Information Sciences Letters

Volume 2
Issue 2 *May 2013*

Article 1

2013

A Real-Time Anomaly Network Intrusion Detection System with High Accuracy

Ahmed A. Elngar
Scince faculty, Al-Azhar University, Cairo, Egypt, elngar_7@yahoo.co.uk

Dowlat A. El A. Mohamed
Math & Computer Science Department, Faculty of Science, Ain-Shams University,
dr_dowlatkma@yahoo.com

Fayed F. M. Ghaleb
Math & Computer Science Department, Faculty of Science, Ain-Shams University, fmghaleb@yahoo.com

Inf. Sci. Lett. **2**, No. 2, 49-56 (2013)

Information Sciences Letters
An International Journal

http://dx.doi.org/10.12785/isl/020201

A Real-Time Anomaly Network Intrusion Detection System with High Accuracy

Ahmed A. Elngar[1,*], *Dowlat A. El A. Mohamed*[2] *and Fayed F. M. Ghaleb*[2]

[1]Computer Science Department, Information & Computer science Faculty, Sinai University, El-Arish ,Egypt
[2]Math & Computer Science Department, science Faculty,Ain-Shams University, Cairo, Egypt

Received: 15 Nov. 2012, Revised: 22 Mar. 2013, Accepted: 23 Mar. 2013
Published online: 1 May. 2013

Abstract: Reliance on Internet and online procedures increased the potential of attacks launched over the Internet. Therefore, network security needs to be concerned to provide secure information channels. Intrusion Detection System (IDS) is a valuable tool for the defense-in-depth of computer networks. However, building an efficient IDS faces a number of challenges. One of the important challenges is dealing with data containing high number of features. This paper is devoted to solve this challenge by proposing an effective PSO-Discritize-HNB intrusion detection system. The proposed PSO-Discritize-HNB IDS combines Particle Swarm Optimization (PSO) and Information Entropy Minimization (IEM) discritize method with the Hidden Naïve Bayes (HNB) classifier. To evaluate the performance of the proposed network IDS several experiments are conducted on the NSL-KDD network intrusion detection dataset. A comparative study of applying Information Gain (IG) which is a well known feature selection algorithm with HNB classifier was accomplished. Also, to validate the proposed PSO-Discritize-HNB network intrusion detection; it is compared with different feature selection methods as Principal Component Analysis (PCA) and Gain Ratio. The results obtained showed the adequacy of the proposed network IDS by reducing the number of features from 41 to 11, which leads to high intrusion detection accuracy (98.2%) and improving the speed to 0.18 sec.

Keywords: Network Security, Intrusion Detection System, Feature Selection, Particle Swarm Optimization, Information Gain, Discritization, Hidden Naïve Bayes

1 Introduction

According to rapid development and popularity of Internet, the potential of network attacks has increased substantially in recent years. Therefore, much attention has been paid to provide secure information channels. However, it is not easy to distinguish the attacks from the normal network actions. To overcome this problem, Anderson in 1980 [1] proposed the concept of Intrusion Detection (ID). ID is a security measure based on the assumption that the behavior of malicious actions is different from a legal user [2]. ID helps to identify a set of actions that compromise the integrity, confidentiality, and availability of information resources [3]. Intrusion Detection System (IDS) becomes an essential key of computer networks security. IDS aim to recognize and notify the unusual access or attack to secure the networks channels, by looking for potential attacks in network traffic and raise an alarm whenever a suspicious activity is detected [4,5].

IDS systems can be divided into two techniques: misuse detection and anomaly detection [6]. Misuse detection can detect the attacks based on well-known vulnerabilities and patterns of intrusions (attacks signatures) stored in a database. It matches the current behavior against the previous knowledge of those known attack patterns [7]. Therefore, this technique may not able to alert the system administrator in case of a new attack. On the other hand, Anomaly detection creates a normal behavior profile and detects the intrusions based on significant deviations from this normal profile [8]. Thus, anomaly detection techniques can detect new types of attack.

Many challenges need to be consider when building an IDS, such as data collection, data preprocessing and classification accuracy. Classification is the prediction of the category labels of instances that are typically

* Corresponding author e-mail: elngar_7@yahoo.co.uk

described by a set of features (attributes) in a dataset. Several classification techniques have been proposed for the development of IDS; including Fuzzy Logic (FL) [9], Neural Networks (NN) [10], Support Vector Machines (SVM) [7,8] and Decision Tree (DT) [11].

Another important problem for constructing an IDS is dealing with data containing large number of features. Data in high dimensional space may lead to decrease the classification accuracy of the IDS. Therefore, feature selection is required as a preprocessing phase for high dimensional data before solving the classification problems. Feature selection aims to reduce the number of irrelevant and redundant features. Different feature selection methods are used to enhance the performance of IDS, including Genetic Algorithm (GA) [9,12], Principal Component Analysis (PCA) [13], Gain Ratio and Information Gain (IG) [14].

This paper proposes an anomaly network intrusion detection system using Particle Swarm Optimization (PSO) feature selection method and Information Entropy Minimization (IEM) discritization with Hidden Naïve Bays (HNB) classifier. The effectiveness of the proposed network IDS is evaluated by conducting several experiments on NSL-KDD network intrusion dataset. The results show that the proposed PSO-Discritize-HNB IDS increases the accuracy and speeds up the detection time. The rest of this paper is organized as follows: Section 2 presents a background of the used methods, including Feature Selection (FS), Information Gain (IG), Particle Swarm Optimization (PSO), Hidden Naïve Bays (HNB). Section 3 describes The NSL-KDD network intrusion dataset. Section 4 introduces the proposed PSO-Discritize-HNB IDS system. Section 5 gives the implementation results and analysis. Finally, Section 6 contains the conclusion remarks.

2 Background

This section gives an overview of Feature Selection (FS), Information Gain (IG), Particle Swarm Optimization (PSO) and Hidden Naïve Bays (HNB).

2.1 Feature Selection (FS)

Feature Selection(FS) is one of the data preprocessing techniques used before classification in IDS [15]. Its purpose is to improve the classification detection accuracy through the removal of irrelevant, noisy and redundant features. FS methods generate a new set of features by selecting only a subset of the original features[16].

There are two main feature selection methods: filter methods [17] and wrapper methods [18]. Filter methods evaluate the relevance of the features depending on the general characteristics of the data, without using any machine learning algorithm to select the new set of features [19]. Frequently used filter methods include Principal Component Analysis (PCA)[13], Gain Ratio and Information Gain (IG)[14]. While, wrapper methods use the classification performance of a machine learning algorithm as the evaluation criterion to select the set of best features [20]. Wrapper methods include PSO algorithm [21] and Genetic Algorithm (GA) [22].

2.2 Information Gain (IG)

The Information Gain (IG) is very easily accessible feature selection method. IG of a given attribute X with respect to the class Y is the reduction in uncertainty about the value of Y, when we know the value of X. It is denoted as $IG(Y \mid X)$.

Let Y and X are discrete variables that take values $Y = \{y_1,...,y_k\}$ and $X = \{x_1,...,x_n\}$. The uncertainty about the value of Y is measured by its entropy, denoted as $H(Y)$. The uncertainty about the value of Y after observing values of X is given by the conditional entropy of Y given X, $H = (Y \mid X)$. Therefore

$$IG(Y \mid X) = H(Y) - H(Y \mid X) \quad (1)$$

Where,

$$H(Y) = -\sum_{i=1}^{k} P(Y = y_i) log_2(P(Y = y_i)) \quad (2)$$

$P(Y = y_i)$ is the prior probabilities for all values of Y.

and

$$H(Y \mid X) = -\sum_{j=1}^{n} P(X = x_j) \sum_{i=1}^{k} P(Y = y_i \mid X = x_j) log_2(P(Y = y_i \mid X = x_j)) \quad (3)$$

where,

$P(Y = y_i \mid X = x_j)$ is the posterior probabilities of Y given the values of X.

According to this measure, an attribute X is regarded more correlated to class Y than attribute Z, if $IG(Y \mid X) > IG(Y \mid Z)$. By calculating information gain, we can select key features based on the correlation rank of each feature to the class [23].

2.3 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) was developed by Kennedy and Eberhart in 1995 [24]. PSO is an evolutionary computation technique which simulates the social behavior of organisms, such as bird flocking. PSO

Inf. Sci. Lett. **2**, No. 2, 49-56 (2013) / www.naturalspublishing.com/Journals.asp

is initialized with a random population (swarm) of individuals (particles), where each particle of the swarm represents a candidate solution in the d-dimensional search space. To find the best solution, each particle changes its searching direction according to: The best previous position of its individual memory (pbest), represented by $P_i = (p_{i1}, p_{i2}, ..., p_{id})$; and the global best position gained by the swarm (gbest) $G_i = (g_{i1}, g_{i2}, ..., g_{id})$ [25].

The d-dimensional position for the particle i at iteration t can be represented as:

$$x_i^t = x_{i1}^t, x_{i2}^t, ..., x_{id}^t \quad (4)$$

While, the velocity (The rate of the position change) for the particle i at iteration t is given by

$$v_i^t = v_{i1}^t, v_{i2}^t, ..., v_{id}^t \quad (5)$$

All of the particles have fitness values, which are evaluated based on a fitness function:

$$Fitness = \alpha.\gamma_R(D) + \beta \frac{|C| + |R|}{|C|} \quad (6)$$

Where, $\gamma_R(D)$ is the classification quality of condition attribute set R relative to decision D and $|R|$ is the length of selected feature subset. $|C|$ is the total number of features. While, the parameters α and β are correspond to the importance of classification quality and subset length, $\alpha = [0,1]$ and $\beta = 1 - \alpha$.

The particle updates its velocity according to:

$$v_{id}^{t+1} = w \times v_{id}^t + c_1 \times r_1(p_{id}^t - x_{id}^t) + c_2 \times r_2(g_{id}^t - x_{id}^t) \quad (7)$$

$$d = 1, 2, ..., D$$

Where, w is the inertia weight and r1 and r2 are random numbers distributed in the range [0, 1]. positive constant c1 and c2 denotes the cognition learning factor (the private thinking of the particle itself) and the social learning factor (the collaboration among the particles). p_{id}^t denotes the best previous position found so far for the i^{th} particle and g_{id}^t denotes the global best position thus far [26].

Each particle then moves to a new potential position based on the following equation:

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (8)$$

$$d = 1, 2, ..., D$$

2.4 Hidden Naïve Bays (HNB)

Hidden Naïve Bays (HNB) was proposed by Jiang et al [27] . HNB is an extended version of the Naïve Bayes (NB) classifier, which relaxes the conditional independence assumption imposed. HNB can avoid the intractable computational complexity for learning and takes the influences from all attributes into account. The HNB classifier creates additional layer that represents a hidden parent of each attribute; this hidden parent combines the influences from all of the other attributes.

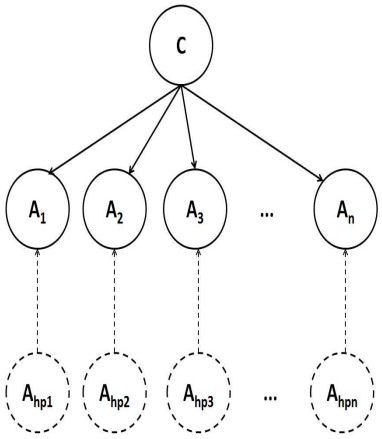

Fig. 1: Hidden Naïve Bays structure

Figure 1 gives the structure of HNB. In this figure C is the class node, and is also the parent of all attribute nodes. Each attribute A_i has a hidden parent A_{hpi} is represented by a dashed circle; where $i = 1, 2, ..., n$. The arc from A_{hpi} to A_i is also represented by a dashed directed line, to differentiate it from regular arcs. Assume an instance E is represented by $E = (a_1, a_2, ..., a_n)$, where a_i is the value of A_i. The joint distribution represented by an HNB is defined as follows:

$$P(A_1, ..., A_n, C) = P(C) \prod_{i=1}^{n} P(A_i \mid A_{hpi}, C) \quad (9)$$

where

$$P(A_i \mid A_{hpi}, C) = \sum_{j=1, j \neq i}^{n} W_{ij} * P(A_i \mid A_j, C) \quad (10)$$

and

$$\sum_{j=1, j \neq i}^{n} W_{ij} = 1 \quad (11)$$

The weights W_{ij}, where $i, j = 1, 2, ..., n$ and $i \neq j$, is compute from the conditional mutual information between two attributes A_i and A_j

$$W_{ij} = \frac{I_p(A_i; A_j \mid C)}{\sum_{j=1, j \neq i}^{n} I_p(A_i; A_j \mid C)} \quad (12)$$

where $I_p(A_i;A_j \mid C)$ is the conditional mutual information given by:

$$I_p(A_i;A_j \mid C) = \sum_{a_i,a_j,c} P(a_i,a_j,c) log \frac{P(a_i,a_j \mid c)}{P(a_i \mid c)P(a_j \mid c)} \quad (13)$$

The classifier corresponding to an HNB is defined as follows:

$$c(E) = arg \max_{c \in C} P(c) \prod_{i=1}^{n} P(a_i \mid a_{hpi}, c) \quad (14)$$

3 Network Intrusion Dataset: NSL-KDD

NSL-KDD dataset [28] is a benchmark used for evaluating network intrusion detection systems. It consists of selected records of the complete KDD'99 dataset [29]. Where, KDD'99 train dataset is five million record of compressed binary TCP dump data from seven weeks of network traffic. Each NSL-KDD connection record contains 41 features (e.g., protocol type, service, and flag) and is labeled as either normal or an attack. The training set contains a total of 22 training attack types, with additional to 17 types of attacks in the testing set. The attacks belong to four categories:

1. **DoS** (Denial of Service) e.g Neptune, Smurf, Pod and Teardrop.
2. **U2R** (user-to-root: unauthorized access to root privileges) e.g Buffer-overflow, Load-module, Perl and Spy
3. **R2L** (remote-to-local: unauthorized access to local from a remote machine)e.g Guess-password, Ftp-write, Imap and Phf
4. **Probe** (probing:information gathering attacks) eg. Port-sweep, IP-sweep, Nmap and Satan.

4 The Proposed Network Intrusion Detection System

The proposed network intrusion detection system is hybrid the Particle Swarm Optimization (PSO) feature selection and (IEM) discritization with Hidden Naïve Bayes (HNB) classifier to detect the network intrusions into five outcomes: normal and four anomaly intrusion types. As shown in Fig 2, the proposed network intrusion detection system consists of four phases.

4.1 Preprocessing Phase

The following three preprocessing stages has been done on NSL-KDD dataset:

1. Convert Symbolic features to numeric value.

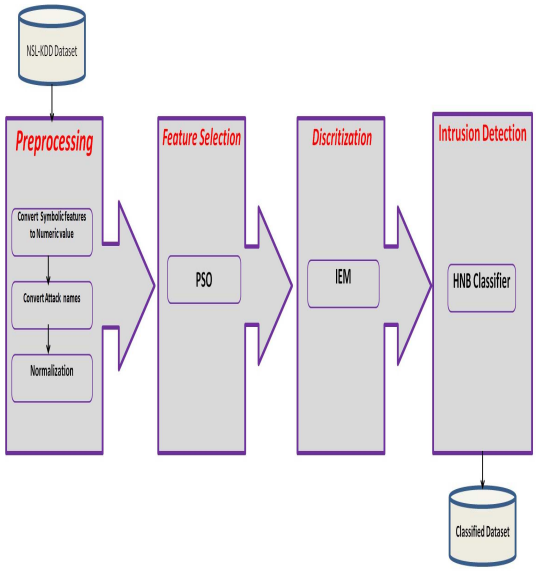

Fig. 2: The structure of the proposed network intrusion detection system

2. Convert Attack names to its category, 0 for *Normal*, 1 for DoS (Denial of service) , 2 for $U2R$ (user-to-root), 3 for $R2L$ (remote-to-local) and 4 for *Probe*.
3. Normaliza the features values, since the data have significantly varying resolution and ranges. The features values are scaled to the range [0, 1], using the following equation:

$$X_n = \frac{X - X_{min}}{(X_{max} - X_{min})} - 1 \quad (15)$$

where, X_{min} , X_{max} are the minimum and maximum value of a specific feature. X_n is the normalized output.

4.2 Particle Swarm Optimization(PSO)Feature Selection Phase

PSO feature selection method efficiently reduced the dimensionality of the NSL-KDD dataset from 41 features to 11 features, which reduces 73.1% of the feature dimension space. At every iteration of the PSO algorithm, each particle X_i is updated by the two best values pbest and gbest. Where, pbest denotes the best solution the particle X_i has achieved so far, and gbest denotes the global best position so far. Algorithm 1 shows the main steps of the PSO algorithm-based feature selection.

4.3 IEM Discritization Phase

Discretization is a process of converting the continuous space of feature into a nominal space [30]. The goal of

Inf. Sci. Lett. **2**, No. 2, 49-56 (2013) / www.naturalspublishing.com/Journals.asp

Algorithm 1 PSO algorithm-based feature selection

Input:
m: the swarm size.
c_1 , c_2 : positive acceleration constants. w: inertia weight.
MaxGen: maximum generation.
MaxFit: fitness threshold.
Output:
Global best position (best features of NSL-KDD dataset)

1: Initialize a population of particles with random positions and velocities on d=1,...,41 NSL-KDD features dimensions $pbest_i$=0, Gbest=0, Iter=0.
2: **while** Iter < MaxGen or gbest < MaxFit **do**
3: **for** i = 1 to number of particles m **do**
4: Fitness(i)=Evaluate(i)
5: **if** fitness(i) > fitness ($pbest_i$) **then**
6: fitness ($pbest_i$)= fitness(i)
7: Update $p_{id} = x_{id}$
8: **end if**
9: **if** fitness(i) > Gbest **then**
10: Gbest=Fitness(i)
11: Update gbest = i
12: **end if**
13: **for** each dimension d **do**
14: Update the velocity vector.
15: Update the particle position.
16: **end for**
17: **end for**
18: Iter= Iter+1
19: **end while**
20: Return the Global best position.

discretization process is to find a set of cut points, these cut points partition the range into a small number of intervals [31].

In this model, the 11 features output from the PSO where discritized by the Information Entropy Minimization (IEM) discretization method. The IEM discretization method was proposed by Fayyad et al. [32], where the cut points should be set between points with different class labels.

Let T partition set S into subsets S_1 and S_2, for k classes $C_1,...,C_k$ the class entropy of a subset S is given by

$$Ent(S) = -\sum_{i=1}^{k} P(C_i,S)log(P(C_i,S)) \qquad (16)$$

where $P(C_i,S)$ is the proportion of examples in S that have class C_i.

For an attribute A, the class information entropy of the partition induced by partition T is defined as

$$E(A,T;S) = \frac{|S_1|}{|S|}Ent(S_1) + \frac{|S_2|}{|S|}Ent(S_2) \qquad (17)$$

4.4 HNB Intrusion Detection Phase

The dataset which has been reducted by PSO method and discritized by IEM method is passed to the HNB classifier to be classified. The algorithm of HNB classifier is described in algorithm 2.

Algorithm 2 Hidden Naïve bayes Algorithm

Input:a set D of training examples
Output:a hidden naive bayes for D

1: **for** each $c \in C$ **do**
2: compute P(c) from training set.
3: **end for**
4: **for** each pair of attributes A_i and A_j **do**
5: **for** each assignment a_i , a_j and c to A_i , A_j and C **do**
6: compute $P(a_i \mid a_j$, c) from training set
7: **end for**
8: **end for**
9: **for** each pair of attributes A_i and A_j **do**
10: compute $I_p(A_i;A_j \mid C)$ and W_{ij} from training set
11: **end for**

5 Implementation Results and Analysis

The proposed hybrid network intrusion detection system is evaluated using the NSL- KDD dataset, where 59586 records are randomly taken. All experiments have been performed using Intel Core i3 2.13 GHz processor with 2 GB of RAM. The experiments have been implemented using Java language environment with a ten-fold cross-validation.

5.1 Performance Measure

The detection effectiveness of the proposed PSO-Discritize-HNB IDS are measured in term of TP Rate, FP Rate and F-measure; which are calculated based on the Confusion Matrix (CM). The CM is a square matrix where columns correspond to the predicted class, while rows correspond to the actual classes. Table 1 gives the confusion matrix, which shows the four possible prediction outcomes [33].

Table 1: Confusion Matrix

	Predicted Class	
Actual Class	Normal	Attake
Normal	TN	FP
Attake	FN	TP

where, **True negatives (TN)**: indicates the number of normal events are successfully labeled as normal.

False positives (FP): refer to the number of normal events being predicted as attacks.

False negatives (FN): The number of attack events are incorrectly predicted as normal.

True positives (TP): The number of attack events are correctly predicted as attack.

$$TPRate = \frac{TP}{TP+FN} \quad (18)$$

$$FPRate = \frac{FP}{FP+TN} \quad (19)$$

$$F-measure = \frac{2*TP}{(2*TP)+FP+FN} \quad (20)$$

5.2 Results and Analysis

–Experiments One: HNB vs. PSO-Discritize-HNB vs. IG-Discritize-HNB

Table 2 shows the accuracy measurements achieved by the HNB classifier using the 41 full dimension features of the NSL-KDD dataset. While, Table 3 gives the accuracy measurements for the proposed anomaly PSO-Discritize-HNB network intrusion detection system with redacted 11 dimension features.

Table 2: HNB accuracy measurements (41-dimension feature)

Class name	TP Rate	FP Rate	F-Measure
Normal	0.965	0.01	0.975
DoS	0.991	0.002	0.993
U2R	0.984	0.005	0.948
R2L	0.973	0.007	0.927
Probe	0.98	0.004	0.974

Table 3: The proposed PSO-Discritize-HNB accuracy measurements (11-dimension feature)

Class name	TP Rate	FP Rate	F-Measure
Normal	0.98	0.014	0.981
DoS	0.993	0.004	0.993
U2R	0.96	0.002	0.964
R2L	0.952	0.004	0.942
Probe	0.973	0.003	0.975

From table 2 and 3, it is clear that the classification accuracy achieved by combining the PSO feature selection and IEM discritization with the HNB classifier is improved than using HNB as a standalone classifier.

To facilitate the comparison, in this paper, we use IG a well known filter based methods of features selection. Table 4 shows the classification accuracy of combinig IG feature selection algorithm and IEM discritization with the HNB classifier.

Table 4: IG-Discritize-HNB accuracy measurements (24-dimension feature)

Class name	TP Rate	FP Rate	F-Measure
Normal	0.967	0.007	0.978
DoS	0.99	0.002	0.993
U2R	0.991	0.006	0.948
R2L	0.983	0.007	0.933
Probe	0.988	0.004	0.981

–Experiments Two: Proposed PSO-Discritize-HNB vs. different feature selection methods

To validate the proposed PSO-Discritize-HNB system; the testing accuracy, feature numbers and timing speed of the proposed system is compared with different feature selection methods. Table 5 shows the comparison results of the proposed PSO-Discritize-HNB system with HNB, IG-Discritize-HNB, PCA-Discritize-HNB and the Gain Ratio-Discritize-HNB network intrusion detection systems. Table 5 illustrate that, the proposed PSO-Discritize-HNB network intrusion detection systems gives the best accuracy performance (98.2%). Also the proposed PSO-Discritize-HNB network intrusion detection systems reduced the feature space to 11, which leads to enhance the timing speed to 0.18 sec which is very important for real time network applications.

Table 5: Testing accuracy, Features Number and Timing comparison

System	Test accuracy	Features no.	Building Time
HNB	97.7%	41	2.47 sec.
IG-Discritize-HNB	97.9%	24	1.09 sec.
PCA-Discritize-HNB	97.1%	25	0.68 sec
Gain Ratio-Discritize-HNB	97.4%	19	0.44 sec
Proposed PSO-Discritize-HNB	98.2%	11	0.18 sec.

6 Conclusions

In this paper we proposed a real-time network intrusion detection system (PSO-Discritize-HNB) with high accuracy.The proposed network intrusion detection system combines PSO feature selection method and IEM discritization with HNB classifier. The NSL-KDD network intrusion benchmark was used for conducting several experiments to test the effectiveness of the proposed network intrusion detection system. Also, a comparative study with applying IG feature selection and IEM discritization with HNB classifier was accomplished.

Inf. Sci. Lett. **2**, No. 2, 49-56 (2013) / www.naturalspublishing.com/Journals.asp

To validate the proposed PSO-Discritize-HNB network intrusion detection; it is compared with different feature selection methods as PCA and gain ratio. The results obtained showed the adequacy of the proposed network IDS by reducing the number of features from 41 to 11, which leads to high detection accuracy (98.2%) and speed up the time to 0.18 sec.

References

[1] J.P. Anderson, "Computer security threat monitoring and surveillance", Technical Report, James P. Anderson Co., Fort Washington, PA, (1980).

[2] W. Stallings, "Cryptography and network security principles and practices", USA, Prentice Hall, (2006).

[3] P. Dokas, L. Ertoz , V. Kumar , A. Lazarevic , J. Srivastava, and P. Tan, "Data mining for network intrusion detection". In Paper presented at the proceedings of the nsf workshop on next generation data mining, Baltimore, (2002).

[4] C. Tsai , Y. Hsu, C. Lin and W. Lin, "Intrusion detection by machine learning: A review", Expert Systems with Applications, **36**, 11994-12000, (2009).

[5] A. A. Elngar, D. A. El A. Mohamed, and F. F. M. Ghaleb, " A Fast Accurate Network Intrusion Detection System", International Journal of Computer Science and Information Security (IJCSIS), **10**, 29-35, (2012).

[6] T. Verwoerd and R. Hunt, "Intrusion detection techniques and approaches", Computer Communications, **25**, 1356-1365, (2002).

[7] S. Mukkamala, G. Janoski and A.Sung, "Intrusion detection: support vector machines and neural networks", In Proc. Of the IEEE International Joint Conference on Neural Networks (ANNIE), St. Louis, MO, 1702-1707, (2002).

[8] S. X. Wu and W. Banzhaf,"The use of computational intelligence in intrusion detection systems: A review", Applied Soft Computing, **10**, 1-35, (2010).

[9] C. Tsang, S. Kwong and H. Wang, "Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection", Pattern Recognition, **40**, 2373-2391, (2007).

[10] G. Wang, J. Hao, J. Ma and L. Huang, "A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering", Expert Systems with Applications, **37**, 6225-6232, (2010).

[11] T. Abbes, A. Bouhoula and M. Rusinowitch, "Protocol analysis in intrusion detection using decision tree", Inform.Technol. Coding Comput. **1**, 404-408, (2004).

[12] K.Y. Chan, C.K. Kwong, Y.C. Tsim, M.E. Aydin and T.C. Fogarty, "A new orthogonal array based crossover, with analysis of gene interactions, for evolutionary algorithms and its application to car door design", Expert Systems with Applications, **37**, 3853-3862, (2010).

[13] R. B. Dubey , M.Hanmandlu and S. K. Gupta, "An Advanced Technique for Volumetric Analysis" International Journal of Computer Applications, **1**, 91-98, (2010).

[14] M. Ben-Bassat, "Pattern recognition and reduction of dimensionality" Handbook of Statistics II, **1**, North-Holland, Amsterdam, (1982).

[15] H. Liu and H. Motoda," Feature Extraction, Construction and Selection: A Data Mining Perspective", Kluwer Academic, second printing, Boston, (2001).

[16] H. F. Eid and A. Hassanien, "Improved Real-Time Discretize Network Intrusion Detection Model", Seventh International Conference on Bio-Inspired Computing: Theories and Application (BIC-TA 2012), December 14-16, Gwalior, India, (2012).

[17] L. Yu and H. Liu," Feature selection for high-dimensional data: a fast correlation-based "lter solution", In Proc. of the Twentieth International Conference on Machine Learning, 856-863, (2003).

[18] Y. Kim, W. Street and F. Menczer "Feature selection for unsupervised learning via evolutionary search", In Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 365-369, (2000).

[19] H. Almuallim and T.G. Dietterich, "Learning Boolean Concepts in the Presence of Many Irrelevant Features", Artificial Intelligence, **69**, 279-305, (1994).

[20] Y. Kim, W. Street and F. Menczer "Feature selection for unsupervised learning via evolutionary search", In Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 365-369, (2000).

[21] L. Chuang, C. Ke and C. Yang, "A Hybrid Both Filter and Wrapper Feature Selection Method for Microarray Classification", In Proc. of the International Multi Conference of Engineers and Computer Scientists (IMECS), Hong Kong, March, **I**, 19-21, (2008).

[22] C. Yang, L. Chuang and C. Hong Yang, "IG-GA: A Hybrid Filter/Wrapper Method for Feature Selection of Microarray Data", Journal of Medical and Biological Engineering, **30**, 23-28, (2009).

[23] W. Wang, S. Gombault and T. Guyet , "Towards fast detecting intrusions: using key attributes of network traffic", The Third International Conference on Internet Monitoring and Protection, IEEE, Bucharest, 86-91, (2008).

[24] R. Eberhart and J. Kennedy, " A new optimizer using particle swarm theory", In Proc. of the Sixth International Symposium on Micro Machine and Human Science, Nagoya, Japan, 39-43, (1995).

[25] G. Venter and J. Sobieszczanski-Sobieski, "Particle Swarm Optimization" (AIAA) Journal, **41**, 1583-1589, (2003).

[26] Y. Liu, G. Wang, H. Chen, and H. Dong, "An improved particle swarm optimization for feature selection", Journal of Bionic Engineering, **8**, 191-200, (2011).

[27] B. Jiang , X. Ding , L. Ma , Y. He , T. Wang and W. Xie, "A Hybrid Feature Selection Algorithm:Combination of Symmetrical Uncertainty and Genetic Algorithms", The Second International Symposium on Optimization and Systems Biology (OSB) 08, China, 152-157, (2008).

[28] M. Tavallaee, E. Bagheri, W. Lu and A. A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set", In Proc. of the 2009 IEEE symposium on computational Intelligence in security and defense application (CISDA), (2009).

[29] KDD'99 dataset, http://kdd.ics.uci.edu/databases, Irvine, CA, USA, July, (2010).

[30] M. Mizianty, L. Kurgan and M. Ogiela, "Discretization as the enabling technique for the Naïve Bayes and semi-Naïve Bayes-based classification", The Knowledge Engineering Review, **25**, 421-449, (2010).

[31] S. Kotsiantis and D. Kanellopoulos, "Discretization Techniques: A recent survey", (GESTS) International Transactions on Computer Science and Engineering, **32**, 47-58, (2006).
[32] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuousvalued attributes for classification learning", In Thirteenth International Joint Conference on Articial Intelligence, 1022-1027, (1993).
[33] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification", JohnWiley & Sons, USA, 2nd edition, (2001).