International Journal of Information and Network Security (IJINS)
Vol.1, No.1, April 2012, pp. 1-9
ISSN: 2089-3299

Synthetic Feature Transformation with RBF Neural Network to Improve the Intrusion Detection System Accuracy and Decrease Computational Costs

[a] **Saeid Asgari Taghanaki,** [b] **Behzad Zamani Dehkordi,** [c] **Ahmad Hatam and** [a] **Behzad Bahraminejad**

[a] Faculty of Electrical and Computer Engineering, Islamic Azad University,
Majlesi Branch, Isfahan, Iran
s.asgari@iaumajlesi.ac.ir, bbahraminejad@ieee.org

[b] Faculty of Computer Engineering, Islamic Azad University,
Shahrekord Branch, Shahrekord, Iran
bzamani@iust.ac.ir

[c] Faculty of Computer Engineering, Hormozgan University,
Hormozgan, Iran
a.hatam@hormozgan.ac.ir

Abstract

With the rapidly growing and wide spread use of computer networks, the number of new attacks and malicious has grown widely. Intrusion Detection System can identify the attacks and protect the systems successfully. However, performance of IDS related to feature extraction and selection phases. In this paper, we proposed new feature transformation to overcome this weakness. For this aim, we combined LDA and PCA as feature transformation and RBF Neural Network as classifier. RBF Neural Net (RBF-NN) has a high speed in classification and low computational costs. Hence, the proposed method can be use in real time systems. Our results on KDDCUP99 shows our proposed method have better performance related to other feature transformation methods such as LDA, PCA, Kernel Discriminant Analysis (KDA) and Local Linear Embedding (LLE).

Keywords*: Intrusion Detection System, Principal Component Analysis, Linear discriminant analysis, RBF Neural Network.*

1. Introduction

The recently growing of local area networks and internet gives a suitable and advance technology for the users. Although the emerging technology is more useful for the users of the computer systems, but the security threads are growing at a high percentage. Network technologies provided the new life and shopping experiences. However, along with network development, there has come a vast increase in network attacks. It not only greatly affects our everyday life, which relies heavily on networks and Internet tools, but also damages computer systems that serve our daily activities including business, learning, entertainment and so on.

Forty million user files of MasterCard and VISA were exposed in 2005 when the company cooperating with Card System solutions was hacked [1], [2]. Many people were forced to renew their credit cards to avoid any financial losses. This event shows the importance of network security. Organizations are using various technologies for system protection and defense such as firewalls, antivirus software's, password protections, etc. to overcome the threads. It is very difficult to provide complete security with these protection techniques. Network accessing and exchanging the data may be easy but providing the security for the information is complex. IDS recognize the unauthorized access to the network, mischievous attacks on the computer systems [3], [4]. To recognize the attacks and detect the intrusions the IDS technology is more useful. The place of IDS in network is presented in Fig. 1.

Intruders categorized into two types: a. external, b. Internal. The unauthorized users who enter the system, make changes to the system, and access the resource in the network without authorization, is an external intruder. The intruder in the network without user accounts trying to attack the system is an internal intruder.

IDS categorize into two types: a. Misuse-detection, b. Anomaly-detection. Intrusion detection with known patterns is called misuse detection. Identifying the abnormalities from the normal network behaviors is called anomaly detection. Hybrid detection systems combine both of these detection systems. Our method work based on Misuse-detection because we work patterns. In the other hand, IDS can classify by the locality of intrusion. The activities with a specific host can be monitored by a host based IDS and monitoring of the network traffic is done by a network based IDS. The host activities such as application logs, system calls, password files, capability/acl databases can be test for intrusion detection by a host based IDS. The network traffic and unique packets for a network tests mischievous traffic based IDS.

Received March 03th, 2012; Revised April 8th , 2012; Accepted April 11th, 2012

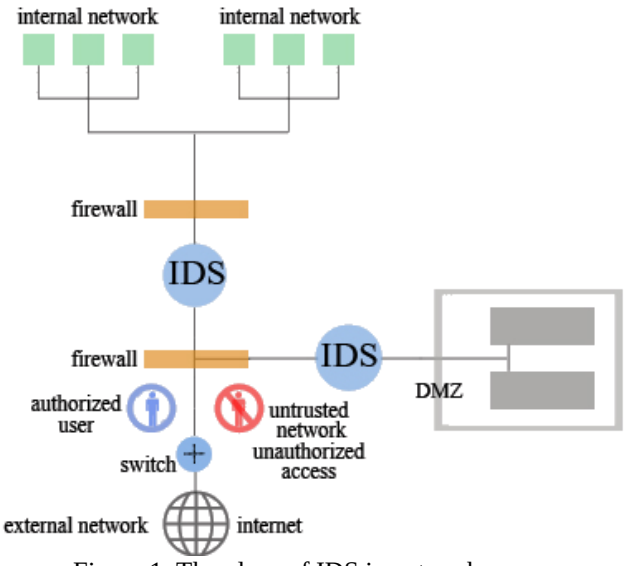

Figure 1. The place of IDS in network

Feature transformation is a process, which creates the new collection of features. In this process, the high-dimensional data has diminished into a meaningful representation of reduced dimensionality. The reduced representation should have a dimensionality that adapt desirably to the innate dimensionality of the data. The innate dimensionality of data is the least number of parameters should be use to determine the observed properties of data. Balancing dimensionality and other undesired properties of high dimensional space [5]. Feature transformation is important in many areas, since it mitigates of dimensionality and other undesired properties of high dimensional spaces [6]. Finally, feature transformation convenience among others, such as classify, visualize and compress of high-dimensional data.

Abdullah et al. used package dump tools such as tcpdump and pcap to collect and evaluate network packets and to recognize network attacks from various network states and packets' distribution [7]. Yu et al. proposed another example of integrating computer forensics with IDS. A knowledge-based system was deployed to collect some features from malicious network behaviors. This system performed reasonably in improving the hit rate of intrusion alerts [8]. Yin et al. presented a method that built a Markov chain to describe users' normal operations. Their method focuses on system calls generated instead of commands submitted [9].

Chau et al. used a pattern extraction technique to identify specific crime data such as segmenting and extracting a suspect from a picture on a security video [10]. Cabrera et al. uses sequential pattern mining to identify attack patterns that hackers often submit, and classified the modus operandi that suspects used in the commission of crimes into predefined crime types [11]. Mohammadi et al. proposed a linear feature transformation method based on class dependent approach for improving the accuracy of intrusion detection systems. In usual class dependent feature transformation methods the mapping process is accomplish using different mapping matrices for different classes of the dataset [12].

In [13], Wei et al proposed an intrusion detection technology, which combines feature extraction with wavelet clustering method. Their intrusion detection model setup has two phases, where the first phase is to project the input data into high dimensional space by using the discriminant vectors extracted by Kernel Fisher Discriminant Analysis. By using KFDA, they can reduce the dimension of the input data and make the dataset more separable. Then the second phase is to set up the detection model based on wavelet clustering. Feng and et al. proposed an incremental kernel principal component analysis algorithm: Data characteristic extraction based on IPCA algorithm (DCEIPCA), which allows efficient processing of large datasets and overcome the insufficient of KPCA. Based on DCEIPCA, they proposed classification expert system for intrusion detection system [14].

Davis et al. reviewed the data preprocessing techniques used by anomaly-based network intrusion detection systems (NIDS), concentrating on which aspects of the network traffic are analyzed, and what feature construction and selection methods have been used [15]. Horng et al. proposed a SVM-based intrusion detection system, which combines a hierarchical clustering algorithm, a simple feature selection procedure, and the SVM technique [16]. In [17], a new method consisting of a combination of discretizers, filters and classifiers was presented which applied to KDD Cup99 dataset. In [18] a framework to optimize data-dependent feature transformations such as PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis) and HLDA (Heteroscedastic LDA) using minimum classification error (MCE) as the main objective are proposed.

Feature transformation is a procedure, which creates the new collection of features. In this process, the non-transformed features have changed into a meaningful representation of transformed features. The transformation should have a dimensionality that adapt desirably to the innate dimensionality of the data. The innate dimensionality of data is the least number of parameters should be use to determine the observed properties of data.

Balancing dimensionality and other undesired properties of high dimensional space .In this process the rudimentary features are transformed to new space, so the accuracy of classifier with the transformed features can increase. Feature transformation is important in many areas, since it mitigates of dimensionality and other undesired properties of high dimensional spaces [19]. We can use from PCA and LDA for transforming the data to new space, also we can use from PCA and LDA and other known methods such as neighborhood component analysis (NCA) as a tools for dimension reduction. So we can reduce the dimension of KDDCUP99 to a useful and lower dimensional dataset for improve the classifying of attacks. Finally, feature transformation convenience among others, such as classify, visualize and compress of high-dimensional data.

In this paper, we propose an intrusion detection system using combination PCA+LDA as feature transformation phase and RBF network as a powerful learner. Our proposed method, combine the PCA and LDA, transform the original features to new mapped space which increasing the ability of detection and decreasing the Time Detection. Therefore, we expect our method have better performance than other linear feature transformation method.

The remaining part of the paper is structured as follows. Section 2 discusses Principal component analysis and linear discriminant analysis as two linear feature transformation methods. After that, in Section 3 our approach has been present. Section 4 presents the results of our experiments on KDDCUP99 dataset. Then, Section 5 concludes the paper, including a discussion on proposed future work.

2. Usage feature transformation methods

There are many possible techniques for transformation of data. Principal component analysis (PCA) and linear discriminant analysis (LDA) [5], [20] are two techniques that commonly used for feature transformation and dimensionality reduction. Linear discriminant analysis method increases the ratio of between-class covariance to the within-class covariance in any specific data set, so, guaranteeing maximal separability.

PCA [21] is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the maximum variance by any projection of the data comes to lie on the primary coordinate (called the first principal component), the second maximum variance on the second coordinate, and so on. The main difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and locality of the original data sets changes when transformed to a different space whereas LDA doesn't change the locality but only attempts to provide more class separability and draw a decision area between the given classes. This method also helps to better understand the allocation of the feature data [21]. In the following of this Section, We explain LDA and PCA, respectively.

.

2.1. Linear Discriminant Analysis

linear discriminant analysis(LDA) is one of the most common supervised linear dimensionality reduction methods, which attempts to find an optimal set of discriminant vectors W=[φ_1,...,φ_d] by enlarge the Fisher principle:

$$J_F(\mathrm{W})=|W^T S_b \mathrm{W}| \,/\, |W^T S_w \mathrm{W}| \quad (1)$$

Here, **S***b* and **S***w* are the between-class scatter matrix and within-class scatter matrix of the training data set, respectively. This can be probable as follows:

$$S_b = \sum_{i=1}^{C} P_i (m_i\text{-m})\,(m_i\text{-m})^T = \sum_{i}^{C-1} \sum_{j=i+1}^{C} P_i P_j (m_i\text{-m})(m_i\text{-m})^T \quad (2)$$

$$S_w = \sum_{i=1}^{C} P_i S_i \quad (3)$$

where *C*, *Pi* ,$\mathbf{m}_i$,**m** and **S***i* represent the whole of classes, a priori probability of class i, the mean vector of class *I*, the mean vector of all training samples and the covariance matrix of class *i*, respectively. Both the original definition and its equivalent pair wise decomposition form can express the between-class scatter matrix Sb [21].

2.2. Principal Component Analysis

Principal Component Analysis (PCA) is a well-established method for dimension reduction. It represents a linear transformation where the data is expressed in a new coordinate basis that corresponds to the maximum variance direction [22], [23]. Suppose that the data set consists of *M* centered observations $X_k \in R^n$, k=1,..., M and $\sum_{k=1}^{m} X_k = 0$, the covariance matrix corresponding to this data set is given by,

$$\mathrm{C}=\frac{1}{M}\sum_{j=1}^{M} X_j X_j^T \quad (4)$$

Diagonal zing C, we obtain the principal components, which are the orthogonal projections onto the eigenvectors are obtained by solving the Eigen value equation:

$$\lambda v = C_V \qquad (5)$$

where $\lambda \geq 0$ and $\mathbf{v} \in R^n \setminus \{0\}^1$.

3. Proposed Method

As we mentioned above, LDA and PCA uses as linear feature transformation methods. These methods have desirable performance in pattern classification system. When data nature is more complex and does not have linearly separability, these methods have poor performance. But each method has their advantages, such as PCA is a good choice to feature reduction application and LDA as supervised method has better performance in classification utility.

In this paper, we combine these methods, which integrated both of these advantages. For this purpose, we use PCA and LDA in form of series. That means we apply LDA transformation after applying PCA. Block diagram of our method shown in figure 2. As you can see, we extract feature from header of each packet. After that, we transformed data to new feature space with PCA and LDA. Then we use transformed feature for training the classifier. In our method, we select the RBF as classifier due to their performance and low computational costs, only training dataset used to compute feature transformation matrix in PCA and LDA methods. The proposed method is simple in use and has powerful performance in detect all of the four type attacks such as (U2R, R2L, DOS and Probe); while the most of recently proposed method have weakness in the delicate attacks such as U2R and probe .Also our method reasonably decreasing the false positive alarms.

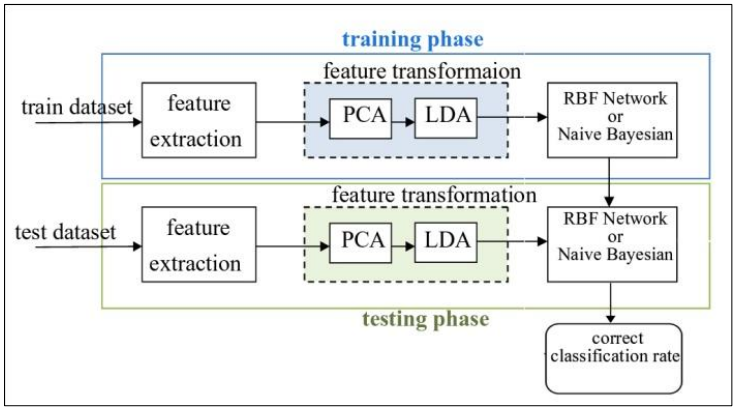

Figure 2. Proposed method block diagram

Two classifiers of RBF network and Naive Bayesian that used in our paper describe as follow of this Section.

3.1. Radial Basis Function Network

RBF network [24], [25] is one of the feed forward neural networks, but has only one middle layer. Typical RBF structural design: Like Back propagation (BP), RBF nets can learn arbitrary mappings: the main difference is in the hidden layer. RBF hidden layer units have a receptive field that has a centre: that is, a specific input value at which they have a maximal output. Their output tails off as the input moves away from this point. Normally, the activation function is a Gaussian: Gaussians with three different standard deviations.

RBF networks are learned by

- deciding on how many middle nodes there should be
- deciding on their sharp nesses (standard deviation (SD)) and centres of their Gaussians
- Training up the output layer.

Generally, the SDs and centres are fixing on first by evaluating the vectors in the training data. The output layer weights are trained by Delta rule. BPNN is the most widely applied neural network. RBFs are gaining in popularity. RBFs have the advantage that one can add extra units with centres near parts of the input which are difficult to classify. Both of BP and RBFs can be used for processing time-varying data [24], [25].

3.1. Naive Bayes classifier

A Naive Bayes classifier is a clean probable classifier based on applying Bayes' theory with strong (naive) independence suppositions. Would be "independent feature model" is more descriptive term for the underlying probability model. In simple terms, a naïve-bayes classifier supposes that the presence (or absence) of a specific feature of a class is not related to the presence (or absence) of any other feature, given the class variable. Even if these features related to each other or upon the existence of the other features, a naïve-bayes classifier considers all of these properties to independently contribute to the probability. Depending on the precise nature of the possibility model, naive Bayes can be trained very professionally in a supervised learning [26], [27].

In many practicable uses, parameter estimation for naïve-bayes models uses the method of maximum likelihood. In other words, one can work with the naïve-bayes model without believing in bayesian probability or using any bayesian methods. In spite of their naive design and apparently over-simplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world applications [26], [27]. An advantage of the naïve-bayes classifier is that it just requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be decided and not the entire covariance matrix [27].

4. Experiment Results

In this section, usage dataset, experiment results and conditions discussed. First, we explain KDDCUP99 dataset used in our experiments. Second, we present our results.

4.1. KDDCUP99 Data Set

To test and to work with the classifiers KDDCUP99 [28] dataset is used. The LAN representing U.S. Air Force LAN was work with the dataset provided by MIT Lincoln Labs, which contains different classes of intrusions present in military networking environment to possess nine weeks of raw TCP/IP data merged with multiple attacks of different types. Every TCP/IP connection with features like duration, protocol type, flag etc., is named as either normal with a particular type of attack such as Smurf, Perl etc., each TCP/IP connection was specially described by different and 41 contiguous. The list of samples for normal class and attack class concluded in 10% of the data set with classification was present in Table2 and number of attacks in training KDDCUP99 dataset w in Table1.

Table1: Number of Attacks in Training KDDCUP99 Dataset

Data Set	Normal	Dos	U2R	R2L	Probe
10%KDD	97277	391458	52	1125	4107
Corrected KDD	60593	229853	70	11347	4106
Whole	972780	3883370	50	1126	41102

Table 2: Attack types and Sample size in 10%KDD Data set

Category	Attack Type(Number of Samples)
Normal	Normal(97277)
DOS	Smurf(280790), Neptune(107201),Back(2203), Teardrop(979), Pod(264), Land(21)
U2R	Buffer_overflow(30), Root kit(10),loadmodule(9), perl(3)
R2L	Warezclient(1020), Guess_passwd(53),Warezmaster(20), Imap(12), ftp_write(8),Multihop(7), Phf(4), Spy(2)
Probe	Satan(1589), Ipsweep(1247),Portsweep(1040), Nmap(231)

The four main classes of attacks are:

- Denial of Service Attacks (DoS-attacks): Where a number of requests were sent by the attacker to the host which he wants to attack.
- User to Root Attacks (U2R-attacks): Getting the right to access from a host by the attacker to obtained the system root (admin) access.
- Remote to User Attacks (R2L-attacks): In this the attacker attempts to access the remote machine through the network and also attempt to control the system operations like a local user.
- Probe (PRB-attacks): The hacker attempt to collect the information and services provided by the machines present in the network to develop the ordinary information.

4.2. Classification results

We use LDA, PCA, Kernel Discriminant Analysis (KDA), Local Linear Embedding (LLE) and LDA+PCA as our method for feature transformation. In addition, we use RBF network and Naïve Bayes as classifiers. We use weka [29] for implementation of classifiers. Figures (3)-(6) shown correct classification rate for two cases of experiments, 66% and 10 folds. As you can see, our method has better performance than LDA, PCA, KDA, and LLE. We have about 8 percent improvement for Naive Bayes classifier in proposed method.

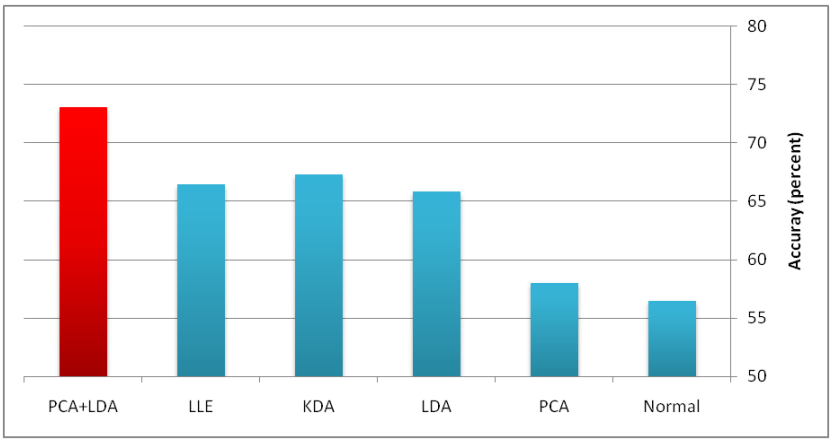

Figure 3. Results of correct classification rate in 66% split case (Naïve Bayes)

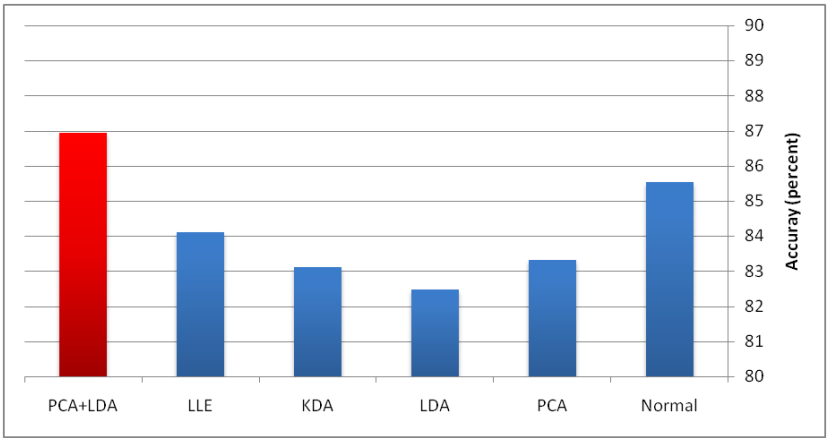

Figure 4. Results of correct classification rate in 66% split case (RBF Network)

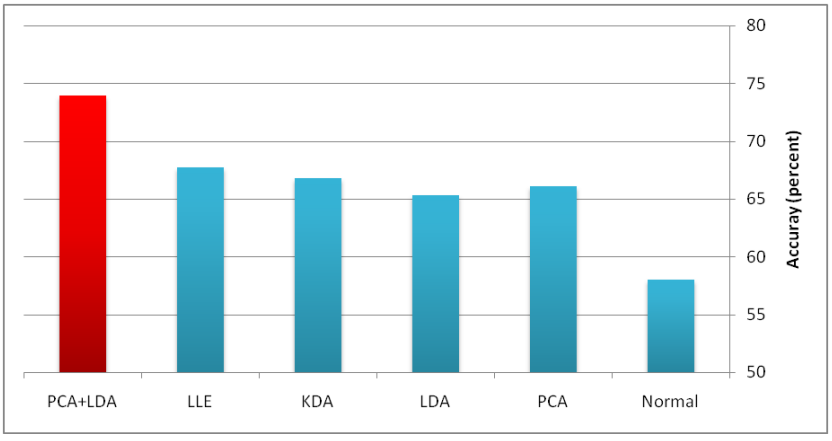

Figure 5. Results of correct classification rate in 10 folds case (Naïve Bayes)

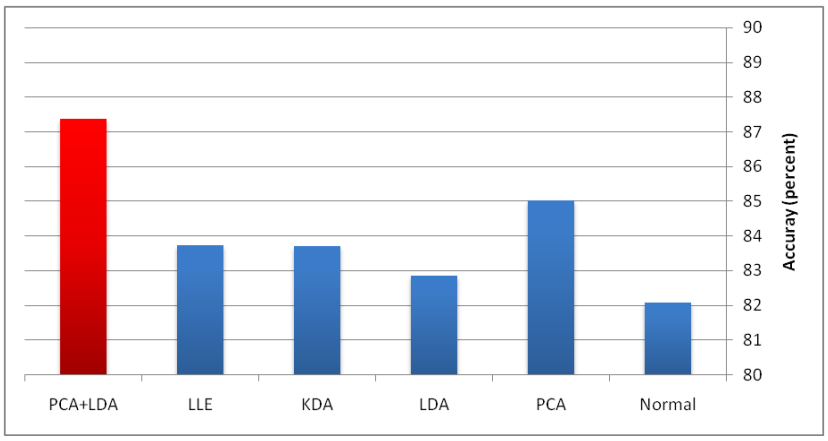

Figure 6. Results of correct classification rate in 10 folds case (RBF Network)

5. Conclusion

In this paper, we proposed new feature transformation based on joint of PAC and LDA. We integrated advantage of these methods. Some when the attacks are very dangerous, so the IDS must act in real-time and identify the attack, so in our method we combined our feature transformation with RBF neural network to produce the best classification accuracy and fast action. For evaluate of proposed method, we done any experiments on KDDCUP99 dataset. In addition, we compare our result with other classifier as naïve Bayes classifier. Results show proposed method has better performance those other transformation techniques such as PCA, LDA, KDA, and LLE. In the future we plan to desig an IDS with higher detection rate with selection the significat features with other method such as Genetic Algorithm and PSO or use from hybrid neural networks. In this paper we worked with supervised classification methods in furure we want to focus on unsupervised learning to desigen an IDS that can detect attacks without training, so we have to remove calculus and time of training.

6. References

[1] B. Schneier, "Card Systems Exposes 40 Million Identities," http://www.schneier.com/blog/archives/2005/06/cardsystems_exp.html

[2] A. P. Mitchell, "40million," http://www.theinternetpatrol.com/cardsystems-compromises data-of-40-million-mastercard-and-visa-cardholders

[3] J. S. Balasubramaniyan, et al., "An architecture for intrusion detection using autonomous agents," in *Proceedings of 14th annual computer security applications conference*, 1998.

[4] L. T. Heberlein, et al., "Towards Detecting Intrusions in a Networked Environment," in *Proceedings of 14th department of energy computer security group conference*, 1991.

[5] K. Fukunaga, "Introduction to Statistical Pattern Recognition", *Academic Press*, San Diego, California, 1990.

[6] L. O. Jimenez, et al., "Supervised classification in high-dimensional space: geometrical, Statistical and asymptotical properties of multivariate data," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 28, Issue 1, pp. 39–54, 1997.

[7] K. Abdullah, et al., "Visualizing Network Data for Intrusion Detection," in *IEEE Workshop on Information Assurance Workshop*, 2005.

[8] J. Q. Yu, et al., "TRINETR :An Intrusion Detection Alert Management System," *SIPLab, Concurrent Engineering Research Center Lane Department of Computer Science and Electrical Engineering*, 2004, pp.235-240.

[9] Q. Yin, et al., "A New Intrusion Detection Method Based on a Behavioral Model," in *World Congress on Intelligent Control and Automation*, 2004, pp. 4370-4374.

[10] M. Chau, et al., "Extracting Meaningful Entities from Police Narrative Reports," in *National Conference on Digital Government Research*, 2002, pp. 271-275.

[11] J. B. D. Cabrera, et al., "Detection and Classification of Intrusion and Faults Using Sentences of System Calls," *SIGMOD Record*, Vol.30, Issue .4, pp.25-34, 2001.

[12] M. Mohammadi, et al, "Class dependent feature transformation for intrusion detection systems," in *19th Iranian Conference on Electrical Engineering (ICEE)*, 2011, pp 1-6.

[13] Y. X. Wei, et al., "KFDA-wavelet cluster based intrusion detection technology," *International Conference on Wavelet Analysis and Pattern Recognition, 2007, ICWAPR '07*, 1899-1903.

[14] Zh. X. Feng, et al., "Expert System Based Intrusion Detection System," *International Conference on Information Management, Innovation Management and Industrial Engineering (ICIII)*, 2010, pp. 404-407.

[15] Davis, et al., "Data preprocessing for anomaly based network intrusion detection: A review," *Computers & Security*, Vol. 30, Issues 6–7, pp. 353–375, 2011.

[16] S. J. Horng, "A novel intrusion detection system based on hierarchical clustering and support vector machines," *Expert Systems with Applications*, Vol. 38, Issue 1, pp. 306–313, 2011.
[17] V. Bolón-Canedo, "Feature selection and classification in multiple class datasets: An application to KDD Cup 99 dataset," *Expert Systems with Applications*, Vol. 38, Issue 5, pp. 5947–5957, 2011.
[18] B. Zamani, et al., "Optimized Discriminative Transformations for Speech Features Based on Minimum Classification Error Direct Link," *Pattern Recognition Letters*, Vol. 32, pp. 948–955, 2011.
[19] K. Fukunaga. "Introduction to Statistical Pattern Recognition", *Academic Press Professional, Inc., San Diego, CA, USA*, 1990
[20] S. Axler, "Linear Algebra Done Right," *Springer-Verlag New York Inc.*, New York, 1995.
[21] M. Loog, et al., "Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion," *IEEE Transaction Pattern Analysis and Machine Intelligence*, Vol. 26, Issue. 6, pp. 732–739, 2004.
[22] B. Scholkopf, et al., "Nonliner component analysis as a kernel eigenvalue problem," *Neural Computation*, Vol. 10, pp. 1299–1319, 1998.
[23] J. Joliffe, "Principal Component Analysis," *Springer*, Berlin, 1986.
[24] L. Fausett, "Fundamentals of Neural Networks," *Prentice-Hall*, 1994.
[25] K. Gurney, "An Introduction to Neural Networks," *UCL Press*, 1997.
[26] H. Zhang, "The Optimality of Naive Bayes," in *proceeding of FLAIRS conference*, 2004.
[27] R. Caruana, et al., "An empirical comparison of supervised learning algorithms," in *Proceedings of the 23rd international conference on Machine learning*, 2006.
[28] KDD cup1999 data, available on http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
[29] http://www.cs.waikato.ac.nz/~ml/weka/

Bibliography of authors

Saeid Asgari Taghanaki received his B.Sc. in 2010 from Iran Islamic Azad University of Shahrekord (IAUSHK) in Software Engineering and now he is a student in last semester of MS in Isfahan, Iran Islamic Azad University of Majlesi in mechatronics trends in Human Machine Interface (HMI). His areas of interests include network security, pattern recognition, image processing, data mining and Soft computing.

Behzad Zamani received his B.Sc. in 2003from Isfahan Uuniversity of technology (IUT) in hardware engineering, MSc. in 2006 from Computer Engineering Department, Iran University of Science & Technology (IUST) and PhD in Artificial Intelligence and working as a Research Assistant in the ASPL and RCIT Labs of Computer Engineering Department, Iran University of Science & Technology (IUST) in 2012. He has published many papers in speech recognition system and related topics. His interest topics include feature transformation and kernel based learning methods.

Ahmad Hatam received his B.Sc. in 1985 from Isfahan, Iran University of technology (IUT) in Electronics Engineering (Electronics), M.Sc. in 1991 from Sharif University of Technology (SUT) in Electrical Engineering (Electronics) and PhD in 2010 from IUT in Electrical Engineering (Communications). He works as a lecturer and inspector in Electrical department at Hormozgan University. In addition, he has worked with industry on several projects in the areas of communications and electronics. His areas of interests are image processing, channel coding, information theory and wireless communications.

Behzad Bahraminejad received his B.Sc. in 2000 from Iran, Shahrood University of technology in Electronics Engineering, M.Sc. in 2004 from K.N.T University of Technology, Tehran, Iran in Bio-Electric Engineering and PhD in 2011 from University Putra Malaysia (UPM), Malaysia in Biomedical Engineering. He focuses his research on biomedical sensors and instrumentation. In recent years, he has been active in research projects in electronic olfaction. His areas of interests are Sensors and Instrumentation, pattern recognition and Soft computing.