An Intrusion Detection Algorithm based on D-S Theory and Rough Set

http://dx.doi.org/10.3991/ijoe.v9iS6.2794

Lifang Wang, Shuhai Zhang
North University of China, Taiyuang, China

***Abstract*—Intrusion detection system is a kind of network security system, which can alarm suspicious transmission or take active response measures when it real-time monitors network transmission and discovers suspicious transmission. But intrusion detection system has many problems such as wrong detection of intrusions, missed intrusions, poor real-time performance. In order to improve the performance of intrusion detection system, this paper proposes an intrusion detection algorithm based on D-S theory and Rough Set. The algorithm uses the attribute reduction algorithm in rough set to eliminate redundant attributes, form the simplest attributes set, overcome the traditional D-S theory relying on expert knowledge to provide evidence and makes each evidence body mutual independence. So it improves the evidence synthesis efficiency, shortens the evidence synthesis time and reduces the conflict phenomenon of evidence synthesis. On this basis, the paper builds an intrusion detection model based on D-S theory and rough set, and the experimental results demonstrate that the model has higher detection rate and lower false detection rate.**

***Index Terms*—D-S (Dempster-Shafer) theory, Rough Set, Intrusion Detection**

I. INTRODUCTION

With the rapid development and the wide application of the computer network technology, network security problems become more and more prominent. Establishment of effective intrusion detection system to protect the security of computer information system has become more and more important. Intrusion detection system (IDS) is a new safety assurance technology following traditional security ways such as firewall and data encryption. IDS can identify and response to improperly use computer and network resource [1-2].Now the methods used to establish the intrusion detection model are mainly neural network, SVM and immune network ,etc. The decision factors of intrusion behavior are often very complicated. Many intrusion detection systems only match and filter a few features to directly led to the high error rate. This paper uses the rough set attribute reduction algorithm to obtain evidence and uses the decision rule strength to obtain the basic probability assignment of evidence. So the method reduces the subjectivity of the basic probability assignment and the conflict phenomena of evidence synthesis. And this paper applies the method in intrusion detection to improve the detection accuracy.

II. D-S THEORY

Let Θ represents a domain set of the evidence *X*'s all values and all elements within the Θ are mutually exclusive, then Θ is X's recognition framework. For the recognition framework Θ, if the function $m:2^{\Theta}\to[0,1]$ meets [3-5]:

$$m(\varnothing)=0,\ \sum_{A\subset\Theta} m(A)=1 \tag{1}$$

In the equation (1), $m(A)$ is $A's$ basic probability assignment. On the basis of the basic probability assignment function, the trust function can be defined as [6-9]:

$$Bel(A)=\sum_{B\subset A} m(A),\ \forall A\subset\Theta \tag{2}$$

In the equation (2), $Bel(A)$ is the possibility measure sum of $A's$ all subsets.

Two evidences on the recognition framework Θ can be efficiently synthesized by the synthesis rules in D-S theory [10]. Two or more evidences can be synthesized two by two by the rules. Let $A=A_i$ and $B=B_j$ be two independent evidence in Θ, $m_1(\cdot)$ and $m_2(\cdot)$respectively denote their basic probability assignment, the synthesized result in Θ is: $m(\cdot)=m_1(\cdot)\oplus m_2(\cdot)$, calculated by equation (3):

$$m(c)=\begin{cases}\dfrac{\sum_{A_i\cap B_j=c} m_1(A_i)m_2(B_j)}{\sum_{A_i\cap B_j=\varnothing} m_1(A_i)m_2(B_j)} & c\neq\varnothing\\ 0 & c=\varnothing\end{cases} \tag{3}$$

The limitations of D-S theory are as follows: the synthesis rules require evidence body mutual independence and are sensitive to the change of the basic probability assignment and require the basic probability assignment value which is reasonably given and can not be randomly assigned.

III. ROUGH SET THEORY

The main idea of rough set theory is as follows [11]: keeping the knowledge base classification ability unchanged, the theory deletes irrelevant or unimportant attributes. Through the knowledge reduction, unnecessary attributes are removed and knowledge representation is simplified and essential information is not lost. So the concept of classification rules can be derived to improve the decision precision.

Definition 1[12]: If Q is an equivalence relation set, $Q \neq \varnothing$, Q is the intersection of all equivalence relations $\cap Q$ and is also an equivalence relation, and is called indistinguishable relation in Q and denoted by $ind(Q)$.

Definition 2[13]: in Knowledge expression system $s = \langle U, A, V, f \rangle$, U is a non-empty finite set of object, A is a non-empty finite set of attributes and called attribute set. Attribute set can be divided into condition attribute set C and decision attribute set D, $A = C \cup D$, $C \cap D = \varnothing$, V is the values range for attribute $a \in A$. $f : U \times A \rightarrow V$ is a information function which makes the attribute a of any element have a unique value in V. And the information system is called decision information system.

Definition 3[14]: For $X \subseteq U$, R is an equivalence relation in U, $\underline{R}X = \{x \in U \mid [X]_R \subseteq X\}$ is R lower approximation set of X. $\overline{R}X = \{x \in U \mid [X]_R \cap X \neq \varnothing\}$ is R upper approximation set of X. $pos_R(X) = \underline{R}X$ is call R positive domain of X. $neg_R(X) = U - \overline{R}X$ is called R negative domain of X.

IV. INTRUSION DETECTION SYSTEM BASED ON D-S THEORY AND ROUGH SET

This paper deals with a large number of network access data in the characteristic level by rough set, and then judges the redundant attributes by the knowledge expression system simplified and gets the simplest combination of the characteristic attributes to simplify the characteristic data. Finally the final invasion decision results can be obtained by the synthesis of D-S theory in decision level. The combination of rough set and D-S theory is as follows:

1. The decision attributes can be obtained by the reduction effect of rough set theory to form evidence.
2. The basic probability assignment of evidence can be obtained by the attribute importance measure of rough set.

The intrusion detection model based on rough set and D-S theory, (as shown in Figure 1).The model first pretreats collection of data, chooses training samples, reduces attributes in decision tables, produces reduced output rules to construct rule base of safe system and intrusion detection detector. The initial intrusion model needs gradually perfect and improvement in subsequent studies to reach the best detection effect.

From intrusion model we can clearly see that intrusion detection algorithm mainly involves some following problems:

1. Intrusion data discrete. IDS analyzes the data which includes network data and host data. The analysis of network packet is a key point in the current intrusion detection study. Compared to host log data, network data is more complex and multiple and thus greatly increase the intrusion difficulty of network attack. To improve the intrusion effects, the large amount of collected data-points needs be dispersed by the method of equal frequency division.
2. Attribute reduction. The attributes of collected data sets are structured and reduced .And redundancy intrusion attributes are removed

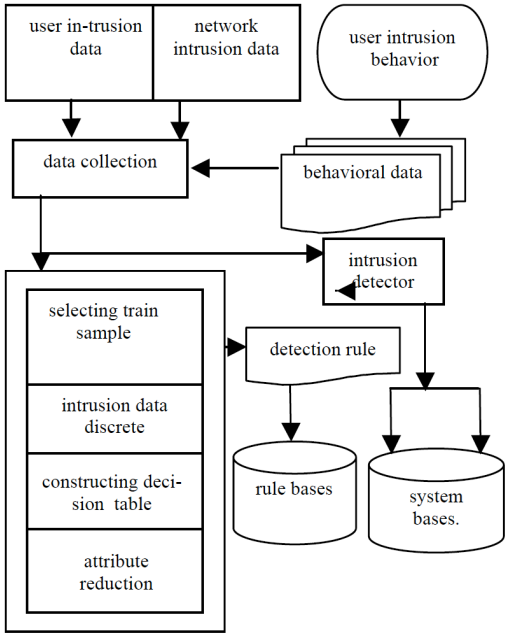

Figure 1. Model based on rough set attribute-weighted

3. Generate intrusion detection rules. After attributes reduction, the model have deleted redundancy attributes, constructed decision table, and derived rules from decision table. Then it detects these rules and puts them into rule database. Based on those rules in rule database, safety detector tests intrusion datum and behaviors.

A. Evidence Formation based on Attribute Reduction

The redundant attributes in decision table can be eliminated by the attribute reduction and the key attributes be retained. The decision table being reduced and the decision table not being reduced have the same knowledge.

This paper presents the relative attribute reduction algorithm which is as follows:

Input: a decision table $S = (U, B, C, V, f)$ which can also be expressed as $S = (U, B, C)$.

Output: a relative reduction $S = (U, R, C, V, f)$ of the decision table, which can be expressed as $S = (U, R, C)$

Step 1: Initialize R to an empty set $\varnothing$.

Step 2: Calculate the condition entropy $H(C \mid B)$ of the decision attribute C and the relative condition attribute B.

Step 3: Calculate the condition entropy $H(C \mid \{b\})$, of each condition attribute b in B corresponding to C, $b \in B$ and putting b corresponding to the minimum condition entropy into R, $R = R + \{b\}$ $A = B - \{b\}$.

Step 4: Calculate $H(C \mid R) - H(C \mid R \cup \{a\})$, $a \in A$ for each attribute of condition attribute set A and putting a corresponding to the maximum difference into R, $R = R + \{a\}$, $A = A - \{a\}$.

Step 5: if $H(C \mid B) = H(C \mid R)$, then go to step 7.

Step 6: Calculate $H(C \mid R - \{r\}) - H(C \mid R)$ for each attribute r corresponding to R (not including a which is the latest attribute added). If the difference value is less than a given threshold value β, make $R = R - \{r\}$, $A = A + \{r\}$, and then go to Step4, and continue to calculating.

Step 7: After reduction, output $S = (U, R, C)$. Condition information entropy is as follows:

$$H(C \mid R) = -\frac{|U - \cup(X_i \cap Y_i)|^2}{|U|^2} \lg \frac{1}{|U|} \quad (4)$$

where $\{X_i\}$ is the lower approximation corresponding to the division $U / C = \{Y_i\}$

B. The Structure of Basic Probability Assignment

In decision table S, $e \in B$, $U / e = \{X_1, X_2, \ldots, X_m\}$, $U / C = \{Y_1, Y_2, \ldots, Y_n\}$, $H = \{e, C\}$, $U / H = \{H_1, H_2, \ldots, Y_j\}$. If $A_t = \left\{ H_i \mid \bigcup_{H_i \subset X_t} H_i = X_t, X_t \in U / e \right\}$, A_t is equivalent class D division of $X_t \mid A_t \mid = k$. The value domain of C is $V_C = \{c_1, c_2, \ldots, c_n\}$, $k \le n$. Calculated the decision rule strength of each object x in H_i, k decision rule strength can be obtained. The definition of decision rule strength is shown in definition 4.

Definition 4[15]: In decision table $S = (U, B, C, V, f)$, B and C are respectively the condition attribute set and the decision attribute set. For $\forall x \in U$, the upper approximation of x corresponding to B is $\overline{B}x$ and the upper approximation of x corresponding to C is $\overline{C}x$. Then the strength of decision rule $f(x, B) \to f(x, C)$ is shown in (5):

$$\eta = | \overline{B}x \cap \overline{C}x | / | \overline{B}x | \quad (5)$$

When $G \subset B$, and many condition attributes are included in G, the strength of decision rule $f(x, G) \to f(x, C)$ is shown in (6):

$$\eta = | \overline{G}x \cap \overline{C}x | / | \overline{C}x | = | [x]_G \cap [x]_C | / | [x]_G | \quad (6)$$

When $e \in B$, the strength of decision rule $f(x, e) \to f(x, C)$ is shown in (7):

$$\eta = | [x]_e \cap [x]_C | / | [x]_e | \quad (7)$$

When $\exists y \in U$ meets $f(x, e) = f(y, e)$ and $f(x, C) \neq f(y, C)$, decision expansion rules are used and the definition of decision expansion rules is shown in definition 5.

Definition 5: in decision table $S = (U, B, C, V, f)$, $e \in B$, $x, y \in U$, there are decision rules $f(x, e) \to f(x, C)$ and $f(y, e) \to f(y, C)$. When $f(x, e) = f(y, e)$ and $f(x, C) \neq f(y, C)$, the decision expansion rule is $f(x, e) \to \{f(x, C), f(y, C)\}$. The strength of decision expansion rule is shown in (8):

$$\eta = \left| \overline{ex} \cap \overline{Cx} \right| / \left| \overline{ex} \right| * \left| \overline{ey} \cap \overline{Cy} \right| / \left| \overline{ey} \right| \quad (8)$$

The theorem determining the basic probability assignment of evidence is as follows:

Theorem if Q is a proposition in recognition framework Θ, $Q \in 2^{\Theta}$, decision table $S = (U, B, C, V, f)$, $e \in B$, the value domain of decision attribute $V_C = \Theta$, $H = \{e, C\}$, $X_j \in U / e$, $H_i \in U / H$, $A_j = \{H_i \mid \cup H_i = X_j, H_i \subset X_j, X_j \in U / e\}$, $|A_j| = k$, the basic probability assignment of proposition Q is shown in (9):

$$w(Q) = \eta_Q / \left(1 + \prod_{i=1}^{k} \eta_i \right) \quad (9)$$

C. Algorithm Application

In order to demonstrate the algorithm application process, this paper uses three attributes to compose condition attribute to judge invasion. These three attributes are respectively expressed as E_1, E_2 and E_3. In decision table of intrusion detection system, E_1, E_2 and E_3 can be respectively used as condition attribute, $B = \{E_1, E_2, E_3\}$. C is the decision attribute and the value domain is the recognition framework Θ of intrusion detection. History intrusion data collected are pretreated and attributes are dispersed to form the decision table of intrusion detection, which is shown in Table I.

TABLE I.
DECISION TALE OF INTRUSION DETECTION

U	*C*	*B*		
		E1	*E2*	*E3*
u1	Dos	1	0	0
u2	R2L	1	0	1
u3	Dos	2	1	1
u4	R2L	2	2	1
u5	R2L	1	1	1
u6	Probe	1	2	0
u7	R2L	2	2	0
u8	Probe	0	2	0
u9	Probe	1	2	0
u10	R2L	0	1	1

When the threshold $\eta = 0.1$, condition entropy $H(C \mid R - \{E_3\}) - H(C \mid R) < 0.1$. Condition attribute E_3 is redundant on the basis of the entropy. The basic probability assignment of E_1 and E_2 can be respectively calculated by (5)~(9). The basic probability assignment of evidence and the basic probability assignment synthesized are shown in Table II.

TABLE II.
BPA OF EVIDENCE AND SYNTHESIZED

evidence	Basic probability assignment(BPA)			
	{Probe,R2L, DoS}	{Probe}	{DoS}	{R2L}
E_1	0.0489	0.2715	0.4299	0.2439
E_2	0.0963	0.2311	0.3891	0.2889
synthe-sized	0.0107	0.2229	0.5166	0.2494

When $\exists E_1, E_2 \subset \Theta$, $p(E_1) = \max\{P(E_i), E_i \subset \Theta\}$ and $p(E_2) = \max\{m(E_i), E_i \subset \Theta and E_i \neq E_1\}$.

When $\begin{cases} p(E_1) - p(E_2) > \eta_1 \\ p(\Theta) < \eta_2 \\ p(E_1) > p(\Theta) \end{cases}$, η_1 and η_2 are the preset threshold. E_1 is judgment result. When the value of η_1 and η_2 is 0.1, intrusion type is DoS which is determined by the above reasoning decision rule. But how to determine the specific type of DoS needs further analysis. It can be seen from the synthesized results that the uncertainty degree is gradually reduced, and the reasoning obviously concentrates to the results set {DoS}. These show that much evidence synthesis performance is better than the single-evidence synthesis.

D.Experimental Results and Analysis

In order to prove the effectiveness of the algorithm, this paper use data sets KDD99 to test. Because the data amount in KDD99 is too large, 10000 data selected from the data set are as the experimental data. Data selected contain as much as possible the common attack method, and ensure each attack method has a certain quantity of data, which includes normal records and 4 class attack record. The proportion of all kinds of attacks in attack record is as follows: Dos is 94%, R2L is 2%, U2R is 2%, Probe is 2%. Therefore, the identification intrusion detection framework $\Theta = \{DoS, \Pr obe, U2R, R2L\}$ can be established.

There are 32 continuous attributes in 41 characteristic attributes of TCP record(http: //kdd.ics.uci. edu/databases /kddcup99/kddcup99.html). First, Naive Scaler algorithm disperses continuous attributes, so all the characteristic attributes are transformed into discrete attributes. Then attribute reduction algorithm reduces attributes of 10% data set selected.

Because the reduction method has been realized in Rosetta software, we use Rosetta software to reduce attributes of the data set. After reduced, normal connection records attribute set has 26 items:{ l,2, 4, 6, 7, 8, 9, 10, 13, 15, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28,34, 35, 36, 37, 38, 40}.

To assess the performance of the intrusion detection method, two performance parameters are introduced: detection rate α and false detection rate β. They are defined as follows: $\alpha = H / L$, $\beta = O / Z$ H is the invasion number of correct detection, O is the invasion number of misjudgment, L is the total invasion number and Z is total non-intrusion record number. Detection rate and false detection rate can measure the performance of the intrusion detection system. Intrusion detection system is always expected to achieve high detection rate and low false detection rate. The experimental results are shown in Table III.

TABLE III.
INTRUSION DETECTION RESULTS

Actual results	Detection results					α	β
	R2L	U2L	DoS	Probe	normal		
R2L	178	2	0	0	17	0.90	0.03
U2L	2	147	0	0	51	0.75	0.04
DoS	0	1	197	0	1	0.99	0.11
Probe	1	1	24	169	3	0.85	0.06
normal	2	1	5	1	9297	0.99	0.01

From table 3, we can see that the network intrusion detection algorithm based on rough set and D-S theory in this paper has high detection rate and low false detection rate. For normal connection, detection rate of DoS and Probe attack is very high, and the result is better than the best detection result of KDD Cup99:97.3% (DoS) and 75.0% (Probe).Therefore, this algorithm is very good to satisfy the security detection requirements of intrusion detection system.

To further test the algorithm effectiveness, lots of comparison experiment in different types of intrusion algorithms, such as data mining(DM), Support Vector Machine (SVM), BP(back propagation)Neural Network, etc are made and experimental results are obtained as follows(See Figure 2 and Figure 3) By comparing the experimental results we can see that the algorithm based on rough set and D-S theory is better than other methods of intrusion detection in the detection rate and the false detection rate. Because data mining applied to intrusion detection requires a large amount of data, and the intrusion methods based on Support Vector Machine and BP neural network need high training speed and large computing amount etc, so the effectiveness and intrusion rate of the detection algorithms on the base of the BP neural networks, support vector machines and data mining is significantly lower than the algorithm based on rough set and D-S theory. All these are fully proved that the rough set and D-S theory used in intrusion detection systems are very effective.

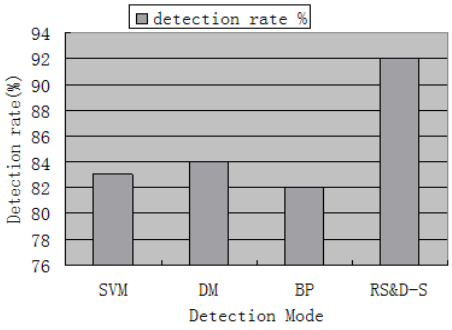

Figure 2. Detection rate comparison

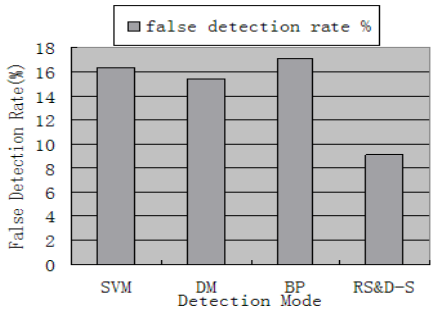

Figure 3. False detection rate comparison

V. CONCLUSION

The intrusion detection algorithm based on rough set and D-S theory can give full play to their respective advantages of the two kinds of uncertainty reasoning theory. Through the experiment on the KDD99 data set, it shows that the hybrid model has higher detection rate and lower false detection rate. Future research will focus on to attribute reduction algorithm optimization to improve the detection speed.

ACKNOWLEDGMENTS

REFERENCES

[1] Kang D K, Fuller D, Hona R V, "Learning classifiers for misuse detection using a bag of system calls representation," *Proc. Of IEEE International Conference on Intelligence and Security Information 2011. IEEE Computer Society, Atlanta, GA, USA*, vol.3, 2011,pp.510-514.

[2] Walker K M, Daniel F S, Lee Badger M. "Confining root programs with domain and type enforcement," *Proc. Of the 14th USENIX Security Symp, Focusing on Applications of Cryptography. San Jose, USENIX*, vol.8,2010,pp.18-25.

[3] Tan K M C, Maxion R A. ""Why 6?" defining the operational limits of stide, an anomaly-based intrusion detector,"*Proceedings of the 2010 IEEE Symposium on Security and Privacy. Washington, DC, USA, IEEE Computer Society*, vol.9,2010,pp.178.

[4] Garo De-bin, Reiter M K, Song D. "Gray-box extraction of execution graphs for anomaly detection,"*Proceedings of the 15th ACM Conference on Computer and Communications Security Washington DC, USA*,vol.7, 2009,pp.88-95

[5] Xia Yongxiang, Shi Zhicai. "An Incremental SVM for Intrusion Detection Based on Key Feature Selection,"*Proc. Of the 3rd International Symposium on Intelligent Information Technology Application. Nanchang, China*,vol.6,2011,pp.205-208.

[6] Wespi A, Dacier M ,Debar H. "An intrusion detection system based on the teiresias pattern-discovery algorithm,"*Proceedings of the 2008 European Institute for Computer Anti-Virus Research Conference*. vol.1,2010,pp.778-792.

[7] Debar H , Dacier M , Nassehi M , et al. "Fixed vs. Variable-length pattern for detection suspicious process behavior," *JESO-RICS 2008. 10th European Symposium on Research in Computer Security. LNCS, Louvain-la-Neuve, Belgium*, vol.1,2008,pp.10-15.

[8] Han Sang-jun, Cho S-B. "Evolutionary Neural Networks for Anomaly Detection Based on the Behavior of a Program," *IEEE Transactions on systems, man, and cybernetics-part B: cybernetics*,vol.45,pp. 555-563, April 2010.

[9] Sujatha P K, Kannan A, "A Behavior Based Approach to Host Level Intrusion Detection using Self-organizing Maps," *First International Conference on Emerging Trends in Engineering and Technology. IEEE Computer Society*,vol.4,2009,pp.268-1103.

[10] Michael C, Ghosh A. "Two state-based approaches to program based anomaly detection," *Proceedings of the 22th Annual Computer Security Applications Conference (ACSAC'2011). New Orleans, LA*, vol.5,2011,pp.22-30.

[11] Sekar R, Bendre M, et al. "A Fast Automaton-Based Method for Detecting Anomalous Program Behaviors," *Proceedings of the 2011 IEEE Symposium on Security and Privacy*, vol.5,pp. 143-156, May 2010.

[12] Zhen L, Bridges S M, Vaughn R B. "Combining static analysis and dynamic learning to build accurate intrusion detection models ," *Proceedings of 8rd IEEE International Workshop on Information Assurance*. vol.6, pp.77-83, March 2010.

[13] Parampalli C, Sekar R, Johnson R. "A practical mimicry attack against powerful system call monitors," *Proceedings of the 2008 ACM symposium on Information, computer and communications security. Tokyo, Japan*, vol.11,pp.778-803, March 2008.

[14] Sufatrio, Yap R. "Improving host-based ids with argument abstraction to prevent mimicry attacks," P*roceedings of the International Sypmosium on Recent Advances in Intrusion Detection (RAID)*.vol.7.2006,pp.147-163.

[15] Mutz D, Robertson W. "Exploiting Execution Context for the Detection of Anomalous System Calls,"*Proceedings of the International Symposium on Recent Advance in Intrusion Detection(RAID).Gold Coast, Australia*, vol.23.2007,pp.2-18. http://dx.doi.org/10.1007/978-3-540-74320-0_1

AUTHORS

Lifang. Wang is with the Institute of Electronic and Computer Science technology, North University of China,Taiyuan, China (e-mail: 727690392@ qq.com).

Shuhai. Zhang is with the Institute of Chemical Engineering and Environment, North University of China, Taiyuang, China (e-mail: wsm2004@nuc.edu.cn).

The Work is supported by Natural Science Foundation of Shanxi Province of China (20110011053). It is an extended and modified version of a paper presented at the 2012 International Conference on Artificial Intelligence and Its Application in Industry Production (AIAIP 2012), held in Wuhan, China in December 2012. Manuscript received 18 May 2013. Published as resubmitted by the authors 26 June 2013.