCLUSION

Experimental results show that to establish data mining-based intrusion detection system, the prior knowledge of security is very important. Data mining can only extract rules from the limited raw data, but not all the rules are reasonable. It is necessary that the expert with computer safety knowledge provides evidence to guide its work. Furthermore, diversification detection means should be promoted. High detection rate in the experiment shows that the methods are feasible such as combining misuse detection and anomaly detection together, data distribution and so on.

REFERENCES

[1] T.F.Lunt , R.Jagannathan,and R.Lee, “IDES:The Enhanced Prototype-A Real-Time Intrusion-Detection Expert System”, Number SRI-CSL-88-12 . Computer Science Laboratory,SRI International,Menlo Park,CA,1988.

[2] Wenke Lee and Salvatore J.Stolfo, “Data Mining Approaches for Intrusion Detecion”,In Proceedings of the 7th USENIX Security Symposium,1998.

[3] Lee,Wenke, Salvatore J. Stolfo, and KuiW. Mok. “A data mining framework for building intrusion detection models.” Security and Privacy, 1999. Proceedings of the 1999 IEEE Symposium on. IEEE, 1999.

[4] W. Lee and S. J. Stolfo. A framework for constructing features and models for intrusion detection systems. ACM Transaction on Information and System Security, 3(4):227-261, Nov. 2000.

[5] Z. A. Othman, A. Bakar and I. Etubal, Improving signature detection classification model using features selection based on customized features. Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on Nov. 29 2010-Dec. 1, 2010.

[6] O. Al-Jarrah and A. Arafat , “Network Intrusion Detection System using attack behavior classification” , 2014 5th International Conference on Information and Communication Systems (ICICS) , pp.1 -6 .

[7] Z. Wanlei, Keynote: Detection of and Defense Against Distributed Denial-of-Service (DDoS) Attacks, Proc. IEEE 11th International Conference onTrust, Security and Privacy in Computing and Communications, 2012

[8] Beniwal, Sunita, and Jitender Arora. “Classification and feature selection techniques in data mining.” International Journal of Engineering Research and Technology. Vol. 1. No. 6 (August-2012). ESRSA Publications, 2012.

[9] Neha G. Relan and Dharmaraj R. Patil. “Implementation of network intrusion detection system using variant of decision tree algorithm.” 2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE) ,2015.

[10] P.Porras, D. Schnackenberg, et al. “The Common Intrusion Detection Framework Architecture, CIDF”, University of California, 1998.