Information Security based on Soft Computing Techniques

Witaya Siripanwattana[1] and Surat Srinoy[2]

***Abstract*— The Traditional intrusion detection systems (IDS) look for unusual or suspicious activity, such as patterns of network traffic that are likely indicators of unauthorized activity. However, normal operation often produces traffic that matches likely "attack signature", resulting in false alarms. One main drawback is the inability of detecting new attacks which do not have known signatures. In this paper we propose an intrusion detection method that proposes rough set based feature selection heuristics and using fuzzy c-means for clustering data. Rough set has to decrease the amount of data and get rid of redundancy. Fuzzy Clustering methods allow objects to belong to several clusters simultaneously, with different degrees of membership. Our approach allows us to recognize not only known attacks but also to detect suspicious activity that may be the result of a new, unknown attack. The experimental results on Knowledge Discovery and Data Mining-(KDDCup 1999) dataset.**

I. Introduction

As defined in [1], intrusion detection is "the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusions. It is also defined as attempts to compromise the confidentiality, integrity, availability, or to bypass the security mechanisms of a computer or network". Anomaly Intrusion Detection Systems (IDSs) aim at distinguishing an abnormal activity from an ordinary one.

Intrusion detection is a critical component of secure information systems. Many approaches have been proposed which include statistical [2], machine learning [3], data mining [4] and immunological inspired techniques [5]. Identification of suspicious activities before they have an impact; to perform situational assessment and to respond in a more timely and effective manner. Events that may not be actual security violations but those that do not fit in the normal usage profile of a user may be termed as suspicious events. Monitoring suspicious activities may help in finding a possible intrusion. There are two main intrusion detection systems. Anomaly intrusion detection system is based on the profiles of normal behaviors of users or applications and checks whether the system is being used in a different manner [6]. The second one is called misuse intrusion detection system which collects attack signatures, compares a behavior with these attack signatures, and signals intrusion when there is a match. The theory of rough sets has been specially designed to handle data imperfections same as in fuzzy logic. 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. Rough sets estimates the relevance of an attribute by using attribute dependencies regarding a given decision class. It achieves attribute set covering by imposing a discernibility relation. It is often impossible to analyze the vast amount of whole data, but one has to focus the analysis on an important portion of the data such as using some criteria, only the classes of interest can be selected for analysis or processing while the rest is rejected. This paper suggests the use rough set as a dimensionality reduction technique to avoid this information loss.

The rest of this paper is organized as follows. In section 2, we discuss the related works; introduce rough set in section 3; explains fuzzy clustering in section 4; evaluate our intrusion detection model through experiments in section 5; and in section 6 ends the paper with a conclusion and some discussion.

II. Related Works

Most intrusion occurs via network using the network protocols to attack their targets. Twycross [7] proposed a new paradigm in immunology, Danger Theory, to be applied in developing an intrusion detection system. Alves et al. [8] presents a classification-rule discovery algorithm integrating artificial immune systems (AIS) and fuzzy systems. For example, during a certain intrusion, a hacker follows fixed steps to achieve his intention, first sets up a connection between a source IP address to a target IP, and sends data to attack the target [6]. Generally, there are four categories of attacks [9]. They are: 1) DoS (denial-of-service), for example ping-of-death, teardrop, smurf, SYN flood, and the like. 2) R2L : unauthorized access from a

Manuscript received October 9, 2007. This work was supported in part by Suan Dusit Rajabhat University, Bangkok, Thailand.

W. Siripanwattana[1] is with Computer Science Department, Suan Dusit Rajabhat University, Bangkok, 10300 THAILAND (corresponding author to provide phone: 662-244-5600; fax: 662-244-5603; e-mail: wasiripa@ yahoo.com).

S. Srinoy[2]. is with the Computer Science Department, Suan Dusit University, Bangkok, 10300 THAILAND, (e-mail: surat_sri@dusit.ac.th).

remote machine, for example guessing password, 3) U2R : unauthorized access to local super user (root) privileges, for example, various “buffer overflow” attacks, 4) PROBING: surveillance and other probing, for example, port-scan, ping-sweep, etc. Some of the attacks (such as DoS, and PROBING) may use hundreds of network packets or connections, while on the other hand attacks like U2R and R2L typically use only one or a few connections.[10]

A. Rough Set

Rough set theory is a formal methodology that can be employed to reduce the dimensionality of datasets as a preprocessing step to training a learning system on the data.Suppose that a dataset is viewed as a decision table *T* where attributes are columns and objects are rows. Let *U* denote the set of all objects in the dataset and *A* the set of all attributes such that $a : U \to V_a$ for every $a \in A$ where V_a is the value set for attribute *a*. In a decision system, *A* is decomposed into the set *C* of conditional attributes and the set *D* of decisions attributes which are mutually exclusive and $C \cup D = A$. For any $P \subseteq A,$ there is an equivalence relation *I(P)* as follows:

$$I(P) = \{(x,y) \in U^2 \mid \forall a \in Pa(x) = a(y)\}.$$

If $(x,y) \in I(P),$ then *x* and *y* are indiscernible by attributes from *P*. The equivalence classes of the *P-indiscernibility equivalence relation I(P)* are denoted $[x]p.$Given an equivalence relation *I(P)* for $P \subseteq C,$ the lower approximation $\underline{P}X = \{x \in U \mid [x]p \subseteq X\}.$ The *C-positive region* of *D* is the set of all objects from the universe *U* which can be classified with certainty into classes of *U/D* employing attributes from *C*, that is,

$$POS_c(D) = \bigcup_{x \in U/D} \underline{C}X.$$

An attribute $c \in C$ is *dispensable* in a decision table *T* if $POS_{(c-\{c\})}(D) = POS_c(D);$ otherwise attribute *c* is *indispensable* in *T*. A set of attributes $R \subseteq C$ is called a *reduct* of *C* if it is a minimal attributes subset preserving the condition: $POS_R(D) = POS_c(D).$ With regard to computational complexity and memory requirements, however, the calculation of all reducts in an NP-hard task [11]. To solve this problem, we use QUICKREDUCT algorithm [12] shown below for feature selection of classification. The algorithm uses the *degree of dependency* $\gamma p(D)$ as follows:

$$\gamma P(D) = \frac{\| POS_P(D) \|}{\| U \|},$$

For any set *A*, ||*A*|| denotes the cardinality of *A*.

As a criterion for the attribute selection as well as a stop condition. This algorithm does not always generate a *minimal* reduct since $\gamma p(D)$ is not a perfect heuristic. It does result in only one close-to-minimal reduct, though it is useful in greatly reducing dataset dimensionality. The average complexity of QUICKREDUCT algorithm was experimentally determined to be approximately *O(n)* for a dimensionality of *n* though the worst-case runtime complexity is *O(n!).*

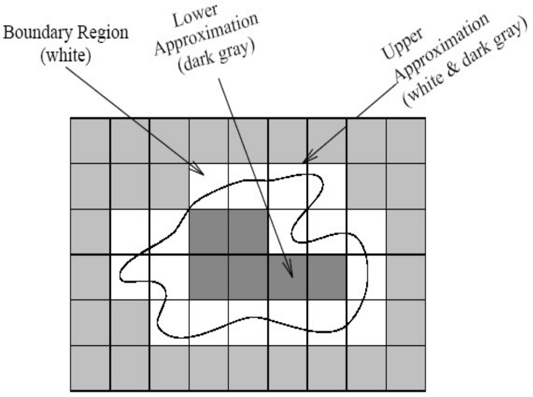
Boundary Region (white)
Lower Approximation (dark gray)
Upper Approximation (white & dark gray)

Fig. 1: Depiction of Rough Sets

```
QuickReduct(C,D,R)
Input: The set C of all conditional attributes
       The set D of decision attributes.
Output: The reduct R of $C(R \subseteq C)$
$R \leftarrow \phi$
do
    $T \leftarrow R$
    $\forall x \in (C - R)$
    if $\gamma R \cup \{x\}(D) > \gamma T(D)$
        $T \leftarrow R \cup \{x\}$
    $R \leftarrow T$
until $\gamma R(D) = \gamma C(D)$
return R
```

Fig. 2: QuickReduct Algorithms

To illustrate the operation of Rough Set Attribute Reduction (RSAR), an example dataset is presented as in Table 1

Table 1: Example Dataset

Instance	Attributes			Decision field
	Service	Count	Srv_count	
1	http	1	4	Yes
2	ftp_data	2	3	Yes
3	Private	1	5	No
4	http	1	1	Yes
5	Domain_u	2	3	No
6	http	0	2	No

Information can be incomplete, inconsistent, uncertain, or all three. We adopted the rough set algorithm for data cleaning as proposed by Sarjon and Mohd. Noor [13]. To use rough sets by the equivalence up to discernibility, this attribute reduction will have to be minimal with respect to content of information.

B. Fuzzy C-mean (FCM) Clustering

Fuzzy *C*-means (FCM) algorithm, also known as fuzzy ISODATA, was introduced by Bezdek [14] as extension to Dunn's [18] algorithm to generate fuzzy sets for every observed feature. The Fuzzy *C*-means clustering algorithm is based on the minimization of an objective function called *C-means functional.*

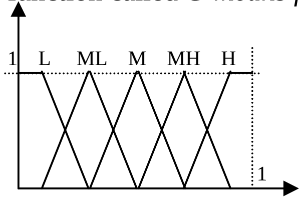

Fig. 3: A Fuzzy Space of Five Membership Function

The fuzzy membership functions corresponding to the informative regions are stored as cases. A collection of fuzzy sets, called fuzzy space, defines the fuzzy linguistic values or fuzzy classes. A sample fuzzy space of five membership function is shown in Fig. 3.

Fuzzy clustering methods allow for uncertainty in the cluster assignments. Rather that partitioning the data into a collection of distinct sets (where each data point is assigned to exactly one set), fuzzy clustering creates a fuzzy pseudo partition, which consists of a collection of fuzzy sets. Fuzzy sets differ from traditional sets in that membership in the set is allowed to be uncertain. A fuzzy set is formalized by the following definitions. Let $X := \{x^1, x^2, ..., x^n\}$ be a set of given data. A *fuzzy subset* of *X* is represented by a mapping $A : X \rightarrow [0,1]$, called the *membership function,* where $A(x^i)$ represents the degree of membership of point x^i . Thus, for example, $A(x^i) = 1$ indicates that x^i is definitely not in the subset, and $A(x^i) = 0.8$ indicates that x^i is likely, but not certainly, in the subset.

A *fuzzy pseudo partition* of X is a family of fuzzy subsets of *X*, denoted by $P = \{A_1, A_2, ..., A_c\}$, which satisfies the equations

$$\sum_{i=1}^{c} A_i(x^k) = 1, \quad \text{for } k = 1, ..., n. \qquad (1)$$

Notice that each point can be assigned a positive degree of membership to several subsets. (1) Assures that the total degree of membership across all subsets is one. This assures that the membership of each point into fuzzy subsets is completely assigned. For example, of $A(x^i) = 0.8$, there is a degree of membership of 0.2 for point x^i that must be assigned to other subsets.

Fuzzy *c*-mean clustering is analogous to the *k*-mean clustering algorithm. The goal is to construct a fuzzy pseudo-partition consisting of *c* fuzzy subsets (or fuzzy clusters), with points assigned (fuzzily) to clusters based on their distance to cluster centers. To make this more concrete, it is necessary to define the concept of a cluster center for a fuzzy cluster. Given a pseudo partition $P = \{A_1, A_2, ..., A_c\}$, the c cluster centers, $v_1, v_2, \ldots, v_c$ associated with the partition are calculated by the formula

$$v_i = \frac{\sum_{k=1}^{n} [A_i(x_k)]^m x_k}{\sum_{k=1}^{n} [A_i(x_k)]^m} \qquad (2)$$

Where $m > 1$ is a real number that governs the influence of membership grades. Observe from (2) that the center v_i is simply a weighted average of all the points in the data set. However, the weight of each point x_k depends on its degree of membership in the fuzzy cluster. The parameter *m* governs how heavily to weigh the degree of membership.

III. Experimental Setup and result

In this experiment, we use a standard dataset the raw data used by the KDD Cup 1999 intrusion detection contest [15]. This database includes a wide variety of intrusions simulated in a military network environment that is a common benchmark for evaluation of intrusion detection techniques. In general, the distribution of attacks is dominated by probes and denial-of-service attacks; the most interesting and dangerous attacks, such as compromises, are grossly under-represented [16]. The data set has 41 attributes for each connection record plus one class label. There are 24 attack types, but we treat all of them as an attack group. A data set of size N is processed. The nominal attributes are converted into linear discrete values (integers). After eliminating labels, the data set is described as a matrix *X*, which has *N* rows and m=41 columns (attributes). There are m_d=8 discrete-value attributes and m_c = 33 continuous-value attributes.

We ran our experiments on a system with a 1.5 GHz Pentium IV processor and 512 MB DDR RAM running Windows XP. All the preprocessing was done using MATLAB®. MATLAB's Fuzzy Logic Toolbox [17] was used for Fuzzy *c*-means clustering, whereas rough set operations were done in ROSETTA [18]. ROSETTA is a software toolkit capable of performing all the operations for data processing and classification. In practice, the number of classes is not always known beforehand. There is no general theoretical solution to finding the optimal number of clusters for any given data set. We choose k = 5 for the study. We will compare five classifiers which have been also used in detecting these four types of attacks

C. Data Preprocessing

A considerable amount of data-preprocessing had to be undertaken before we could do any of our modeling experiments. It was necessary to ensure though, that the reduced dataset was as representative of the original set as possible. The test dataset that previously began with more than 300,000 records was reduced to approximately 18,216 records. Table 2 shows the dataset after balanced among category for attack distribution over modified the normal and other attack categories. Preprocessing consisted of two steps. The first step involved mapping symbolic-valued attributes to numeric-valued attributes and the second step implemented non-zero numerical features. We reduce the dimensionality of this data set (by using rough set) from 41 to 10 attributes are *duration, service, src_bytes, dst_byte, count, srv_count, serror_rate, dst_host_srv_count, dst_host_diff_srv_rate, and dst_host_same_src_port_rate.*

D.Future Selection

Feature selection techniques aim at reducing the number of unnecessary features in classification rules. Rough set theory has been used to define the necessity of features.

Feature selection is an optimization process in which one tries to find the best feature subset, from the fixed set of the original features, according to a given processing goal and a feature selection criterion. A pattern's features, from the point of view of processing goal and type, may be irrelevant (having no effect on processing performance) or relevant (having an impact on processing performance). Features can be redundant (correlated, dependent) [19]. When we process volumes of data, it is necessary to reduce the large number of features to a smaller set of features. There are 42 fields in each data record and it is hard to determine which fields are useful or which fields are trivial. Jin et al [8] suggest correlation coefficients between fields by using SPSS. They propose that if the correlation coefficients of fields *i* and *j*, *R(i,j)*, is larger than 0.8, then there is a strong correlation between fields *i* and *j*, and will select either one of them to represent these two fields. Rough sets allow us to determine (for a discrete attribute data set) a set called a core, containing strongly relevant features, and reducts, containing core plus additional weakly relevant features, such that each reduct is satisfactory to determine concepts in the data set. Based on a set of reducts for a data set some criteria for feature selection can be formed, for example a selecting feature from a reduct containing the minimal set of attributes [19].

E. Performance measure

Standard measures for evaluating IDSs include *detection rate, false alarm rate, trade-off between detection rate and false alarm rate* [20], *performance* (Processing speed + propagation + reaction), and *Fault Tolerance* (resistance to attacks, recovery, and subversion). Detection rate is computed as the ratio between the number of correctly detected attacks and the total number of attacks, while false alarm (false positive) rate is computed as the ratio between the numbers of normal connections that are incorrectly misclassified as attacks [21]. These are good indicators of performance, since they measure what percentage of intrusions the system is able to detect and how many incorrect classifications are made in the process.

Anomaly detection amounts to training models for normal traffic behavior and then classifying as intrusions any network behavior that significantly deviates from the known normal patterns and to construct a set of clusters based on training data to classify test data instances. In fig. 4 is result from our experiment.

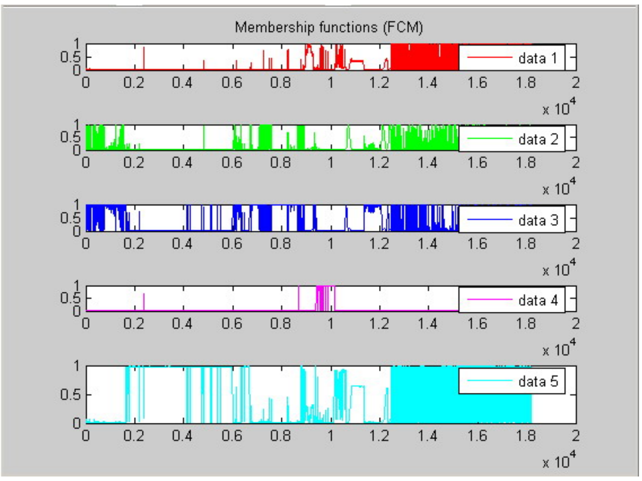

Fig. 4: Membership Functions of Each Cluster

IV. CONCLUSION AND FUTURE WORK

In this paper we apply fuzzy *c-means* methods to intrusion detection to avoid a hard definition between normal class and certain intrusion class and could be considered to be in more than one category (or from another point of view it allows representation of overlapping categories).We introduce the current status of intrusion detection systems (IDS) and rough set based feature selection heuristics , and present some possible data mining based ways for solving problems. Rough set based methods with data reduction for network security are discussed. 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 in intrusion detection in our future work

ACKNOWLEDGMENT

This research was supported in part by grants from Suan Dusit Rajabhat University www.dusit.ac.th at Bangkok, Thailand.

REFERENCES

[1] D.S Bauer, M.E Koblentz, (1988), “NIDX- An Expert System For Real-Time Network Intrusion Detection”, *Proc. of the Computer Networking Symposium*, pp. 98-106.

[2] R. Bace and P. Mell, (2001), “Intrusion Detection Systems”, *NIST Special Publication on Intrusion Detection System*, 31 November 2001.

[3] A.Sundaram, (1996), “An Introduction To Intrusion Detection”, *Crossroads: The ACM student magazine*, Volume 2 (Issue 4).

[4] D. Denning, (1986), “An Intrusion-Detection Model”, *In IEEE computer society symposium on research in security and privacy*, pp. 118-131.

[5] T.Lane, (2000), “Machine Learning Techniques For The Computer Security”, PhD thesis, Purdue University.

[6] W. Lee and S. Stolfo, (1998), “Data Mining Approaches For Intrusion Detection”, *Proc. of the 7th USENIX security symposium*.

[7] D.Dagupta and F. Gonzalez, (2002) “An Immunity-Based Technique To Characterize Intrusions In Computer Networks”, *IEEE Transactions on Evolutionary Computation*, Volume 6, June 2002, pp. 281- 291.

[8] H. Jin, J. Sun, H. Chen, and Z. Han, (2004), “A Fuzzy Data Mining Based Intrusion Detection System”, *Proc. of 10 th International Workshop on future Trends in Distributed Computing Systems (FTDCS04) IEEE Computer Society*, Suzhou, China, May 26-28, pp. 191-197.

[9] J. Twycross , (2004), “Immune Systems”, *Danger Theory and Symposium on Immune System and Cognition*, Leeds, U.K., March.

[10] R.T. Alves, M.R.B.S. Delgado, H.S. Lopes, A.A. Freitas, (2004), “An artificial immune system for fuzzy-rule induction in data mining”, Lecture Notes in Computer Science, Berlin: Springer-Verlag, Volume 3242, pp.1011-1020.

[11] Q. Shen and A. Chouchoulas, (2001), “Rough Set-Based Dimensionality Reduction For Supervised And Unsupervised Learning”, *International Journal of Applied Mathematics and Computer Science*, Volume 11 (Issue 3), pp. 583–601.

[12] J. Katzberg and W. Ziarko, (1996), “Variable Precision Extension Of Rough Sets”, *In W. Ziarko* (ed.) *Fundamenta Informaticae, Special Issue on Rough Sets*, Volume 27 (Issue 2-3), pp. 155–168.

[13] D. Sarjon and Mohd Noor Md Sap, (2002), “Association Rules Using Rough Set and Association Rule Methods”, *Proc.of 7th Pacific Rim International Conference on Artificial Intelligence (PRICAI-02)*,Tokyo, Japan, August 18-22, pp. 238-243.

[14] J. Bezkek, (1999), “Pattern Recognition With Fuzzy Objective Function Algorithms”, Plenum Press, USA, 1981.

[15] KDD Data Set, 1999; http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[16] P. Laskov, K. Rieck, C. Schäfer, K.R. Müller, (2005), “Visualization Of Anomaly Detection Using Prediction Sensitivity”, *Proc.of Sicherheit*, April 2005, pp.197- 208.

[17] Math Works, (2001), “Statistical Toolbox For User’s Guide”, Math Works.

[18] K. Cios, W. Pedrycz, R. Swiniarski, (1998), “Data Mining-Methods For Knowledge Discovery”, Kluwer Academic Publishers, London.

[19] W. Chimphlee, Abdul Hanan Abdullah, Mohd Noor Md Sap and S. Chimphlee, (2005), “Unsupervised Anomaly Detection With Unlabeled Data Using Clustering”, *Proc. of. International Conference on ICT-Mercu Buana ICT2005*, pp.42-49.

[20] A. Lazarevic, A. Ozgur, L. Ertoz, J. Srivastava, and V. Kumar, (2003), “A Comparative Study Of Anomaly Detection Schemes In Network Intrusion Detection”, *In SIAM International Conference on Data Mining*, 2003.

[21] T. Wakaki, H. Itakura, and M.Tamura, (2004), “Rough Set-Aided Feature Selection For Automatic Web-Page Classification”, *Proc. of the IEEE/WIC/ACM International Conference on Web Intelligence (WI’04).*