Research on Anomaly Intrusion Detection Based on Rough Set Attribute Reduction

Liu Cuijuan, Li Yuanyuan
Shijiazhuang University of Economics
Shijiazhuang, China

Qin Yankai
Vocational and Technical training center of Hebei
electric power corporation
Shijiazhuang, China

***Abstract*—With growing amount of network information flowing, the limitations of the traditional network IDSs are more and more outstanding, it couldn't adapt to the trend of increasing novel network attacks and data quantity. Rough Set's reduction theory can effectively avoid redundancy, reduce the extra attributes in the large information. In order to find the novel attacks in IDSs, this paper presents the method of how to measure distance with rough set theory, and the results show that it is efficient to find anomaly.**

Keywords-Rough set, Attribute reduction, Anomaly Intrusion Detection, Classification

I. INTRODUCTION

As network-based computer systems play increasingly vital roles in modern society, information already becomes the key to drive economic and social developments; the problem of information security is gradually outstanding while people share resources highly. In the network environment, it has become the target of intrusion by our enemies and criminals. As the second safety wall after firewall, Intrusion Detection Systems (IDSs) could collect information from network or system and analyze information, so that it can identify who are using a computer system without authorization and who have legitimated access to the system but are abusing their privileges.

Intrusion detection techniques can be categorized in two broad groups: misuse detection and anomaly detection [1]. Misuse detection systems use patterns of well-known attacks to match and identify known intrusions; it is useless to unknown attacks. Anomaly detection systems can be effective against novel or unknown attacks through the detection model, since no a priori knowledge about specific intrusion is required. Network Anomaly Intrusion Detection is an important problem and an active area of research.

The information in network as high as dozen of dimensions even hundreds of dimensions, and rough set as a strong mathematical tool to analyze uncertain knowledge, it can effectively avoid redundancy. IN this paper, through attribute reduction algorithm, the application of RS to the anomaly detection is explained in detail. We have tested our approach using KDD Cup 1999[2] dataset and it significantly reduces the time of process while maintaining the classification result.

II. ANOMALY DETECTION

Anomaly detection focuses on a very small percentage of data objects, which are often ignored or discarded as noise [3].For example, some algorithms in data mining have considered anomaly, but it is the exceptional cases which may interest us. The identification of anomaly can lead to the discovery of truly unexpected knowledge in areas such as electronic commerce, credit card fraud, and even the analysis of performance statistics of professional athletes.

The anomaly is the data deviated from most of the data, this deviation is so large that make people suspect that the data is not out of the random factors, but produced by a completely different mechanism. The key of anomaly detection is how to measure the abnormal data and how to mining the anomaly.

The notion of anomaly studied here is defined as follows: an object O in a dataset S is a DB (p,d), if at least fraction p of the objects in S lies greater than distance d from O.

III. REALATION OF ANOMALY DETECTION AND CLASSIFICATION

Anomaly detection can be thought of as a classification problem, anomaly detection is a process of classification to detect unknown behavior, if unknown behaviors are similar to the known normal behaviors, then the behavior is normal behavior, otherwise it is the intrusion behavior.

The classification accuracy depends directly on the training set of attributes set, classification process is not necessary and cannot cover all the attributes. To reduce the computational cost and immediate response to the invasion, it is necessary to narrow the attributes set without loosing important information. Therefore, determining the appropriate attributes set to classify is very important in terms of the performance of the IDSs.

IV. ROUGH SETS

Rough sets theory has been used successfully as a selection tool to discover data dependencies and reduce the number of attributes contained in a dataset by purely structural method [4].

Rough sets remove superfluous information by examining attribute dependencies. It deals with inconsistencies, uncertainty and incompleteness by imposing an upper and a lower approximation to set membership. Given a dataset with discretized attribute values, by the use of rough sets it is possible to find a subset of the original attributes using rough sets that are the most informative; all other attributes can be removed from the dataset with minimal information loss [5]. It can shorten the training time, and it is a significant analysis method in the intelligence science.

Published by Atlantis Press, Paris, France.

In rough sets, the datasets we process are represented as a table, a two dimension array, where rows stand for records and columns for their attributes.

A. Basic concepts of rough sets

Definition 3.1.1 S=<U, R, V, f> is an information system, where U is a non-empty finite set of objects, and R=C ∪ D is a non-empty finite set of attributes, C is a set of condition attributes, D is a set of decision attributes, where V= $\cup_{r\in R} V_r$ is the set of attribute value, where Vr is the range of value, and f:U×R→V is an information function.

Definition 3.1.2 Let S=<U, R, V, f> be an information system, then a subset of attributes B⊆R defines an equivalence relation (called an indiscernibility relation) on U, given by

IND(B)={(x,y)|(x, y)∈U2: ∀ b∈ B,f(x,b)=f(y,b)}

Definition 3.1.3 Let S=<U, C ∪ D, V, f> be an information system, let X⊆U is a group of objects, and B⊆C is a group of attributes, we can define the B-positive region POSB(D) in the relation IND(D) as:

POSB(D)= ∪{B_ (X) :X∈ IND(D)}

Where B_ (X)={x|(x ∈ U ∧ [x]B ⊆ X)} is called the family of sets bottom approximation sets of the set X as to B. Definition 3.1.4 If every attribute of A is B-indispensable, then A is indispensable with respect to B. The set of all B-indispensable attributes from the set of A is called a B-relative core of A, and it is denoted by $\mathrm{CORE}_B(A)$,

$$\mathrm{CORE}_B(A) = \{ a\in A : \mathrm{POS}_A(B) \neq \mathrm{POS}_{A-\{a\}}(B)\}$$

B. Attributes reduction

Attributes may be irrelevant (having no effect on the performance) or relevant (having an impact on the processing performance). An attribute may have a different discriminatory or predictive power. Some attributes of an information system may be redundant with respect to a specific classification. By virtue of the dependency properties of attributes, one can find a reduced set of the attributes by removing superfluous attributes, without loss of the classification power of the reduced information system.

We can take the following steps to calculate the set of attributes reduction.

```
Input: decision table S=<U, R, V, f>
Output: the set of attributes reduction R
Steps:  For each a∈ C
            To calculate $|POS_{\{a\}}(D)|$;
        Find the attribute a which has the maximum value of $|POS_{\{a\}}(D)|$;
        R={a};
        While $|POS_R(D)| \neq |POS_C(D)|$ Do
            For each a∈ C-R;
                To calculate $|POS_{R\cup\{a\}}(D)|$;
            Find the attribute a which has the maximum value of $|POS_{R\cup\{a\}}(D)|$;
            R=R ∪ {a}; //take attribute a in the rear of set of R
        For each a∈ R
            R=R-{ai };
            If $|POS_R(D)| \neq |POS_C(D)|$
                R=R ∪ {ai };
```

C. The calculation method of distance

Definition Let S=<U, R, C ∪ D, f> be an information system, the distance is defined as:

$$VD(x,y)=\sum_{a\in C} D_a(x,y)$$

$$D_a(x,y)=\sum_{E\in U/IND(D)} f\left|\frac{|[x]_a\cap E|}{|[x]_a|}-\frac{|[y]_a\cap E|}{|[y]_a|}\right|$$

Where x and y are the objects in U, f is absolute value function, for all a∈ C, $[x]_a$ is the equivalence class.

As table Ⅰ show, we can compute any of the distance between two objects. For example, we compute the distance between u1 and u2.

TABLE I. DECISION TABLE

U	a	b	c	e
U1	0	0	1	0
U2	1	1	1	2
U3	2	0	0	1
U4	2	1	0	0
U5	0	2	0	0

$U/IND(D)$ ={E1,E2,E3}={{u1,u4,u5},{u2},{u3}}

VD(u1,u2)= $D_a(u1,u2)+D_b(u1,u2)+D_c(u1,u2)$

$[u1]_{\{a\}}$={u1}, $[u2]_{\{a\}}$={u2}

$$D_a(u1,u2) = f\left|\frac{|\{u1\}\cap E1|}{|\{u1\}|}-\frac{|\{u2\}\cap E1|}{|\{u2\}|}\right| + f\left|\frac{|\{u1\}\cap E2|}{|\{u1\}|}-\frac{|\{u2\}\cap E2|}{|\{u2\}|}\right| + f\left|\frac{|\{u1\}\cap E3|}{|\{u1\}|}-\frac{|\{u2\}\cap E3|}{|\{u2\}|}\right| =0+1+0=1$$

$[u1]_{\{b\}}$={u1,u3}, $[u2]_{\{b\}}$={u2,u4}

$$D_b(u1,u2) = f\left|\frac{|\{u1,u3\}\cap E1|}{|\{u1,u3\}|}-\frac{|\{u2,u4\}\cap E1|}{|\{u2,u4\}|}\right| + f\left|\frac{|\{u1,u3\}\cap E2|}{|\{u1,u3\}|}-\frac{|\{u2,u4\}\cap E2|}{|\{u2,u4\}|}\right| + f\left|\frac{|\{u1,u3\}\cap E3|}{|\{u1,u3\}|}-\frac{|\{u2,u4\}\cap E3|}{|\{u2,u4\}|}\right| =0+\frac{1}{2}+\frac{1}{2}=1$$

In the same way, $D_c(u1,u2)$=0

VD(u1,u2)=1+1+0=2

So the distance between u1 and u2 is 2, we can compute all of the distance between two objects.

V. Experiments

A. Data Sets

KDD Cup 1999 is network data which simulate intrusion in the military network environment, it collected data for 9 weeks, and its training data included 5000 thousand network connection, each connection record was a TCP packet sequence, which finished in the special protocol and stipulate time, these sequences transmit data between a fixed source IP address and destination IP address. It has 41 linear attributes and one categorical attribute, which can be divided into 5 categories: normal, u2r_r2l (u2r: unauthorized access to local super user privileges, r2l: unauthorized access from a remote machine), dos (denial of service), probe (surveillance and other probing).

B. Attribute Reduction

The record in KDD Cup 1999 includes two types of attribute values: continuous and discrete type. Continuous type, e.g., the number of seconds of the connection, number of data bytes from the source to destination, erc., discrete type, e.g., protocol type, network service on the destination, erc.,and there is 32 continuous attributes and 9 discrete attributes.

Rough sets can't deal with continuous attributes directly, so we must discrete the continuous attributes. We adopted Naive Scalar algorithms [6].

We extract randomly 15000 non-repetitive connection records, and store into training database. Applied the attribute reduction algorithm, the original 42 attributes reduce to 15.

In order to verify the ability of classification of data sets after reduction, we have adopted the following test methods:

We use Weka 3 software [7], which is a collection of machine learning algorithms for data mining tasks. We take it as the tool for classification, and select four classifiers: RBF Network, Bayes Net, ID3, and Decision Table. The classification of the training set is used to estimate the rate of detection. The detection rate of each classifier is calculated as percentage of the successful classification records.

In Table Ⅱ, the quality of the reduction set is shown. The rate of detection is the same or slightly higher than before reduction. We find a smaller set of attributes with the close classificatory power as the original set. Our goal is to reduce the number of attributes without loosing the accuracy of detection. And the result confirms that the number of attributes not have too great impact on detection rates. Meanwhile, our methods shorten the process time. It is very important to network intrusion detection which needs real-time response.

TALBE II. DCOMPARING DETECTION RATES (IN %) AND RUNNING TIME(IN SECOND)

Methods	Attribute number	Detection rates (%)	Running time(second)
RBF Network	42	94.25	40.51
	15	95.08	22.42
Bayes Net	42	93.13	0.61
	15	94.87	0.23
ID3	42	100	2.94
	15	99.90	0.98
Decision Table	42	99.32	33.29
	15	99.29	8.87

C. Experimental results and discussion

In this section, we see how the distance and the percentages affect performance. We believe that: the farther the distance, the lower the detection rates; the closer the distance, the higher false alarm rates. So the distance depends on the characteristics of dataset.

Varying the value of p in d =3, we record the detection rates and the false alarm rates. In Table Ⅲ, the rate of detection is slightly higher when p =0.98, but the false alarm rate is also higher. That is because a large number of data considered as anomaly.

TABLE III. STATISTICS FOR VARIABLE p

p	**Detection rates**	**false alarm rates**
0.95	0.912	0.005
0.96	0.945	0.010
0.97	0.988	0.018
0.98	0.997	0.135
0.99	0.991	0.127

VI. Summaries

As a strong mathematical tool to analyze uncertain knowledge, rough sets' reduction theory can effectively avoids redundancy, more importantly, it can't loose any vital information, so it can shorten the processing time. In order to find novel attacks, anomaly detection based on distance is approved, with can detect anomaly effectively. The whole process provides a convenient and rapid method of guarding against the new and diversified attacks.

The experiment results show that rough sets method is suitable for attribute selection. This method is efficient and reduces the amount of data set for handle data. Intrusion detection model is a composition model that needs various theories and techniques. One or two models can hardly offer satisfying results. We plan to apply other theories and techniques to operate in a high accurate and low false alarm rate in our future work.

REFERENCES

[1] Ming Zhu, Ming Ming, and Jun Wang.Network Intrusion Detection Method Based on Outlier Mining, Computer Engineering [J], Vol 29, No. 13, ,2003,PP:125-127

[2] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[3] E.M.Knorr and Raymond T.Ng.Algorithms for Mining Distance-Based Outliers in Large Datesets[C],Proceedings of the 24th VLDB Conference New York,USA,1998

[4] Chouchoulas A. and Shen Q.,"Rough Set-Aided Keyword Reduction for Text Categorization,"Application Artificial Intelligence,vol. 15,no. 9,pp. 843-873,2001

[5] Jensen R. and Shen Q.,"Fuzzy-Rough Data Reduction with Ant Coliny Optimization,"Fuzzy Sets and Systems,vol. 149,pp. 5-20,2005

[6] Zhang Wenxiu,Wu Weizhi.An Introduction an a Survey for Studies of Rough Set Theory,Fuaay Systems and Mathematics[J], 2000, 14(4), PP:1-12

[7] http://www.cs.waikato.ac.nz/~ml/weka/index.html

Published by Atlantis Press, Paris, France.