Asian Journal of Information Technology 12 (2): 70-76, 2013
ISSN: 1682-3915

Enhancing Intrusion Detection Using Layered Approach with PCA

[1]B. Ben Sujitha and [2]V. Kavitha
[1]Ponjesly College of Engineering,
[2]University College of Engineering, Nagercoil, Tamilnadu, India

Abstract: An Intrusion Detection System is now an inevitable part of any computer network. There is a high increase in the type of attacks including new and earlier unseen attacks. An Intrusion Detection System must reliably detect malicious activities in a network and must perform efficiently to cope with the large amount of network traffic. In this system, PCA algorithm is used for feature selection. PCA is employed to reduce the high dimensional data vectors and thus, detection is handled in a low dimensional space with high efficiency and low use of system resources. Then, the Layered approach is introduced for identifying the attack which has less computational overheads. Thus, the new system addresses the two issues namely accuracy and efficiency. Thus, the system is robust and is able to handle noisy data without compromising the performance. The advantage of the system is that has reduced the computation time and false alarm rate.

Key words: Intrusion detection, PCA, neural network, KDD, agent

INTRODUCTION

Intrusion detection as defined by the SysAdmin, Audit, Networking and Security (SANS) Institute is the art of detecting in appropriate, inaccurate or anomalous activity. Today, intrusion detection is one of the high priority challenging tasks for network administrators and security professionals. More sophisticated security tools mean that the attackers come up with newer and more advanced penetration methods to defeat the installed security systems. Thus, there is a need to safeguard the networks from known vulnerabilities and at the same time take steps to detect new and unseen but possible, system abuses by developing more reliable and efficient Intrusion Detection Systems (IDS). Any Intrusion Detection System has some inherent requirements. Its prime purpose is to detect as many attacks as possible with minimum number of false alarms, i.e., the system must be accurate in detecting attacks. However, an accurate system that cannot handle large amount of network traffic and is slow in decision making will not fulfill the purpose of an Intrusion Detection System. A system that is desired that it detects most of the attacks gives very few false alarms, copes with large amount of data and is fast enough to make real-time decisions. Intrusion detection started in the year 1980s after the influential study from Anderson (2010). Intrusion Detection Systems are classified as network based, host based or application based depending on their mode of deployment and data used for analysis.

Additionally, Intrusion Detection Systems can also be classified as signature based or anomaly based depending upon the attack detection method. A knowledge-based (Signature-based) Intrusion Detection Systems references a database of earlier attack signatures and known system vulnerabilities. The meaning of word signature is that the recorded evidence of an intrusion or attack. Each intrusion leaves a footprint behind (e.g., nature of data packets, failed attempt to run an application, failed logins, file and folder access, etc.). These footprints are called Signatures and can be used to identify and prevent the same attacks in the future. Based on these signatures Knowledge-based (Signature-based) IDS identify intrusion attempts. The disadvantages of Signature-Based Intrusion Detection Systems are signature database must be continually updated and maintained and Signature-based Intrusion Detection Systems may fail to identify unique attacks. Behavior-based (Anomaly-based) Intrusion Detection Systems references a baseline or learned pattern of normal system activity to identify active intrusion attempts. Deviations from this baseline or pattern cause an alarm to be triggered. Higher false alarms are often related with Behavior-Based Intrusion Detection Systems. There is another approach used for detecting both the normal and the known anomalous patterns for training a system and then performing classification on the best data. Such a system is known as the Hybrid System. Hybrid Systems can be very efficient, subject to the classification method used and can also be used to label unseen or new

Corresponding Author: B. Ben Sujitha, Ponjesly College of Engineering, Nagercoil, Tamilnadu, India

instances as they assign one of the known classes to every test instance. It is observed that the proposed system performs significantly better than other systems. Also, the robustness of the proposed method is identified by introducing the noise in the system.

LITERATURE REVIEW

Large amount of research has been done in the area of intrusion detection and a number of techniques including Data Mining approaches such as clustering, Naive Bayesian classifiers, Bayesian Networks, Hidden Markov Models, Decision Trees, Artificial Neural Networks, Support Vector Machines, Genetic Algorithm, agent based approaches and many others have been described in order to Detect Intrusion. The detection techniques are described here particularly with regards to the data they analyze before they label any event as intrusion. Data Mining based approaches for intrusion detection are based on building classifiers based on discovering relevant patterns of program and user behavior. Association rules (Agrawal *et al*., 1993) and frequent episodes are used to learn the record patterns that describe user behavior (Lee and Stolfo, 1998; Lee *et al*., 1998). Data Mining approaches can deal with symbolic data and the features can be defined in the form of packet and connection details. The mining of features is limited to entry level of the packet and requires the number of records to be large and sparsely populated otherwise they tend to produce very large number of rules which increases the complexity.

Clustering of data has been applied extensively for intrusion dsetection using various clustering methods including kmeans, fuzzy c-means and many others (Portnoy *et al*., 2001; Shah *et al*., 2003). However, one of the main drawbacks of clustering techniques is that it is based on calculating the distance between the observations and hence the attributes of the observations must be numeric. Symbolic attributes cannot be used for clustering which results in inaccuracy. Naive bayes classifiers are also proposed by Amor *et al*. (2004) however, they make very strict independence assumption between the attributes (Valdes and Skinner, 2000). The Bayesian network (Kruegel *et al*., 2003) is used to remove the threshold and combine the results of individual models to reach a final result. However, they tend to be attack specific and build a decision network based on special features of each attack. Thus, the size of the bayesian network increases rapidly as the number of features considered increases and the type of attacks modeled increases. Hidden Markov Models have also been used in intrusion detection (Warrender *et al*., 1999; Du *et al*., 2004; Wang *et al*., 2004). The use of Hidden Markov Models for modeling the normal sequence of system calls (Forrest *et al*., 1996) of a privileged process which can then be used to detect anomalous traces of sequence calls. However, modeling the system calls alone may not always provide accurate classification as in such cases various connection level features are ignored. Further, Hidden Markov Models are generative models and fail to model long range dependency between the observations. Decision Trees (Amor *et al*., 2004) have also been used for intrusion detection. The problem with the Decision Trees is to select the best attribute for each decision node during the construction of the tree. One such criterion is to use the gain ratio as in C4.5. The Decision Trees suffer from similar problems as the Bayesian Networks. The Decision Trees tend to grow in size and complexity as the number of attributes increases. Decision Trees can be easily used for building the misuse detection systems but it is very difficult to construct anomaly detection system using Decision Trees. Debar discusses the use of Artificial Neural Networks for Network Intrusion Detection. Though the Neural Networks can work effectively with noisy data but they require large amount of data during training and it is often hard to select the best possible neural network architecture. Support Vector Machines (SVM) (Mukkamala *et al*., 2002) which maps real valued input feature vector to higher dimensional feature space through non-linear mapping have been used for detecting intrusions. The SVM's provide real time detection capability and can deal with large dimensionality of data. However, they are used effectively for binary class classification only. Along with these other techniques for detecting intrusion includes the use of genetic algorithms and agent based approach including autonomous agents for intrusion detection and probabilistic agent based approach for intrusion detection which are generally aimed at a distributed Intrusion Detection System.

The 1999 KDD Intrusion Detection data set which is a version of the 1998 DARPA intrusion detection data set prepared and managed by the MIT Lincoln lab and the system call data set collected at the University of New Mexico have been widely used to report various experimental results on Intrusion Detection. The DARPA data set presents data as a collection of records where each record presents a summary of a connection or sequence of packets between a specific source and target IP address at certain well defined times while the system call data is the traces of system calls generated by certain selected routines such as send mail where each trace is just a sequence of system call and its corresponding process id (Andrew and Srinivas, 2003).

All the above mentioned techniques for detecting intrusions are primarily targeted at ensuring availability.

There are methods by Zhong *et al.* (2005), Hu and Panda (2004) and Lee *et al.* (2002) which are meant to ensure confidentiality and integrity of the data stored in databases. They use the database logs either to build the normal user profiles (Hu and Panda, 2004) or to extract signatures for detecting known attacks as discussed by Lee *et al.* (2002).

However, to ensure that a network is secure there is a need to provide confidentiality and integrity along with availability. The framework suggested in this study aims at providing all the three (confidentiality, integrity and availability) together in a single system.

FEATURE SELECTION AND REDUCTION

Feature selection is one of the key topics in IDS, it improves classification performance by searching for the subset of features which best classifies the training data (Andrew and Srinivas, 2003). In problem of high dimensional feature space, some of the features may be redundant or irrelevant. Removing these redundant or irrelevant features is very important; hence they may deteriorate the performance of classifiers. Feature Selection involves finding a subset of features to improve prediction accuracy or decrease the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features (Koller and Sahami, 1996). This is very important if real-time detection is desired. Principal Component Analysis (PCA) is an essential technique in data compression and feature selection (Oja, 1992) which has been applied to the field of ID (Kuchimanchi *et al.*, 2004; Shyu *et al.*, 2003). PCA (Cureton and D'Agostino, 1983) is an efficient method to reduce dimensionality by providing a linear map of n-dimensional feature space to a reduced m-dimensional feature space. In this study, Principal Component Analysis is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data.

PCA produces a set of principal components which are ortho-normal eigen value/eigenvector pairs. In mathematical terms the principal components of the distribution of the connection records is found or the eigenvectors of the covariance matrix of the set of the connection records (Jolliffe, 2002). These eigenvectors can be thought of as a set of features which together characterize the variation between records connections. Each connection record can be presented exactly in terms of linear combination of the eigen connections. Each connection can also be approximated using only the best eigen connections those that have the largest eigen values and which therefore account for the most variance within the set of connection records. The best n eigen connections span an n dimensional sub connection connection space of all possible connection records. This new space is generated by an information theory method called Principal Component Analysis (PCA) (Jolliffe, 2002).

Calculation of eigen connection: Suppose $x_1, x_2, \ldots, x_M$ are Nx_1 are the training set of connection vectors. The average profile $\overline{x}$ of this set is defined by:

$$\overline{X}=\frac{1}{M}\sum_{i=1}^{M}X_i \tag{1}$$

Each connection record vector X_i differs from the average X_i by:

$$\phi_i = X_i - \overline{X} \tag{2}$$

The eigen connections are the eigenvectors of the covariance matrix C where:

$$C_{(n\times n)}=\frac{1}{M}\sum_{N=1}^{M}\phi_n\phi_i^T = AA^T \tag{3}$$

$$A_{(n\times m)}=\frac{1}{\sqrt{M}}[\phi_1\ \phi_2 \ldots \phi_M] \tag{4}$$

Let U_k be the kth eigen vector of C , λ_k the associated eigen values then:

$$CU_k = \lambda_k U_a \tag{5}$$

Such that:

$$U_\kappa^T U_1 = \begin{cases} 1, \text{if } k=1 \\ 0 \text{ if } k\neq 1 \end{cases}$$

To reduce the dimensionality of the vectors, corresponding to the K largest eigen values:

$$\left(X-\overline{X}\right)=\frac{1}{M}\sum_{i=1}^{K}b_iU_i \tag{6}$$

The value of K is chosen to reduce the dimensionality is based on the following criterion:

$$\text{If } \frac{\sum_{i=1}^{K}\lambda_i}{\sum_{i=1}^{N}\lambda_i} > \text{Threshold}$$

This method has proven to be an exceedingly popular technique for dimensionality reduction and is discussed

at length in most texts on multivariate analysis. The set of n different measures are collected in a vector called connected record vector representing the corresponding connection.

LAYERED APPROACH

The goal of the layered model as shown in Fig. 1 is to reduce computation and the overall time required to detect anomalous events.

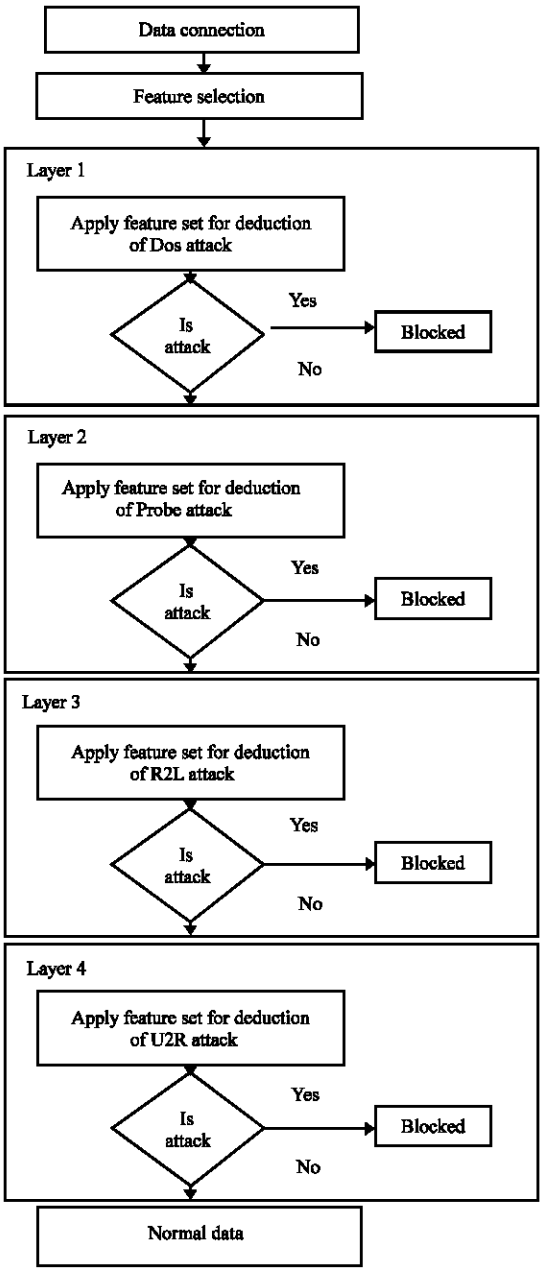

Fig. 1: The working of proposed layered model

The time required to detect an intrusive event is significant and can be reduced by eliminating the communication overhead among different layers. This can be achieved by making the layers autonomous and self-sufficient to block an attack without the need of a central decision-maker. Every layer in the LIDS framework is trained separately and then deployed sequentially. The four layers are defined which correspond to the four attack groups mentioned in the data set. They are Probe layer, DoS layer, R2L layer and U2R layer. Each layer is then separately trained with a small set of relevant features. Feature selection is significant for Layered approach and discussed in the earlier study. In order to make the layers independent some features may be present in more than one layer. The layers essentially act as filters that block any anomalous connection thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion. The effect of such a sequence of layers is that the anomalous events are identified and blocked as soon as they are detected. The second goal is to improve the speed of operation of the system. The speed of the running of the system has been improved since, the features are so much reduced with the help of component analysis.

This results in significant performance improvement during both the training and the testing of the system. In many situations there is a trade-off between efficiency and accuracy of the system and there can be various avenues to improve system performance. The certainly increases system efficiency but it severely affects the accuracy. To balance this trade-off, the PCA is more accurate, though expensive. The Layered approach implemented improves overall system performance. The performance of the proposed system, Layered PCA is comparable to that of the Layered CRF, Kmeans and C4.5 and has higher attack detection accuracy.

INTEGRATING LAYERED APPROACH WITH PCA

In Introduction, researchers discussed two main requirements for an Intrusion Detection System; accuracy of earlier the PCA can be effective in improving the attack detection accuracy by reducing the number of false alarms while the Layered Approach can be implemented to improve the overall system efficiency. Hence, a natural choice is to integrate them to build a single system that is accurate in detecting attacks and efficient in operation. Given the reduced set of feature as the input to the layered approach, researchers first select four layers corresponding to the four attack groups (DoS, Probe, R2L, and U2R). Each attack type falls exactly into one of the following four categories:

Probing: Surveillance and other probing e.g., port scanning

DOS: Denial-of-service, e.g., syn flooding

U2R: Unauthorized access to local super user (root) privileges, e.g., various buffer overflow attacks

R2L: Unauthorized access from a remote machine, e.g., password guessing

This proposed system identify the attack category once the system detects an event as anomalous. Layered approach has the additional capability to identify the type of attack once it detected because every layer is trained to detect only a particular category of attack.

EXPERIMENTAL RESULTS

Dataset: In this experiment, the benchmark KDD'99 intrusion data set. This data set is a version of the original 1998 DARPA intrusion detection evaluation program which is prepared and managed by the MIT Lincoln Laboratory. The data set contains about five million connection records as the training data and about two million connection records as the test data. In the experiments, researchers use 10% of the total training data and 10% of the test data (with corrected labels) which are provided separately. This leads to 494,020 training and 311,029 test instances.

Evaluation: Each record in the data set represents a connection between two IP addresses starting and ending at some well-defined times with a well-defined protocol. Further, every record is represented by 41 different features. Each record represents a separate connection and is hence considered to be independent of any other record. The input data is selected and reduced and given as input to the first layer DOS layer. The reason for choosing the Dos layer first is that majority of the attack is DoS attack. In the first layer itself most of the attack can be blocked. Thus, the computational time for the succeeding layers is less. The training data is either labeled as normal or as one of the 24 different kinds of attack. These 24 attacks can be grouped into four classes; Probing, DoS, R2L and U2R. Similarly, the test data is also labeled as either normal or as one of the attacks belonging to the four attack groups. It is important that the test data is not from the same probability distribution as the training data and it includes specific attack types not present in the training data. This makes the intrusion detection task more realistic.

For the results, researchers give the Precision, Recall and F-value and not the accuracy alone as with the given data set it is easy to achieve very high accuracy by carefully selecting the sample size. From Table 1, researchers note that the number of instances for the U2R, Probes and R2L attacks is very low. Hence, if we use accuracy as a measure for testing the performance of the system, the system can be biased and can attain accuracy. However, Precision, Recall and F-value are not dependent on the size of the training and the test samples. They are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{7}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{8}$$

$$\text{F-value} = \frac{(1+\beta^2)\times \text{Recall} \times \text{Precision}}{\beta^2 \times (\text{Recall} + \text{Precision})} \tag{9}$$

where, TP, FP and FN are the number of True Positives, False Positives and False Negatives, respectively and corresponds to the relative importance of precision versus recall and is usually set to 1.

Table 1 gives the number of instances for each group of attack in the data set and shows the detection accuracy. In this experiment, hybrid detection, i.e., both the normal and the anomalous connections is used for training the model. When all the 41 features are considered, the time taken to test all the 250,436 attacks was 51 sec.

Table 2 gives the number of instances for each group of attack in the data set and shows the time taken to detect the attack in each layer. In this experiment, hybrid detection, i.e., both the normal and the anomalous connections is used for training the model. When all the 41 features are considered, the time taken to test all the 250,436 attacks was 46 sec.

Comparison of results: In this study, researchers compare the research with other well-known methods

Table 1: Accuracy to Detect Attack in each layer

	Detection accuracy	
Layers	Total (%)	Cumulative (%)
DoS	91.817	91.817
Probe	3.586	95.403
R2L	1.566	96.969
U2R	1.004	97.973

Table 2: Time taken to detect attack in each layer

	Testing time	
Layers	Total (sec)	Cumulative (sec)
DoS	27	27
Probe	12	39
R2L	5	43
U2R	3	46

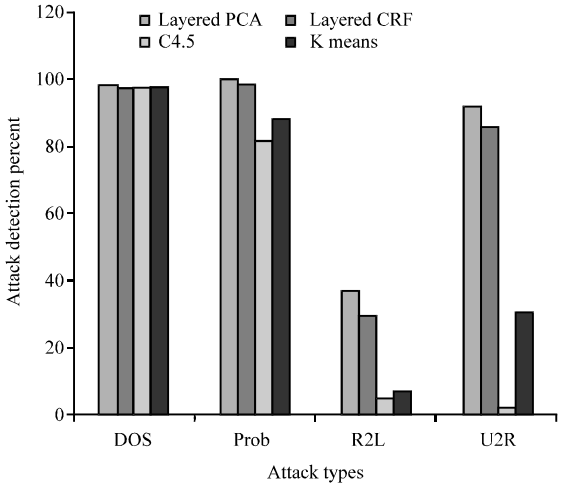

Fig. 2: Performance comparison for detection of attack

Table 3: Performance comparison for detection of attack

Parameters	DoS	Probe	R2L	U2R
Layered PCA				
PD	98.50	98.80	36.701	91.302
FAR	0.05	0.86	0.321	0.040
Layered CRF				
PD	97.40	98.60	29.600	86.300
FAR	0.07	0.91	0.350	0.050
C4.5				
PD	97.00	80.80	4.600	1.800
FAR	0.30	0.70	0.005	0.002
K means				
PD	97.30	87.60	6.400	29.800
FAR	0.40	2.60	0.100	0.400

based on the anomaly intrusion detection principle. The anomaly-based systems primarily detect deviations from the learnt normal data by using Statistical Methods, Machine Learning or Data Mining approaches. Standard techniques such as the Decision Trees and Naive Bayes are known to perform well. Table 3 represents the Probability of Detection (PD) and False Alarm Rate (FAR) in percent for various methods including the KDD'99 cup winners. From the Table 3, it is observed that the Layered PCA perform significantly better than the previously reported results including the winner of the KDD'99 cup and various other methods applied to this data set. The most impressive part of the Layered PCA is the margin of improvement as compared with other methods. Layered PCA have very high attack detection of 98.50% for DoS and 98.80% detection for Probe. They outperform by a significant percentage for the R2L and the U2R attacks. Table 3 shows that the proposed system performs well with the unknown attacks due to the training algorithm. Such that the false alarm rate is decreased better than the earlier models. Figure 2 shows the improvement in the attack detection. The graph is plotted against the different attacks and percent of attack detection (PD).

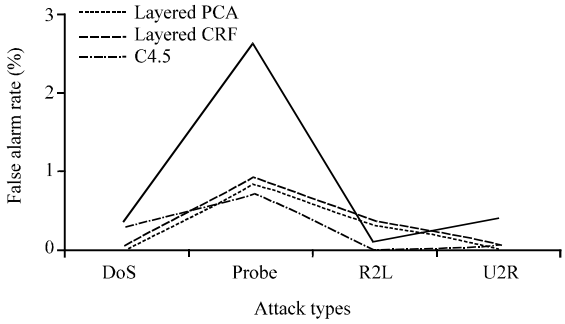

Fig. 3: Performance comparison of reduction of false alarm rate

The best IDS System should produce less false alarm rate. The proposed has the goal and achieved better than the earlier approaches. Figure 3 shows the improvement in the reduction of False Alarm Rate (FAR). Figure 3 is drawn for showing the false alarm rate comparison. The graph is plotted against the different attacks and percent of reduction of False Alarm Rate (FAR).

CONCLUSION

Current Intrusion Detection Systems (IDS) examines all data features to detect intrusion or misuse patterns. Some of the features may be redundant or contribute little (if anything) to the detection process. The purpose of this research is to identify important input features in building IDS that is computationally efficient and effective.

The experimental results show that the proposed model gives better and robust representation of data as it was able to reduce features resulting in a 97% data reduction and approximately 94% time reduction in training with almost same accuracy achieved in detecting new attacks. Meantime, it significantly reduce a number of computer resources both memory and CPU time, required to detect an attack. This shows that the proposed system is reliable in intrusion detection. This system will be improved with the computation by modifying the layered as parallel.

REFERENCES

Agrawal, R., T. Imielinski and A. Swami, 1993. Mining association rules between sets of items in large databases. Proceedings of the ACM SIGMOD International Conference on Management of Data, May 25-28, 1993, Washington, DC., USA., pp: 207-216.

Amor, N.B., S. Benferhat and Z. Elouedi, 2004. Naive bayes vs decision trees in intrusion detection systems. Proceedings of the ACM Symposium on Applied Computing, March 14-17, 2004, Nicosia, Cyprus, pp: 420-424.

Anderson, J.P., 2010. Computer Security Threat Monitoring and Surveillance. James P. Anderson Co., Washington.

Andrew, H.S. and M. Srinivas, 2003. Identifying important features for intrusion detection using support vector machines and neural networks. Proceedings of the 2003 Symposium on Applications and Internet, January 27-31, 2003, IEEE Xplore, London, pp: 209-216.

Cureton, E.E. and R.B. D'Agostino, 1983. Factor Analysis: An Applied Approach. Vol. 1, Routledge, London, ISBN: 9780805815467, Pages: 457.

Du, Y., H. Wang and Y. Pang, 2004. A hidden markov models-based anomaly intrusion detection method. Proceedings of the 5th World Congress on Intelligent Control and Automation, June 15-19, 2004, China, pp: 4348-4351.

Forrest, S., S.A. Hofmeyr, A. Somayaji and T.A. Longstaff, 1996. A sense of self for Unix processes. Proceedings of the IEEE Symposium on Securiry and Privacy, May 6-8, Oakland, CA. USA., pp: 120-128.

Hu, Y. and B. Panda, 2004. A data mining approach for database intrusion detection. Proceedings of the ACM Symposium on Applied Computing, March 14-17, 2004, Nicosia, Cyprus, pp: 711-716.

Jolliffe, I.T., 2002. Principal Component Analysis. 3rd Edn., Springer, New York.

Kruegel, C., D. Mutz, W. Robertson and F. Valeur, 2003. Bayesian event classification for intrusion detection. Proceedings of the 19th Annual Computer Security Applications Conference, December 8-12, 2003, Barbara, CA., pp: 14-23.

Kuchimanchi, G.K., V.V. Phoha, K.S. Balagami and S.R. Gaddam, 2004. Dimension reduction using feature extraction methods for Real-time misuse detection systems. Proceedings from the 5th Annual Information Assurance Workshop, June 10-11, 2004, Ruston, LA., pp: 195-202.

Lee, S.Y., W.L. Low and P.Y. Wong, 2002. Learning fingerprints for a database intrusion detection system. Proceedings of the 7th European Symposium on Research in Computer Security Zurich, October 14-16, 2002, Switzerland, pp: 264-279.

Lee, W. and S. Stolfo, 1998. Data mining approaches for intrusion detection. Proceeding of the 7th USENIX sec. Symposium, San Antonio, Texas, January 26-29, 1998, USENIX Association, Berkeley, CA, USA., pp: 1-16.

Lee, W., S. Stolfo and K. Mok, 1998. Mining audit data to build intrusion detection models. Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining, August 27-31, 1998, New York, USA., pp: 66-72.

Mukkamala, S., G. Janoski and A. Sung, 2002. Intrusion detection using neural networks and support vector machines. Proceedings of IEEE International Joint Conference on Neural Network, May 12-17, Honolulu, HI, USA., pp: 1702-1707.

Oja, E., 1992. Principal components, minor components and linear neural networks. Neural Networks, 5: 927-935.

Portnoy, L., E. Eskin and S. Stolfo, 2001. Intrusion detection with unlabeled data using clustering. Proceedings of ACM CSS Workshop on Data Mining Applied to Security, November 5-8, 2001, Philadelphia, PA., pp: 5-8.

Shah, H., J. Undercoffer and A. Joshi, 2003. Fuzzy clustering for intrusion detection. Proceedings of the 12th IEEE International Conference on Fuzzy Systems, Volume 2, May 25-28, 2003, Baltimore, MD., pp: 1274-1278.

Shyu, M.L., S.C. Chen, K. Sarinnapakorn and L. Chang, 2003. A novel anomaly detection scheme based on principal component classifier. Proceedings of the 3rd IEEE International Conference on Data Mining, November 19-22, 2003, Melbourne, Florida, USA., pp: 172-179.

Valdes, A. and K. Skinner, 2000. Adaptive, model-based monitoring for cyber attack detection. Proceedings of the 3rd International Workshop on RAID 2000, October 2-4, 2000, Toulouse, France, pp: 80-92.

Wang, W., X.H. Guan and X.L. Zhang, 2004. Modeling program behaviors by hidden markov models for intrusion detection. Proceedings of the International Conference on Machine Learning and Cybernetics, Volume 5, August 26-29, 2004, China, pp: 2830-2835.

Warrender, C., S. Forrest and B.A. Pearlmutter, 1999. Detecting intrusions using system calls: Alternative data models. Proceedings of the Symposium on Security and Privacy, May 9-12, 1999, Oakland, CA., USA., pp: 133-145.

Zhong, Y., Z. Zhu and X.L. Qin, 2005. A clustering method based on data queries and its application in database intrusion detection. Proceedings of the 4th International Conference on Machine Learning and Cybernetics, Volume 4, August 18-21, 2005, Guangzhou, China, pp: 2096-2101.