International Journal of Security and Its Applications
Vol. 10, No. 8 (2016), pp.263-276
http://dx.doi.org/10.14257/ijsia.2016.10.8.23

A Novel Rough Set Methodology and Machine Learning based Novel Network Intrusion Detection System: Theoretical Analysis and Applications

Yang Hui-jun[1]

[1]*Department of Information Service, Anhui Institute of International Business, Hefei AnHui, 231131, China*

Abstract

Data mining for intrusion detection is one of the most cutting-edge researches which focus on network security, database, and information decision-making. Due to the emergence of new forms of attacks and intrusion on the network, we need a new intrusion detection system which would be able to detect new and unknown attacks. Nevertheless, because of the complexity and diversity of network security alarm data, ordinarily, it is difficult to analyze and evaluate network security situation accurately. Intrusion detection is to protect network system from attacks and defend its security. We used machine learning technology in intrusion detection system in order to improve system performance effectively. In the paper, by studying the characteristics of network data intrusion, we put forward a intrusion detection system based on Rough set theory, and detect anomaly action in network. This method can extract detection rule model of network connection data, dealing with incomplete data and discrete data exit in data mining effectively. The basic ideas and techniques of data mining-based intrusion detection and the architecture of a real time data mining-based IDS are discussed. Meanwhile, we mainly analyzed the basic structure of intrusion detection system and application of several Machine Learning methods in intrusion detection which include Bayesian Classification-based method, neural networks-based method, Support Vector Machine-based method(SVM). The experiments results show that, models, methods and generation framework proposed in this paper can effectively detect network intrusion.

Keywords*: Network Intrusion, Rough Set Theory, Machine Learning, SVM*

1. Introduction

With the continuous development of network technology, people's work and life has been closely linked with the network. Nevertheless, network has become a major target of attackers, and the risks of network intrusion are increased. Network security problem has become unavoidable. Intrusion detection system[1] is the key technology which can protect network security, and it protect network form attacks initiatively. Thus, intrusion detection has become an important technology which get more attention around the world.

Intrusion detection technology is divided into two categories[13]: anomaly detection and misuse detection. Anomaly detection uses quantitative way to describe acceptable behavior characteristics, distinguishing abnormal behavior characteristics and those contrary to normal behavior action. The advantage of anomaly detection is it can discover new and unknown intrusion, simultaneously, is has a certain learning ability. Misuse detection uses foregone system and attack pattern of application to detect intrusions. The feature of misuse detection is it can detect intrusion with high accuracy, but misuse detection can only discover foregone intrusions[2].

In 1980, James P. Anderson first proposed the concept of intrusion detection[3]. After that, in 1987, Denning proposed a generic intrusion detection model, and this model

ISSN: 1738-9976 IJSIA

became the basis of development in intrusion detection. In 1990, Heberlein exploited the fist Network Security Monitor(NSM)[4], from now on, two basic structure of intrusion system(host-based and network-based) is determined. Passed through twenty years, there were many different intrusion detection systems based on various methods or theories. However, these systems all lack accuracy and ability of detect new and unknown attacks, thus, it is inconvenient for us to make distributed analyses or team working. In order to solve these problems, we applied intelligent technology and machine learning into intrusion detection systems. Wenke Lee and his colleagues proposed a data mining framework used in IDS[5], trying to mine network audit data in order to discover a new model and improve adaptability of IDS. Forrest *et al*. referred to Artificial Immune Theory, and established normal call database, the basic of anomaly detection. Normal operating stale of processes are described by the system call sequence model generated when processes running normally[6]. T. Lane *et al.*, established the framework use for normal behavior, and applied case study method and pattern matching to detect intrusions[7], however, the cost of this algorithm is pretty large.

In this paper, we introduce Rough set theory[8], and establish network intrusion detection model based on Rough set theory. Simultaneously, we mainly describe Machine Learning based on Bayesian Classification, neural networks-based, and association rules mining intrusion detection technology. We proposed Anomaly Detection System based on Data mining(ADSDM). ADSDM is able to mine suspicious behavior within port numbers, application layer data, and network data protocol. In data mining, we mainly notice association rules data mining method based on weak rule. This method is used to detect attacks those have less abnormal operation and may not be readily detected. At the same time, the influence between network communication time, direction, port number, and host address is used to establish Bayesian networks with various attributes nodes, then regards the networks as abnormality discriminator. The discriminator further distinguishes suspicious behavior discovered in association rule mining. By the way, in this paper, on the basis of summarization for the recent development of network security situation awareness, we proposed a network security situation model based on knowledge discovery, and we applied into network security situation awareness system which was denoted as Net-SSA[13]. During the experiment, we figured out that association between several different data mining methods works more effectively than usage of one method. For instance, combination of Rough set theory and genetic algorithm can improve intrusion detection accuracy.

2. Rough Set Theory

Rough set theory assumes that knowledge is a kind of ability to classify objects, and knowledge must be associated with various classification models among specific or abstract parts of the world. These parts are called universe discussed.

Definition 1: Information system is denoted as a quadruple: $IS = \{U, A \cup \{d\}, V, f\}, U \neq \phi$. A is all attributes set of object. $V = \bigcup_{a \in A} V$ is attribute value set, and V_a is range of attribute $a \in A$. Information function $f : U \times A \to V$, this function specifies values of x in U.

Definition 2: For every attributes subset $B \subseteq A$, we define binary relations as $IND(B) = \{(x, y) \mid (x, y) \in U^2, b \in B, s.t. b(x) \neq b(y)\}$, $IND(B)$ is indiscernible. Obviously, $IND(B)$ is an equivalent relation, and $IND(B) = \bigcap_{Ib \in B} IND(\{b\}); \forall x \in U$. We denote equivalence class x as $[x]_B$.

Definition 3: $T = \{U, A \cup \{d\}, V, f\}$ is a decision-making system. U, A, V, f are the

same to Definition 1, and $\{d\}$ is decision attribute, A is condition attribute. We assume P is equivalence relation set of U^2, if $IND(Q) = IND(P)$, then Q is reducts of P. We define all necessary relations sets as *core*, denoted as *CORE*(P).

Definition of decision-making attribute $D = \{d\}$:

$$M_D = M_D(i,j)_{m\times n} = \begin{Bmatrix} \{c_k \mid c_k \in C \wedge c_k(x_i) \neq c_k(x_j), d(x_i) \neq d(x_j)\} \\ \phi, d(x_i) = d(x_j) \end{Bmatrix}$$

2.1. Network Intrusion Detection Technology based on Rough Set Theory

Network intrusion detection system includes two stages: pattern generation and pattern detection. During pattern generation, network data collection module is to collect network connection data, and data selection module is to select target data from all connection data including dimension, attribute, and data type. Rough set theory is used in data preprocessing module and reducing knowledge module. During pattern detection stage, the system uses generated detection rules to detect current data and alarm for abnormal behavior.

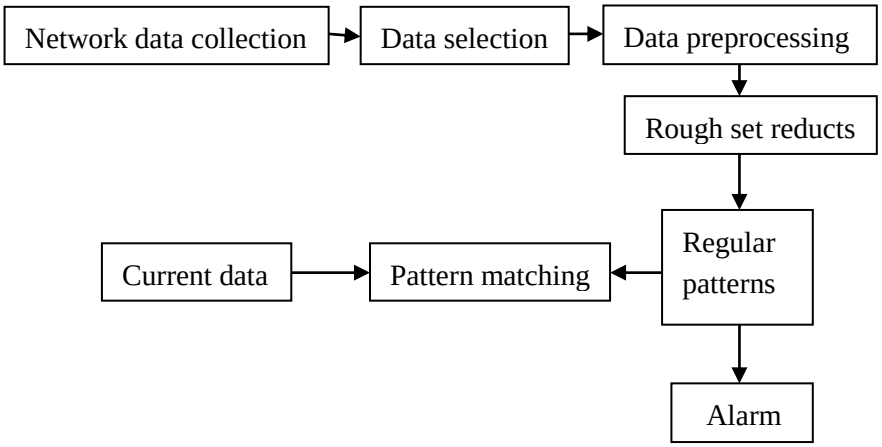

Figure 1. Network Intrusion Detection System based on Rough Set Theory

2.2. Data Preprocessing

We choose KDD CUP 99 Data Sets as data source, and there are 41 attribute values in each network connection record which equal to condition attribute in information system. The data in this data source are complete.

There are 30 numerical attribute in this data set, thus, we use Semi Navie Scaler Algorithm to make data discretization. The selection of the breakpoint is completely determined by the information of data, and persist the indiscernible relation between original data. The algorithm is as follows:

For each $a \in A$, breakpoint set $C_a = \phi$, we have three steps:

Step 1: Sequence range V_a, where x_i,x_{i+1} are interfacing records, and $v_a^i \leq v_a^{i+1}, v_a^i, v_a^{i+1} \in V_a$

Step 2: Compute decision value sets with maximum frequency D_i,D_{i+1}:

$$D_i = \{v \in V_d \mid v = \arg\max_{v'} \mid \{x \in [x_i]_a \mid d(x) = v'\} \mid\};$$

$$D_{i+1} = \{v \in V_d \mid v = \arg\max_{v} \mid \{x \in [x_j]_a \mid d(x) = v'\} \mid\};$$

Step 3: if $(D_i \subseteq D_{i+1}) or (D_{i+1} \subseteq D_i)$ $C_a = C_a \cup \{(v_a^i + v_a^{i+1})/2\}$;

2. 3. Knowledge Reduction Algorithm

During knowledge reduction, we use Genetic Algorithm(GA) to search object space in order to speed up this process. The fitness function defined as:

$$f(B) = (1-p) \times \frac{|C| - |B|}{|C|} + \rho \times \min\{r, \frac{|[\mathrm{Sin}\, S \mid S \cap B \neq \phi]|}{|S|}\},$$

where ρ is weighting function, C is condition attribute set, $S = \{M_D(i,j)_{n \times n} \mid M_D(i,j)_{n \times n} \neq \phi\}$, and r is the minimum precision control value. This algorithm process shows as follows:

(1) Randomly select Initial Population P, and cycle following operation:
(2) P -Selection, then we obtain P_1, P_2, P_3;
(3) P_1-Crossover, then we have Q_1;
(4) P_2-Mutation, then we have Q_2;
(5) Inversion- P_3, then we have Q_3;
(6) Compute new population P in terms of (2),(3),(4),(5);
(7) If genetic fitness in P no longer increases, end;
(8) Else if, return (2).

3. Structure and Function of Anomaly Detection System based on Data Ming(ADSDM)

ADEDM has three parts: data preprocessing, data mining, and anomaly behavior detection. Training stage includes data preprocessing and data mining, while detection stage includes anomaly behavior detection.

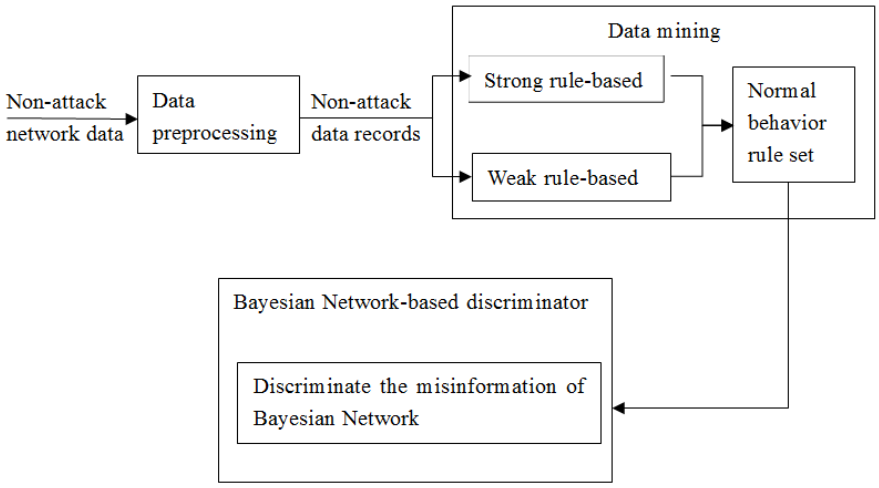

Figure 2. Training Stage of ADSDM

Figure 1 shows the training stage of ADSDM. Data preprocessing transforms non-attack data to data records that can be mined in data mining process. During data mining process, the system mines out all rule sets from the preprocessing records, and select normal behavior rule set in terms of future knowledge.

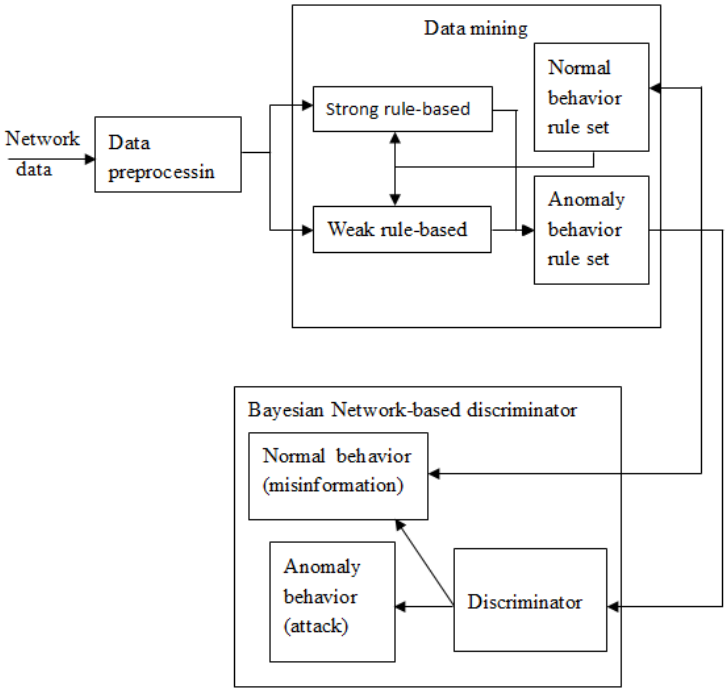

Figure 3. Detection Stage of ADSDM

3.1. Data Mining-based Anomaly Detection Technology

ADSDM based on weak rule data mining technology which is similar to recent data mining method[9]. In this paper, we mainly introduce weak rule-based technology.

3.1.1. Weak Rule-based Data Mining: Definition[10]: $I=\{i_1,i_2,\ldots,i_m\}$ is attribute set, each item $T\subseteq I$, and D is itemset. Association rules form: $A\rightarrow B$, s, c, where $A\subseteq I$, $B\subseteq I$, $A\cap B=\phi$, s is the support of rule $A\rightarrow B$, and probability denoted as $P(A\cap B)$. c is the confidence of rule $A\rightarrow B$, and conditional probability is $_c$.

For instance, association rule "SIP=202.119.36.5$\rightarrow$$DIP$=32.119.5.6.3, 0.05, 0.3", 5% records are accessing data between source host "202.119.36.5" and destination host "32.119.56.3".

Weak rule-based data mining, that is to select data records set whose support is lower than while confidence level is bigger than given threshold.

This data mining method selects low happen probability records. Due to these records are difficult to calculate, thus, we have to change values of support and confidence, and then multiply adjusting coefficient, $\alpha: Adapt\_Para(x)=Para(x)\times\alpha$

For instance, if α =100, for slow scan of a particular host, $s(dip=10.10.10.10, dport=80)=10\%$, $s(dip=10.10.10.10, dport=1259)=0.1\%$, then adjustment value $Adapt\_s(dip=10.10.10.10, dport=80)=10$, $Adapt\_s(dip=10.10.10.10, dport=1259)=0.1$.

3.1.2. Anomaly Behavior Discrimination: Attributes of network behavior are not independent each other[11]. Bayesian network describes associated conditional probability distribution, and shows relationship between each variables. There are two parts make up Bayesian network, that is Directed acyclic Graph(DAG) and Conditional Probability Table(CPT).

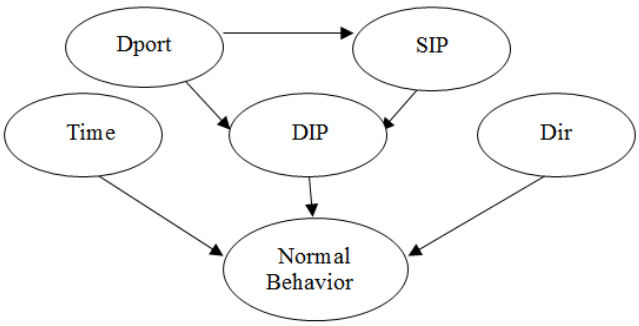

Figure 4. Bayesian Network of Anomaly Behavior Discrimination

Where $CPT_{ij} = P(node = j \mid parent\_node = i)$, $CPT_{ij} \geq 0(\forall i, j)$, and $\sum_j CPT_{ij} = 1(\forall j)$. State value of father node is pluralistic while Probability of child node is determined by probability of father node and *CPT*: $P(node = j) = P(parent\_node = i) \times CPT_{ij}$.

3.2. Data Mining-based Algorithms

Algorithm 1: Association rule training algorithm for normal data without attacks.

Input: Attribute records data set after data preprocessing: FR_T , and application layer data set: DR_T .

Output: Weak rule-based flow attribute data set: $WFAR = \{wfa_1, wfa_2, \ldots, wfa_k\}$, and weak rule-based application layer data set: $WAAR = \{waa_1, waa_2, \ldots, waa_q\}$.

Step 1: By means of first-in first-out(FIFO), add FR_T into flow attribute data record queue FR_QUEUE , and add DR_T into DR_QUEUE .

Step 2: Where $WFAR = WAAR = \phi$.

Step 3: Mine out all association rules from FR_QUEUE and DR_QUEUE , including support, time, and confidence. Then generate flow rule set FAR_TMP and data rule set AAR_TMP .

Step 4: For $\forall v1 \in FAR\_TMP$, if $v1$ is lower than support threshold while bigger than confidence threshold, then $WFAR = WFAR \cup \{v1 \times \alpha\}$.

Otherwise, for $\forall v2 \in AAR\_TMP$, if $v2$ is lower than support threshold while bigger than confidence threshold, then $WAAR = WAAR \cup \{v2 \times \alpha\}$.

Algorithm 2: Data Ming-based Algorithm for Intrusion Detection.

Input: FR_T , DR_T , $WFAR = \{wfa_1, wfa_2, \ldots, wfa_k\}$, and $WAAR = \{waa_1, waa_2, \ldots, waa_q\}$.

Output: Suspicious association rule set *SRR*, and support-suspicious association rule set *SDR*.

Step 1: By means of FIFO, add FR_T into flow attribute data record queue FR_QUEUE, and add DR_T into DR_QUEUE.

Step 2: Where $SRR = SDR = \phi$.

Step 3: For each $wfa_i \in WFAR$, $Rule = wfa_i.Rule$, mine out all association rules of data records in terms of $Rule$, and generate association rule set SC_TMP.

Step 4: For each $v \in SC\_TMP$,

if $(v.Time \in wfa_i.Time)$ then // wfa_i has different support and confidence in different time.

if (support of v is lower than support of wfa_i, and confidence of v is bigger than confidence of wfa_i at the same time)

then $SRR = SRR \cup \{v\}$ // v is suspicious rule

$SDR = SDR \cup \{$data records of support $v\}$

endif

else

if (support of v is lower than support of $Rule$, and confidence of v is bigger than confidence of $Rule$)

then $SRR = SRR \cup \{v\}$ // v is suspicious rule

$SDR = SDR \cup \{$data records of support $v\}$

endif

endif

Step 5: Add $WAAR$ into DR_QUEUE, and mine out suspicious rule set according to Step 3 and 4.

Step 6: Output SRR and SDR.

Algorithm 3: Discriminate the Probability: whether suspicious behavior is *Normal behavior* or not.

Input: SRR and SDR

Output: Alarm Information

For each $Rule \in SRR$

Step 1: Compute the probability in terms of CPT values and $Rule$-support data:

$P(SIP) = P(SIP | DIP) port(P,$

c'

$$P(Normal-behavior) = P(Normal-behavior \mid Time, DIP, Dir) \times P(Time) \times P(DIP) \times P(Dir)$$

Step 2: If $P(Normal-behavior) \geq P\_Threshold$

then{// $P\_Threshold$ is normal, thus, this suspicious behavior is normal(misinformation)

{// c', where β is weight

}

}

else

Alarm Information

endif

4. Network Security Situational Awareness System

In order to avoid confusion and convenience of narration, we introduce some related definition.

(1) Security situation: Security situation information includes time dimension and spatial distribution dimension.

(2) Security event: Alarm information generated from network security situation sensor, and initiated by intrusion behavior. Security event can be denoted as a multivariable equation:

$$e_i = \{DetectTime, eventType, attack_i, srcIP_i, desIP_i, srcPort_i, desPort_i, protocol_i, sensorID_i, confidence_i, severity_i, other_i\}.$$

(3) Security situation modeling: Analyze all alarm events generated from security sensor, and then from network security situation.

4. 1. Network Security Situation Generation

4.1.1. Knowledge Discovery-based Association Rule Mining: Data source used for knowledge discovery derives from two parts: alarm event sets generated from simulation attacks and historical alarm event sets. Knowledge discovery means discovery and selection situation relation knowledge from alarm event sets.

(1) Reduct and Filtering for Alarm Events.

Evidence Theory(DS)-based[12] establishment of alarm event filtering mechanism is a statistical analysis for program in terms of alarm event confidence. Firstly, program automatically count distribution of different alarm events. Secondly, based on DS, and delete insignificant events.

(2) Knowledge Discovery[14,15,16]

We obtain situation knowledge from alarm event sets in terms of frequent pattern and sequential pattern. Frequent pattern is relation between several event attributes, and its aim is to get the regularity of event attributes. Sequential pattern is sequence relation between events, and its aim is to figure out time series relation and causal relation. In general, we consider relation between attack, sip, dip, sport, protocol.

4.1.2. Network Security Situation Generation Algorithm: We calculate the risk level of network nodes by using alarm events after fusion. Mainly considered following factors: alarm confidence *c*, alarm level *s*, and resource impact degree *m*. In addition, we also considered security protection level p_n and recovery factor r_n.

Security situation assessment of single node is computed as follows:

$$S_n(t) = \frac{\sum_{t_i \in T_t} c_i s_i m_i}{p_n r_n} \tag{1}$$

Because different nodes have different position and function in network, thus, we have to consider weight of nodes w_n:

$$S_N(t) = \sum_{n \in N} w_n S_n(t) \tag{2}$$

Attacking index of node is computed as follows:

$$A_n(t) = \sum_{t_i \in T_t} c_i s_i \tag{3}$$

By using these formula, we can draw the assessment-time curve. Nevertheless, it is still insufficient to figure out the space distribution of security situation. Thus, we use graph theory method to describe the distribution of security situation. Definition is as follows:

Definition: Network Security Situation Graph G is a 5-tuple: $G = \{V(G), E(G), A(G), W(G), S(G)\}$.

(1) $V(G) = \{v_1, v_2, \ldots, v_n\}$ is vertex set, and each vertex match one type of network node. $NodeType = \{rtr, sur, hst\}$, where *rtr* is router, *sur* is server, *hst* is host.

(2) $E(G) = \{e_1, e_2, \ldots, e_m\}$ is directed edge set determined by edge type and ordered pair of vertexs.

(3) $A(G) = \{A(v_1), A(v_2), \ldots, A(v_n)\}$, and $v_i \leftrightarrow A(v_i)$. Where $A(v_i)$ is alarm event set, and each element in $A(v_i)$ is a alarm event.

(4) $W(G) = \{w_1, w_2, \ldots, w_n\}$ is weight of nodes, $v_i \leftrightarrow w(v_i)$, and $w(v_i) \in [0,1]$.\

(5) $S(G) = \{s_1, s_2, \ldots, s_m\}$ is assessment of nodes, and $v_i \leftrightarrow s(v_i)$.

Network Security Situation Generation Algorithm is as follows:
Input : Network topology model, alarm events, DS, association rule.
Output: Network Security Situation Graph
BEGIN

(1) Let T represent network topology mode, G represent network security situation graph.
(2) G=initialize(T).
(3) G.W=weigh(G,T)
(4) Let Eve represent the alert events, S represent network security situation sensors.
(5) Eve=fromSensor(S).
(6) Eve$'$=D-S(Eve).
(7) Let C represent the correlation rules.
(8) Eve$''$=correlate(Eve$'$,C).
(9) For each node $v_i \in G$
(10) G.A(v_i)=append(v_i, Eve$''$).
(11) For each node $v_i \in G$
(12) G.S(v_i)=assess(v_i, A,W$'$)
(13) Where A is the alarm event of v_i, W is the weight of v_i.
(14) Return G.
END

5. Experiments and Analyses

This section is divided into three parts. The first is experiment based on Rough Set Theory. The second parts shows the result of ADSDM experiment. The third parts evaluates the performance of network security situation.

5.1. Rough Set Theory-based Experiment

We chose KDD CUP 99 Data Sets as data source, and there are 41 attribute values in each network connection record which equal to condition attribute in information system.

In this experiment, we used 31 attributes in these 41 attributes values. There are 7 basic attributes, 11 connection content attributes, 18 attributes based on host and time. The training data set has 13,107 connected records, and test data set has 26,214 records. Distribution of data connection is displayed in Table 1.

Table 1. Distribution of Data

	Training data	Test data
Type	Percent(%)	Percent(%)
Normal	59.952796	9.727629
Probe	1.213095	1.224527
DoS	38.735037	87.308521
U2R	0.015259	0.003815
R2L	0.083925	1.735714

Table 2. Confusion Matrix of Test Data

Type	Normal	Probe	DoS	U2R	R2L	%correct
Normal	2113	0	437	0	0	0.83
Probe	74	1	246	0	0	0.003
DoS	24	11	22852	0	0	0.99
U2R	1	0	0	0	0	0.0
R2L	25	1	429	0	0	0.0

In Table 2, we can figure out that Normal and DoS keep high connection detection accuracy. Because DoS can pose a great danger to network, thus, the result of this experiment significant.

5.2. ADSDM Experiment

We used data set proposed by Defense Advanced Research Projects Agency (DARPA)[11]. In this experiment, there all totall 57 different kind of attacks. Within 180 attacks, the system detected 141 attacks, so the detection rate is 78.3%. This weak rule-based system can detect IP sweep and Port sweep satifactorily.

During this test, we selected three different speed to evaluate the program: normal speed, tenfold speed, and 1/10 speed. The result shows that, at tenfold speed, system has missed detection, while at 1/10 speed, system has more misinformation. However, system can decrease the attacks misinformation after discriminator.

Table 3. Results of ADSDM Experiment

Type	Number of attacks	Number of detection	%Rate
Probe	37	33	89.2
DoS	59	52	88.1
R2L	49	36	73.5
U2R	31	19	61.3
Data	4	1	25.0
Total number	180	141	78.3

5.3. Network Security Situation Experiment

In this part, Table 4 display the experiment results. The alarm events those are shown in Table 4 represent the alarm information after data fusion. The target addresses are (the host addressed) under attacks. In the test data set, the victim host is $host_v(131.84.1.31)$, and controlled hosts are $host_l(172.16.112.10)$, $host_m(172.16.115.20)$ and $host_n(172.16.112.50)$.

Table 4.

Target Addresses Amount	Event Types	c	s	m
172.16.112.0/24 15	RPC sadmind UDP PING	0.8	1	1
172.16.112.0/24 15	RPC portmap sadmind request UDP	0.7	2	2
172.16.113.0/24 15	RPC sadmind UDP PING	0.8	1	1
172.16.113.0/24 15	RPC portmap sadmind request UDP	0.7	2	1
172.16.114.0/24 15	RPC sadmind UDP PING	0.8	1	1
172.16.114.0/24 15	RPC portmap sadmind request UDP	0.7	2	2
172.16.115.10 90	NETMGT_PROC_SERVICE	0.4	3	3
172.16.115.10 90	CLIENT_DOMAIN overflow attempt RPC sadmind query with root	0.5	2	3
172.16.115.10 90	credential attempt UDP RPC portmap sadmind requset UDP	0.7	2	2
172.16.112.20 90	RPC sadmind UDP NETMGT_PROC_SERVICE	0.4	3	3
172.16.112.20 60	CLIENT_DOMAIN overflow attempt RPC sadmind query with root	0.5	2	3
172.16.112.20 60	credential attempt UDP RPC portmap sadmind requset UDP	0.7	2	2

		RPC sadmind UDP			
172.16.112.50		NETMGT_PROC_SERVICE	0.4	3	3
60					
		CLIENT_DOMAIN overflow attempt			
172.16.112.50		RPC sadmind query with root	0.5	2	3
60					
		credential attempt UDP			
172.16.112.50		RPC portmap sadmind requset UDP	0.7	2	3
60					
202.77.162.213		RESERVICES rsh root	0.6	3	2
60					
202.77.162.213		RESERVICES rsh root	0.6	3	2
45					
202.77.162.213		RESERVICES rsh root	0.6	3	2
30					
131.84.1.31		DDOS	0.4	3	3
10767					

Figure 6. Shows the Dynamic Changes of Network Security Situation Accurately

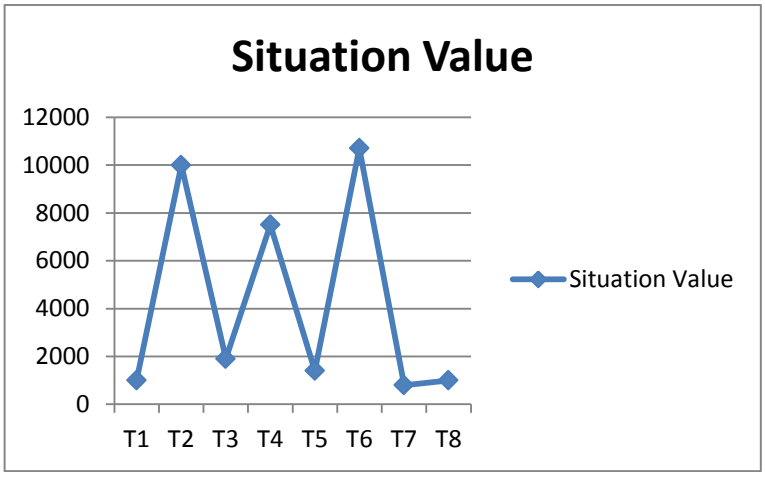

Figure 4. Variable Graph of Network Security Situation

6. Conclusion

People's work and life has been closely linked with the network. However, there are a lot of potential threats on the internet, and our security need protection. In this paper, we proposed several method to detect network intrusion, and we also introduced some algorithm to figure out this network problem. During the experiment, we figured out that association between several different data mining methods works more effectively than usage of one method. For instance, combination of Rough set theory and genetic algorithm can improve intrusion detection accuracy. The results of our experiments show that the methods and system proposed in this paper effectively detect attacks from attacker or other IP addresses, simultaneously, these methods can also protect our security.

References

[1] Denning D. E., "An intrusion-detection model", IEEE Trans on Software Engineering, vol. 13, no. 2, **(1987)**, pp. 222-232.
[2] Portnoy L., Eskin E. and Stolfo S., "Intrusion detection with unlabeleddata using clustering", **(2001)**.
[3] Anderson, "Computer Security Threat Monitoring and Surveillance: [Technical report]", **(1980)**.
[4] Heberlein L. T., Dias G., Levitt K., Mukherjee B., Wood J. and Wolber n, "A Netork Security Monitor", In: Proceedings of 1990 Symposium on Research in Security and Privacy, **(1990)**.
[5] Lee W. and Stoifo S. J., "Data mining approaches for intrusion detection", In: Proc. of the 7th USENIX Security Symposium. San Antonio, TX, **(1998)** Jan.
[6] Forrest S., Perelson A. S. and A1leu L., "Self-nonself discrimination in a computer", In: Proc. of the 1994 IEEE Symposium on Research in Security and privacy, **(1994)**.
[7] Lane T. and Brodley C. E., "Detecting the Abnormal: Machine Learning in Computer Security: [Technical Report ECE-97-1], **(1997)**.
[8] L. Renpu and W. Zheng-ou., "Mining classification rules using rough sets and neural networks", European Journal of 0perational Research, **(2003)** Sept.
[9] Barbara D., Wu N. and Jajodia S., "Detecting Novel Network Intrusions using Bayes Estimators", First SIAM International Conference on Data Mining, **(2001)**.
[10] Han J. and Kamber M., "DATA MINING: Concepts and Techniques, Higher Education Press, **(2001)**.
[11] Lippmann R., "The 1999 DARPA Off-Line Intrusion Detection Evaluation", Computer Networks, vol. 34, no. 4, **(2000)**, pp. S79-595.
[12] Hart J. W., Pei J. and Yin Y. W., "Mining frequent patterns without candidate generation[J]", ACM SIGMOD Record, vol. 29, no. 2, **(2000)**, pp. 1-12.
[13] Dempster A., "Upper and lower probabilities induced by multi-valued mapping[J]", Annals of Mathematical Statistics, vol. 38, no. 2, **(1967)**, pp. 325-339.
[14] Boyer R. S. and Moore J. S., "A Past String Searching Algorithm", Com munications of the ACM, **(1977)**.
[15] Roeseh M., "Snort: Lightweight Intrusion Detection for Networks", In: Proceedings of the USENIX LISA Systems Administration Conference, **(1999)**.
[16] Vapnik V. N., "Statistical learning theory", Adaptive and learning systems for signal processing, communications and control, New York: Wiley, **(1998)**.

Acknowledgments

This work was supported by the Natural Science Foundation of higher education in Anhui Province (KJ2016A127), and the Professional leaders Research Foundation of Higher Vocational Colleges in Anhui Province (2014 Year).

Author

Yang Hui-jun, He received his M.S. degree in computer application from Southeast university of Computer science and engineering in Nanjing, China. He is currently an assistant professor of the Department of information and services of Anhui Institute of International Business. His research interest is mainly in the area of applications from Internet of Things, Cloud computing. He has published several research papers in scholarly journals in the above research areas and has participated in several books.