International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:9, 2007

Network Anomaly Detection using Soft Computing

Surat Srinoy, Werasak Kurutach, Witcha Chimphlee, and Siriporn Chimphlee

***Abstract*—**One main drawback of intrusion detection system is the inability of detecting new attacks which do not have known signatures. In this paper we discuss an intrusion detection method that proposes *independent component analysis* (ICA) based feature selection heuristics and using *rough fuzzy* for clustering data. ICA is to separate these independent components (ICs) from the monitored variables. Rough set has to decrease the amount of data and get rid of redundancy and Fuzzy 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 activity that may be the result of a new, unknown attack. The experimental results on Knowledge Discovery and Data Mining-(KDDCup 1999) dataset.

***Keywords*—**Network security, intrusion detection, rough set, ICA, anomaly detection, independent component analysis, rough fuzzy .

I. Introduction

INTRUSION detection is based on the assumption that intrusion activities are noticeably different from normal system activities and thus detectable. 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. Many approaches have been proposed which include statistical [2], machine learning [3], data mining [4] and immunological inspired techniques [5]. 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. Independent component analysis (ICA) aims at extracting unknown hidden factors/components from multivariate data using only the assumption that the unknown factors are mutually independent. 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 ICA as a dimensionality reduction technique to avoid this information loss. The rest of this paper is organized as follows. In section II, we discuss the related works and independent component analysis; introduce rough set, fuzzy set and rough fuzzy in section III; explains experimental design in section IV; evaluate our intrusion detection model through experiments in section V; and in section VI ends the paper with a conclusion and some discussion.

II. Related Work

A. Feature Reduction

In a classification problem, the number of features can be quite large, many of which can be irrelevant or redundant. Since the amount of audit data that an IDS needs to examine is very large even for a small network, classification by hand is impossible. Feature reduction and feature selection improves classification by searching for the subset of features, which best classifies the training data. Some of the important features an intrusion detection system should possess include refer in Srilatha et al. [20]:

- Be fault tolerant and run continually with minimal human supervision. The IDS must be able to recover from system crashes, either accidental or caused by malicious activity.
- Possess the ability to resist subversion so that an attacker cannot disable or modify the IDS easily. Furthermore, the IDS must be able to detect any modifications forced on the IDS by an attacker.

Manuscript received 1 November, 2005.

Surat Srinoy is with a PhD student under faculty of computer engineering at Mahanakorn University of Technology, Bangkok, Thailand (phone: 662-244-5225; fax: 662-668-7136; e-mail: surat_sri@dusit.ac.th).

Asst. Prof. Dr. Werasak Kurutach is dean of information science and technology at Mahanakorn University of Technology, Bangkok, Thailand (e-mail: werasak@mut.ac.th).

Witcha Chimphlee is with a PhD student under faculty of computer science and information system at University Technology of Malaysia (e-mail:witcha_chi@yahoo.com).

Siriporn Chimphlee is with a PhD student under faculty of computer science and information system at University Technology of Malaysia (e-mail:siriporn_chi@yahoo.com).

- Impose minimal overhead on the system to avoid interfering with the normal operation of the system.
- Be configurable so as to accurately implement the security policies of the systems that are being monitored. The IDS must be adaptable to changes in system and user behavior over time.
- Be easy to deploy: this can be achieved through portability to different architectures and operating systems, through simple installation mechanisms, and by being easy to use by the operator.
- Be general enough to detect different types of attacks and must not recognize any legitimate activity as an attack (false positives). At the same time, the IDS must not fail to recognize any real attacks (false negatives).

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 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]

B. Independent Component Analysis (ICA)

A relevant feature is defined in [5] as one removal of which deteriorates the performance or accuracy of the classifier, and an irrelevant or redundant feature is not relevant. These irrelevant features could deteriorate the performance of a classifier that uses all features since irrelevant information is included inside the totality of the features. Thus the motivation of a feature selector is (i) *simplifying* the classifier by the selected features; (ii) *improving or not significantly reducing* the accuracy of the classifier; and (iii) *reducing* the dimensionality of the data so that a classifier can handle large values of data [6]. Many approaches as feature selectors have been proposed.

Independent component analysis (ICA) for dimension reduction is to separate these independent components (ICs) from the monitored variables. Introduction of ICA concepts in the early 1980s in the context of neural networks and array signal processing. ICA was originally developed to deal with problems that are closely related to the real world 'cocktail-party' problem. ICA is a method for automatically identifying the underlying factors in a given data set. Dimension reduction using ICA is based on the idea that these measured variables are the mixtures of some independent variables. When given such a mixture, ICA identifies those individual signal components of the mixture that are unrelated. Given that the only unrelated signal components within the signal mixture are the voices of different people. ICA is based on the assumption that source signals are not only uncorrelated, but are also 'statistically independent' [7].

ICA techniques provide statistical signal processing tools for optimal linear transformations in multivariate data and these methods are well-suited for feature extraction, noise reduction, density estimation and regression [8]. The ICA problem can be described as follows, each of h mixture signal $x_1(k), x_2(k),...,x_h(k)$ is a linear combination of q independent components $s_1(k),s_2(k),...,s_h(k)$, that is , X = AS where A is a mixing matrix. Now given X, to compute A and S. Based on the following two statistical assumptions, ICA successfully gains the results: 1) the components are mutual independent; 2) each component observes nongaussian distribution. By X = AS, we have $S = A^{-1}X=WX$ (where $W = A^{-1}$). The take is to select an appropriate W which applied on X to maximize the nongaussianity of components. This can be done in an iteration procedure.

Given a set of *n*-dimensional data vectors $[X^{(1)},X^{(2)},\ldots,X^{(N)}]$, the independent components are the directions (vectors) along which the statistics of projections of the data vectors are independent of each other. Formally, if A is a transformation from the given reference frame to the independent component reference from then

$$X = As$$

Such that

$$p(s) = \prod p_a(s_i),$$

where $p_a(.)$ is the marginal distribution and p(s) is the joint distribution over the *n*-dimensional vector *s*.

Usually, the technique for performing independent component analysis is expressed as the technique for deriving one particular W,

$$Y = Wx,$$

Such that each component of y becomes independent of each other. If the individual marginal distributions are non-Gaussian then the derived marginal densities become a scaled permutation of the original density functions if one such *W* can be obtained. One general learning technique [9; 10] for finding one *W* is

$$\Delta W = \eta(I - \phi(y)y^T)W,$$

Where $\phi(y)$ is a nonlinear function of the output vector y (such as a cubic polynomial or a polynomial of odd degree, or a sum of polynomials of odd degrees, or a sigmoidal function) [11].

III. Rough Sets, Fuzzy Set and Rough Fuzzy

A. Rough Sets

Rough sets are characterized by their ability for granular computation. In rough set theory a concept *B* is described by its "lower" $(\underline{B})$ and "upper" $(\overline{B})$ approximations defined with respect to some indiscernibility relation. Rough set

theory [12] provides an effective means for analysis of data by synthesizing or constructing approximations (upper and lower) of set concepts from the acquired data. The key notions here are those of ''information granule'' and ''reducts''. Information granule formalizes the concept of finite precision representation of objects in real life situation, and reducts represent the core of an information system (both in terms of objects and features) in a granular universe [13].

Let $X = \{x_1, ..., x_n\}$ be a set of *U* and *R* an equivalence relation on *X.* As usual, *X/R* denotes the quotient set of equivalence classes, which form a partition in *X,* i.e. *xRy* means the *x* and *y* cannot be took apart. The notion of *rough set* [14] born to answer the question of how a subset *T* of a set *X* in *U* can be represented by means of *X/R.*. It consists of two sets:

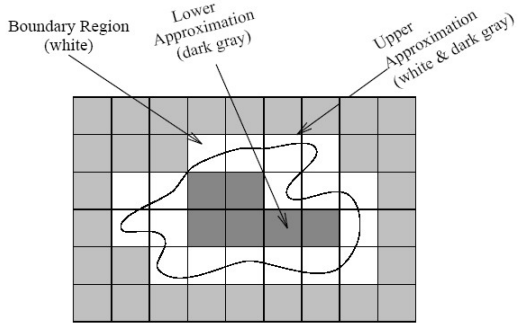

Fig. 1 Rough Representation of a Set with Upper and Lower Approximations

$$RS^*(T) = \{[x]_R \,|\, [x]_R \cap T \neq 0\} \quad (1)$$

$$RS_*(T) = \{[x]_R \,|\, [x]_R \subseteq T\} \quad (2)$$

where $[x]_R$ denotes the class of elements $x, y \in X$ such that *xRy.* $RS^*(T)$ and $RS_*(T)$ are respectively the *upper* and *lower approximation* of T by *R,* i.e.

$$RS_*(T) \subseteq T \subseteq RS^*(T) \quad (3)$$

Other operations over rough sets include:

- Negative region of $X : U - RS^*(X)$.
- Boundary region of $X : RS^*(X) - RS_*(X)$.
- Quality of approximation of *X* by RS_* and

$$RS^* : \mu_R S(X) = \frac{card(RS^*(X))}{card(RS_*(X))}$$

B. Fuzzy Sets

Fuzzy theory provided a mechanism for measuring the degree to which an object belongs to a set by introducing the "membership degree" as a characteristic function $\mu_A(x)$ which associates with each point *x* a real number in the range [0,1]. The nearer the value of $\mu_A(x)$ to unity, the larger the membership degree of *x* in the set *A.*

Let assume *X* be a set, then two different *crisp* versions of a fuzzy set *A* can be define, namely $\overline{A} = \{(x, \mu_{\overline{A}} \,|\, x \in X\}$ and $\underline{A} = \{(x, \mu_{\underline{A}} \,|\, x \in X\}$ where

$$\mu\overline{A}(x) = \begin{cases} 1 & \mu_A(x) \geq 0.5 \\ 0 & \mu_A(x) < 0.5 \end{cases} \quad (4)$$

and

$$\mu_{\underline{A}}(x) = \begin{cases} 1 & \mu_A(x) < 0.5 \\ 0 & \mu_A(x) \geq 0.5 \end{cases} \quad (5)$$

Denote $A \subset X$ and $B \subset X$ two fuzzy sets, i.e. $A = \{(x_i, \mu_A(x_i)), i = 1, ..., n\}$ and $B = \{(x_i, \mu_B(x_i)), i = 1, ..., n\}$, the operations on fuzzy sets are extensions of those used for conventional sets (intersection, union, comparison, etc.). The basic operations are the intersection and union as defined as follows:

The membership degree of the *intersection* $A \cap B$ is

$$\mu_{A \cap B}(x) = \min\{\mu_A(x), \mu_B(x)\} \quad x \in X \quad (6)$$

The membership degree of the *intersection* $A \cup B$ is

$$\mu_{A \cup B}(x) = \max\{\mu_A(x), \mu_B(x)\} \quad x \in X \quad (7)$$

Furthermore, a common measure of similarity between two fuzzy sets *A* and *B* is the l^p-distance, defined as follows [15]. The l^p-distance between two fuzzy sets *A* and *B* is given by

$$l^p(A, B) = (\sum_{i=1}^{n} | \mu_A(xi) - \mu_B(xi) |^p)^{\frac{1}{p}} \quad (8)$$

if p=1 the l^p-distance reduces to the fuzzy Hamming distance.

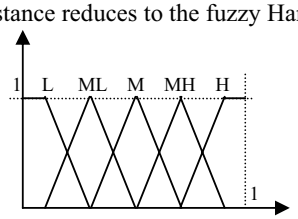

Fig. 2 A fuzzy space of five membership function

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. 2.

C. Rough Fuzzy Sets

In any classification task the aim is to form various classes where each class contains objects that are not noticeably different. These indiscernible or non-distinguishable objects can be viewed as basic building blocks (concepts) used to build up a knowledge base about the real world [16].

In this paper we propose the rough fuzzy sets, realizing a system capable to efficiently cluster data coming from image analysis tasks. The hybrid notion of rough fuzzy sets comes from the combination of two models of uncertainty like vagueness by handling rough sets and fuzzy sets. Rough sets embody the idea of indiscernibility between objects in a set,

while fuzzy sets model the ill-definition of the boundary of a sub-class of this set.

The *rough-fuzzy set* is the generalization of rough set in the sense that here the output class is fuzzy. Let X be a set, R be an equivalence relation defined on X, and the output class $A \subseteq X$ be a fuzzy set. The rough-fuzzy set is a tuple $\langle \underline{R}(A), \overline{R}(A) \rangle$, where the lower approximation $\underline{R}(A)$ and the upper approximation $\overline{R}(A)$ are fuzzy sets of X/R, with membership functions defined in [17, 18] by

$$\mu_{\underline{R}(A)}([x]_R) = \inf\{\mu_A(x) \mid x \in [x]_R\} \quad \forall x \in X \quad (9)$$

and

$$\mu_{\overline{R}(A)}([x]_R) = \sup\{\mu_A(x) \mid x \in [x]_R\} \quad \forall x \in X \quad (10)$$

Here, $\mu_{\underline{R}(A)}(x)$ and $\mu_{\overline{R}(A)}(x)$ are the membership values of $[x]_R$ in $\underline{R}(A)$ and $\overline{R}(A)$, respectively.

The rough-fuzzy membership function of a pattern $x \in X$ for the fuzzy output class $A_c \subseteq X$ is defined as

$$l_{Ac}(x) = \frac{\| F \cap A_c \|}{\| F \|}, \quad (11)$$

Where $F = [x]_R$, and $\| A_c \|$ implies the cardinality of the fuzzy set A_c. Important properties of the rough-fuzzy membership functions that can be exploited in classification task.

TABLE I
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

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

IV. EXPERIMENTAL DESIGN

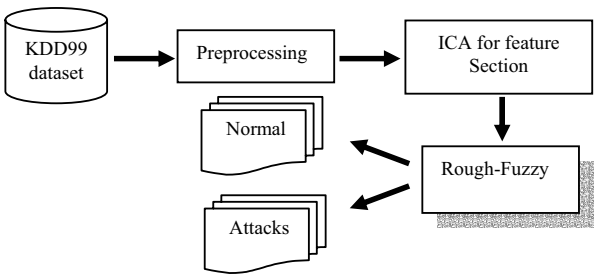

Fig. 3 Step for classification

In our method have three steps (Fig. 3). First step for cleaning and handle missing and incomplete data. Second step for select the best attribute or feature selection using ICA and the last step for classification group of data using rough fuzzy.

V. EXPERIMENTAL SETUP AND RESULTS

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.

A. 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 II shows the dataset after balanced among category for attack distribution over modified the normal and other attack categories refer in [17]. 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.

TABLE II
DATASET FOR ATTACK DISTRIBUTION

Attack Category	% Occurrence	Number of records
normal	31.64	5,763
probe	11.88	2,164
DoS	19.38	3,530
U2R	0.38	70
R2L	36.72	6,689
Summary	100	18,216

B. Features 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) [17]. 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.

C. Performance Measure

Standard measures for evaluating IDSs include *detection rate*, *false alarm rate*, *trade-off between detection rate and false alarm rate* [18], *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 [19]. 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.

VI. Conclusions

Independent component analysis (ICA) aims at extracting unknown hidden factors/components from multivariate data using only the assumption that the unknown factors are mutually independent. The algorithm attempts to maximize the independence among extracted features as well as the mutual information between extracted features and a target variable. We introduce the current status of intrusion detection systems (IDS) and ICA based feature selection heuristics, and present rough fuzzy based ways for solving problems. ICA based methods with data reduction for network security is discussed. A rough fuzzy set comes from the combination of two models of uncertainty like vagueness by handling rough sets and fuzzy sets. Rough sets embody the idea of indiscernibility between objects in a set, while fuzzy sets model the ill-definition of the boundary of a sub-class of this set. These interest methods for handle datasets with an abundance of irrelevant or redundant attributes. 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.

References

[1] D.S Bauer, M.E Koblentz,. NIDX- an expert system for real-time network intrusion detection, *Proceedings of the Computer Networking Symposium*, 1988. pp. 98-106.

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

[3] A.Sundaram, An introduction to intrusion detection, *Crossroads: The ACM student magazine*, 2(4), April 1996.

[4] D. Denning, An intrusion-detection model, *In IEEE computer society symposium on research in security and privacy*, 1986, pp. 118-131.

[5] T.Lane, Machine Learning techniques for the computer Security, PhD thesis, Purdue University, 2000.

[6] W. Lee and S. Stolfo, Data mining approaches for intrusion detection, *Proc. of the 7th USENIX security symposium*, 1998.

[7] D.Dagupta and F. Gonzalez, An immunity-based technique to characterize intrusions in computer networks, *IEEE Transactions on Evolutionary Computation*, 6, June 2002, 28- 291,

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

[9] J. Twycross , Immune Systems, Danger Theory and Intrusion Detection, *presented at the AISB 2004 Symposium on Immune System and Cognition,* Leeds, U.K., March 2004.

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

[11] Q. Shen and A. , Chouchoulas. Rough set-based dimensionality reduction for supervised and unsupervised learning. *International* Journal of APPLIED MATHEMATICS AND COMPUTER SCIENCE, 11 (3), 2001, 583–601,

[12] J. Katzberg and W. Ziarko, Variable precision extension of rough sets, In W. Ziarko (ed.) *Fundamenta Informaticae, Special Issue on Rough Sets*, 27, (2-3), 1996,155–168.

[13] D. Sarjon and Mohd Noor Md Sap, 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, 2002, 238-243.

[14] J. Bezkek, 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, "Visualization of anomaly detection using prediction sensitivity", *Proc.of Sicherheit,* April 2005, 197- 208.

[17] W. Chimphlee, Abdul Hanan Abdullah, Mohd Noor Md Sap, S. Chimphlee, and S. Srinoy, *Unsupervised Clustering methods for Identifying Rare Events in Anomaly Detection*, 6th Internation Enformatika Conference (IEC2005), October 26-28, 2005, Budapest, Hungary.

[18] A. Lazarevic, A. Ozgur, L. Ertoz, J. Srivastava, and V. Kumar, A comparative study of anomaly detection schemes in network intrusion detection. *In SIAM International Conference on Data Mining*, 2003.

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

[20] S. Chebrolu, A. Abraham, J. P. Thomas, Feature deduction and ensemble design of intrusion detection systems, Computer & Security (2004).

Surat Srinoy is with a PhD student under faculty of computer engineering at Mahanakorn University of Technology, Bangkok, Thailand (www.mut.ac.th)He is a lecturer at Suan Dusit Rajabhat University. He complete master degree in 1999 from faculty of computer engineering Chulalongkorn University. He holds vice director of office of academic resources and information technology at Suan Dusit Rajabhat University .His recent research interests include network security, intrusion detection, and immunoinformatics.