Enhancing performance of anomaly based intrusion detection systems through dimensionality reduction using principal component analysis

Basant Subba , Santosh Biswas, Sushanta Karmakar
Department of Computer Science & Engineering
Indian Institute of Technology, Guwahati
Assam, India 781039
Email: s.basant@iitg.ernet.in, santosh_biswas@iitg.ernet.in , sushantak@iitg.ernet.in

***Abstract*—Anomaly based Intrusion Detection Systems (IDSs) are capable of detecting wide range of network attacks. However, they are characterized by high computational overhead due to large number of redundant or highly correlated features in the input data being analyzed by them. In this paper, we propose a model to minimize the computational overhead of anomaly based IDSs through dimensionality reduction technique called Principal Component Analysis (PCA). PCA reduces the high dimensional data using the dependencies between the input features to represent it in a more tractable, lower dimensional form, without losing any significant information contained in the original data. Experimental results on the benchmark NSL-KDD dataset shows that applying PCA can significantly reduce the dimensionality of the data being processed by anomaly based IDSs and thereby minimize their computational overhead without adversely affecting their performances.**

Keywords*—*Principal Component Analysis (PCA), Intrusion Detection System (IDS), Naive Bayes, C4.5, Support Vector Machine (SVM), NSL-KDD dataset.

I. Introduction

Easy availability and widespread use of automated attack tools has made network intrusions a common phenomenon. Today, a plethora of exploits plague computer networks ranging from a stealthy crafted zero-day exploits to computer worms and viruses capable of inflicting mass level infections and damages. All these factors has made network security an indispensable aspect for any given organization. Several preventive mechanisms like encryption, authentication, policy management, firewall etc. have been proposed in the literature as defensive measures against network intrusions. While these preventive mechanisms significantly strengthens the network security, they cannot generally rule out the possibility of network attacks. Attackers are increasingly finding their way around these preventive mechanisms to gain unauthorized access to the network and launch attacks. Moreover, these preventive mechanisms are ineffective against the insider attacks, where the intruder is a legitimate part of the network and possesses the secure cryptographic keys. Therefore, in addition to these preventive measures, a complementary second line of defense in the form of Intrusion Detection System (IDS) is required to provide a comprehensive security against various network attacks [1] [2] [3].

Based on their detection technique, IDSs can broadly be classified into following two categories: misuse based and anomaly based. Misuse based detection approaches [4] [5] use a set of predefined attack signatures to detect network intrusions. They provide an effective defense against known attacks but fail to detect novel and unknown attacks. Moreover, there might be a significant time lapse between the discovery of a new attack and deployment of its corresponding signature. This delay is often too large and there have been numerous instances of network attacks, notably zero-day exploits and computer worms [6] where misuse based detection approaches have been found to be ineffective. In addition, the developed signatures need to be managed, distributed and kept up-to-date by the security administrator. All these factors put a significant constraint on the effectiveness of the misuse based detection approaches.

On the other hand, anomaly based detection approaches [7] [8] initially models the normal network behavior during the training phase and then deploys the learned model to monitor the network traffic for sign of intrusions. Any deviation of the network traffic from the learned model is considered as an anomaly and an alarm is raised whenever any such anomalous network traffic is detected. The main advantage of anomaly based detection approaches over the misuse based detection approaches is that they do not require any prior attack signatures to detect intrusions and therefore they are able to detect novel and unknown attacks. However, anomaly based detection approaches are characterized by high computational overhead which makes them unsuitable for real time intrusion detection in high speed networks.

In this paper, we address the issue of high computational overhead of anomaly based IDSs through application of dimensionality reduction technique called Principal Component Analysis (PCA). PCA is a statistical technique that allows reduction of a complex high dimensional data onto a lower dimensional subspace through removal of highly correlated and redundant features in the data. The dimensionality reduced data can then be processed by various anomaly based IDSs which reduce their overall computational overhead without adversely affecting their performances.

The rest of the paper has been structured in following ways. Section II discusses the related works. Section III outlines the mathematical overview of the PCA and an overall description of how it can be applied for dimensionality reduction of the data. Section IV provides a brief description about various type

of classifiers used for developing anomaly based IDSs. Section V provides the overall description of the proposed intrusion detection model. Section VI provides the experimental results. Conclusion and future work are provided in Section VII.

II. Related Works

Anomaly detection system based on back-propagation Multi Layer Perceptron (MLP) to identify normal users' profile was proposed by *Ryan et al.* [9]. Their MLP model evaluates the users' commands for possible intrusions at the end of each log session. The top 100 important commands used by the user throughout the session was used to determine the user's behavior. They used a 3 layer MLP model with two hidden layers and found that their MLP model was able to correctly identify 22 cases out of 24. Similarly, a process-based intrusion detection approach that provide the ability to generalize from previously observed behavior to recognize future unseen behavior was proposed by *Ghosh et al.* [10]. Their framework employs artificial neural networks (ANNs) and can be used for both anomaly detection in order to detect novel attacks and misuse detection in order to detect known attacks and their variations.

A standalone system that uses a three layer neural network for classification of network connection into normal or misuse class was proposed by *Cannady et al.* [11]. The feature vector used in their study comprised of feature values describing the connection and the commands used. A dataset with ten thousand connection records including one thousand simulated attacks was used in their study. Their two class classifier was successful in correctly classifying the connection records in 89-91% of the cases. In another similar study [12], the authors reported results of about 99.25% correct classification for their two class (normal and attack) problem using a three and four layer neural networks. A novel online monitoring approach to distinguish between attacks and normal activity in SIP-based Voice over IP environments was proposed by *Nassar et al.* [13]. Their approach uses a set of 38 features in VoIP flows and Support Vector Machines for classification of network activity as normal or attack.

Intrusion detection models based on Self-Organizing Maps (SOM) are discussed in [14] [15]. These models operate on real-time data without extensive off-line training and with minimal expert knowledge and are shown to achieve high performance in terms of accuracy and detection rate. Intrusion detection model based on the Principal Component Analysis (PCA) and Support Vector Machine (SVM) is proposed in [16]. They used PCA and SVM along with feature reduction technique to develop anomaly based intrusion detection model and showed that their proposed model is able to achieve high performance.

Summarizing the related works, we found that although anomaly based intrusion detection systems based on complex learning algorithms like SVM [16], ANN [9] [10] and SOM [14] [15] achieves high detection rate and accuracy, they also introduce a significant computational overhead which puts severe limitation on their practical application and real time deployment in high speed networks. The high computational overhead of anomaly based IDSs can primarily be attributed to large number of input features in the data being analyzed by these IDSs. In this paper, we aim to address the issue of high computational overhead of anomaly based IDSs through application of data dimensionality reduction technique called PCA. Applying PCA can significantly reduce the number of features of the high dimensional data by projecting it onto a lower dimensional subspace without losing any significant information contained in the original high dimensional data. The dimensionality reduced data can then be used by various anomaly based IDSs, which significantly reduces their computational overhead.

III. Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique used for finding patterns in data of high dimensionality. PCA transforms a set of correlated features in the high dimensional data onto a smaller subset of uncorrelated features called principal components, thereby projecting the higher order n-dimensional data into a low order k-dimensional data ($n > k$) without losing any significant information contained in the original data. PCA achieves this transformation by finding k eigenvectors onto which to project the n-dimensional data so as to minimize the overall projection error. This is illustrated with an example in Fig. 1. As shown in the figure the two dimensional data points (represented by features X_1 and X_2) can be projected onto any of the two lines (Line 1 and Line 2). However, PCA chooses Line 1 over Line 2 for projection, since the overall orthogonal projection error for projecting the data points onto Line 1 is much smaller compared to orthogonal projection error for projecting the same data points onto Line 2.

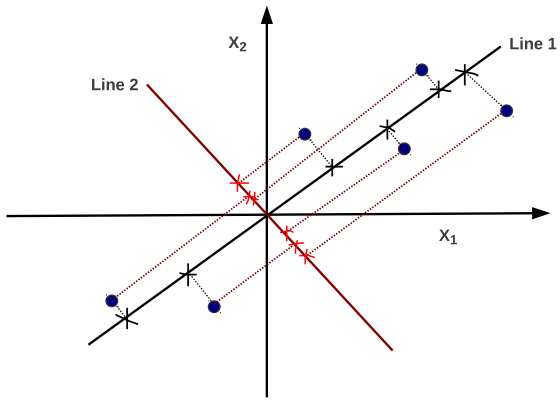

Fig. 1: Projection of two dimensional data points onto a one dimensional line using PCA

In the subsequent section, we provide a detail description of how PCA can be applied for dimensionality reduction of a complex high dimensional data onto a lower sub-dimensional data.

The first step in dimensionality reduction process using PCA is normalization of feature values of the dataset. Given an n-dimensional dataset (A) with m instances ($n < m$), the dataset (A) is mean normalized to obtain the normalized dataset (A^*).

$$A = \begin{bmatrix} a_1^1 & .. & a_{(n)}^1 \\ a_1^2 & .. & a_{(n)}^2 \\ . & .. & .. \\ a_1^m & .. & a_{(n)}^m \end{bmatrix} \; ; \; A^* = \begin{bmatrix} a_1^{1*} & .. & a_{(n)}^{1*} \\ a_1^{2*} & .. & a_{(n)}^{2*} \\ . & .. & .. \\ a_1^{m*} & .. & a_{(n)}^{m*} \end{bmatrix}$$

where $a_i^{j*} = \frac{a_i^j - \mu^j}{max_j - min_j}$, with μ^j, max_j and min_j being the mean, maximum and minimum values, respectively of the j^{th} column of dataset A.

In the next step, the covariance matrix (A^*_{cov}) of the mean normalized matrix A^* is calculated. A^*_{cov} is a $n \times n$ dimensional matrix.

$$A^*_{cov} = A^{*T} A^* = \begin{bmatrix} a_1^1 & .. & a_{(n)}^1 \\ a_1^2 & .. & a_{(n)}^2 \\ . & .. & .. \\ a_1^n & .. & a_{(n)}^n \end{bmatrix}$$

where A^{*T} is the transpose of the mean normalized dataset A^*. Singular value decomposition (SVD) [17] is then used to compute the eigenvectors of the covariance matrix A^*_{cov}. These eigenvectors are stored in a $n \times n$ dimensional matrix E_{vec}, where each column of E_{vec} represents one of the n eigenvectors of the covariance matrix A^*_{cov}.

$$E_{vec} = \begin{bmatrix} e_1^1 & e_2^1 & .. & e_{(n)}^1 \\ e_1^2 & e_2^2 & .. & e_{(n)}^2 \\ . & .. & .. & .. \\ e_1^n & e_2^n & .. & e_{(n)}^n \end{bmatrix}$$

The eigenvectors in the matrix E_{vec} are sorted from left to right with eigenvector corresponding to the largest eigenvalue represented by the first column of E_{vec} and the eigen vector corresponding to the smallest eigenvalue represented by the last column of E_{vec}. The top k eigenvectors ($k < n$) corresponding to first k columns of the matrix E_{vec} are then selected and stored in the $n \times k$ dimensional matrix A^*_{eig}. The value of k is chosen in such a way that enough eigenvectors are retained to maintain a variance of 90-98% of the original dataset (A^*).

$$A^*_{eig} = \begin{bmatrix} e_1^1 & e_2^1 & .. & e_{(k)}^1 \\ e_1^2 & e_2^2 & .. & e_{(k)}^2 \\ . & .. & .. & .. \\ e_1^n & e_2^n & .. & e_{(k)}^n \end{bmatrix}$$

Finally to project the higher order $\mathcal{R}^n$ dimensional dataset (A^*) onto a lower order $\mathcal{R}^k$ dimensional dataset Z ($k < n$), we multiply A^* with A^*_{eig}.

$$Z = A^* * A^*_{eig} = \begin{bmatrix} z_1^1 & z_2^1 & .. & z_{(k)}^1 \\ z_1^2 & z_2^2 & .. & z_{(k)}^2 \\ . & .. & .. & .. \\ z_1^m & z_2^m & .. & z_{(k)}^m \end{bmatrix}$$

The accuracy of the PCA depends on the number of principal components (k) retained during the dimensionality reduction process. To obtain these k principal components, we perform the singular value decomposition of the covariance matrix (A^*_{cov}) to obtain three independent $\mathcal{R}^{n \times n}$ dimensional matrices U, S and V, where U and V are orthogonal matrices containing the singular vectors, and S is a diagonal matrix containing the singular values of A^*_{cov}, such that A^*_{cov} = U*S*V^T. Finally, the dimensionality reduction of the data is achieved by retaining enough eigenvectors to account for some percentage of the variance of the original dataset (A^*). For instance, to retain a 98% variance of the original dataset, we choose the first k column vectors of the matrix S such that:

$$\frac{\sum_{i=1}^{k} S_{ii}}{\sum_{i=1}^{n} S_{ii}} \geq \; 0.98$$

In general, applying PCA and retaining about 90-95% variance of the original dataset can significantly reduce the number of features in the dimensionally reduced dataset. PCA achieves this by transforming the set of correlated features in the data onto a much smaller subset of uncorrelated features called principal components. This allows PCA to project the original high dimensional dataset onto a much smaller subspace without losing much of the information contained in the original dataset.

The dimensionality reduced dataset obtained after PCA are then analyzed by various classifiers namely, Naive Bayes, C4.5 decision tree, Support Vector Machine (SVM) and Multilayer Perceptron (MLP).

IV. Proposed IDS Framework

Fig 2 shows the overall architecture of the proposed IDS framework. A brief description about each component of the proposed framework is provided in subsequent sub-sections:

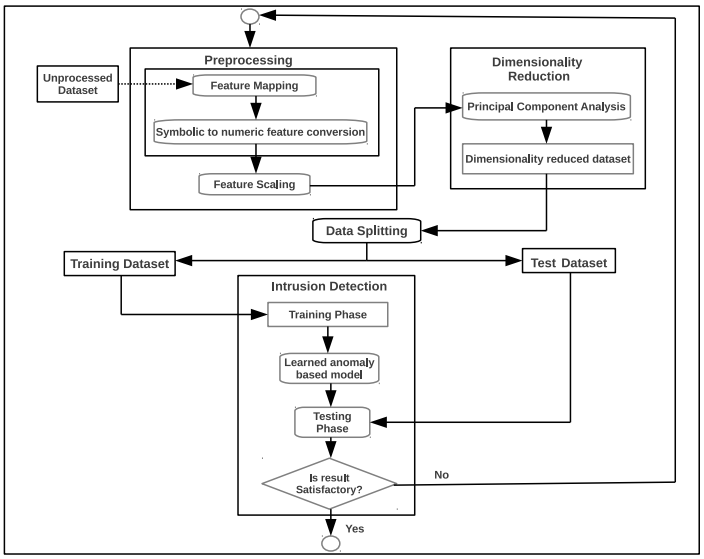

Fig. 2: Proposed Intrusion Detection System architecture

A. Preprocessing phase

Since most of the anomaly based IDSs can only process numeric dataset, some amount of data preprocessing is required. Dataset preprocessing consists of following steps:

- Conversion of symbolic attributes in the dataset to numeric attributes.
- Normalization of numeric attributes in the dataset to enhance the performance of anomaly based IDSs. Let x_j^i denote the j^{th} column feature value corresponding to the i^{th} row of the dataset. And let μ_j be the mean value of the j^{th} column feature. The feature value x_j^i is then mean normalized using the following method:

$$x_j^i = \frac{x_j^i - \mu_j}{(Max - Min)\ value\ of\ j^{th}\ column\ feature}$$

- Represent each data class in the dataset using a unique integer value.

B. Dimensionality Reduction phase

During this phase, the high dimensional dataset $\in \mathbb{R}^n$ is projected onto a lower dimensional dataset $\in \mathbb{R}^k$ ($n > k$) through application of PCA as described in Section III. The dimensionality reduction process is carried out in such a way that the new *k* features in the dimensionally reduced dataset maintains a variance of 90 - 98% of the original *n* dimensional dataset. Finally the dimensionality reduced dataset is split into 65% Training and 35% Test dataset.

C. Intrusion Detection phase: Parameter selection

During this phase, various parameters of different classifiers are set. Table I shows the parameter settings of different classifiers used in our study. The classifiers are initially trained using the dimensionality reduced training dataset to develop a learned intrusion detection model of the network. Finally, the trained classifier models are tested on the test dataset to evaluate their performances. If the performance of the trained model on the test dataset is unsatisfactory, the entire process is revisited and executed again with different set of parameters.

We have used accuracy and detection rate as two key parameters for evaluation of different intrusion detection models. Accuracy is defined as the fraction of elements correctly classified as true alarms out of all the elements the intrusion detection model classified as positive, whereas detection rate is the fraction of elements correctly classified as positive out of all positive elements in the dataset.

$$Accuracy = \frac{TP}{TP + FP}$$

$$Detection\quad Rate = \frac{TP}{TP + FN}$$

where, *TP, FP* and *FN* represents true positives, false positives and false negatives, respectively.

TABLE I: **List of classifiers with their parameter settings**

Classifiers	**Parameter settings**
Support Vector Machine (LibSVM)	Radial Basis Function (RBF) kernel, gamma = 0.4, loss = 0.002
Multilayer Perceptron (MLP)	No. of hidden layer nodes = No. of (Attribute + class)/2, Learning rate = 0.15
C4.5 Decision Tree (J48)	confidenceFactor = 0.25
Naive Bayes	——

V. Experimental Results

We have used the NSL-KDD dataset [18] [19] for analyzing the performance of our proposed intrusion detection framework. The NSL-KDD dataset consists of 41 features with varying proportion of normal and attack data as shown in Table II. PCA was applied to reduce the dimensionality of the NSL-KDD dataset by retaining 85% to 98% variance of the original dataset. In our study, we found that only 17 features were required in the dimensionally reduced dataset to retain a 98% variance of the original dataset. The performance of various IDS models based on SVM, MLP, C4.5 and Naive Bayes classifiers were then evaluated on both the original NSL-KDD dataset containing the entire 41 features and the dimensionally reduced dataset obtained after applying PCA.

A. Anomaly based IDS parameters selection

The default implementation of SVM, MLP, C4.5 and Naive Bayes (NB) classifiers available in Weka [20] were used in our analysis. The parameter settings of various classifiers used in our study is shown in Table I.

TABLE II: Data distribution of the NSL-KDD Train and Test dataset

Count of Type	Train set	Test set
Probe	6372	2940
DoS	70631	55768
U2R	1782	4302
R2L	2745	3176
Normal	17389	13703

TABLE III: Performance of various IDS models on the binary class NSL-KDD dataset without dimensionality reduction

Anomaly detection Model	No. of Features	Acc (%)	DR (%)
SVM	41	99.63	99.16
MLP	41	97.16	96.77
C4.5	41	97.35	97.98
Naive Bayes	41	95.16	91.65

*Acc : Accuracy *DR : Detection rate

Table III and Table IV shows the performance comparison of various anomaly based IDS models on the original binary class NSL-KDD dataset and the dimensionally reduced NSL-KDD dataset obtained after applying PCA with 98% variance retention, respectively. It can be observed from these tables that the accuracy and detection rate of the anomaly based IDSs obtained using the dimensionally reduced dataset is comparable to that obtained using the original higher dimensional dataset.

TABLE V: Performance of various IDS models on the multi-class NSL-KDD dataset without dimensionality reduction

		SVM		Naive Bayes		C4.5		MLP	
		Acc (%)	DR (%)	Acc (%)	DR (%)	Acc (%)	DR (%)	Acc (%)	DR (%)
Class Labels	Normal	95.44	99.13	92.38	71.15	93.69	97.15	94.74	88.48
	DoS	99.83	99.81	96.98	93.89	92.43	98.72	97.39	99.23
	U2R	99.61	63.53	73.67	60.62	94.51	97.96	97.67	72.24
	R2L	70.54	94.35	65.61	84.59	96.02	99.27	72.96	87.94
	Probe	99.73	98.67	37.42	98.09	99.02	99.62	78.90	87.52
	Average	97.93	97.56	91.52	87.97	95.31	98.74	95.05	95.05

*Acc : Accuracy *DR : Detection rate

TABLE VI: Performance of various IDS models on the multi-class NSL-KDD dataset after dimensionality reduction using PCA with 98% variance retention (17 features)

		SVM		Naive Bayes		C4.5		MLP	
		Acc (%)	DR (%)	Acc (%)	DR (%)	Acc (%)	DR (%)	Acc (%)	DR (%)
Class Labels	Normal	95.13	98.93	92.18	70.85	93.19	96.95	94.34	88.17
	DoS	99.13	99.33	96.26	93.17	92.11	98.36	96.93	98.94
	U2R	98.97	63.11	73.19	60.47	93.96	97.35	97.07	71.87
	R2L	69.87	93.79	65.21	84.17	95.78	98.91	72.37	87.29
	Probe	99.47	98.23	36.92	97.81	98.53	98.92	78.16	87.11
	Average	97.61	97.27	91.35	87.67	95.13	98.57	94.91	94.87

TABLE IV: Performance of various IDS models on the binary class NSL-KDD dataset after dimensionality reduction using PCA with 98% variance retention

Anomaly detection Model	No. of Features	Acc (%)	DR (%)
SVM	17	99.13	98.68
MLP	17	96.76	96.13
C4.5	17	96.85	97.23
Naive Bayes	17	94.56	90.75

However, the number of features analyzed by these IDSs in the dimensionality reduced dataset is significantly less (17) compared to that in the original dataset (41), which greatly decreases their computational overhead.

Table V and Table VI shows the performances of various anomaly based IDS models on the original multi-class NSL-KDD dataset and on the dimensionality reduced NSL-KDD dataset with 98% variance retention, respectively. Again it can be observed that the accuracy and detection rate of these anomaly based IDS models on the dimensionally reduced dataset is comparable to that obtained using the original dataset with the entire 41 features. The detection rate of various classifiers on the dimensionality reduced and original NSL-KDD dataset is shown in Fig. 3. Similarly, the accuracy of various classifiers on the dimensionality reduced and original NSL-KDD dataset is shown in Fig. 4

Fig. 5 and Fig. 6 shows the detection rate and accuracy of various anomaly based IDS models on the dimensionality reduced NSL-KDD dataset with different degree of variance retention of the original dataset. It can be observed from these figures that the performance of different anomaly based IDS models on the dimensionality reduced dataset with 98% variance retention (17 features) is comparable to the performance of these IDS models on the original NSL-KDD dataset with the entire 41 features.

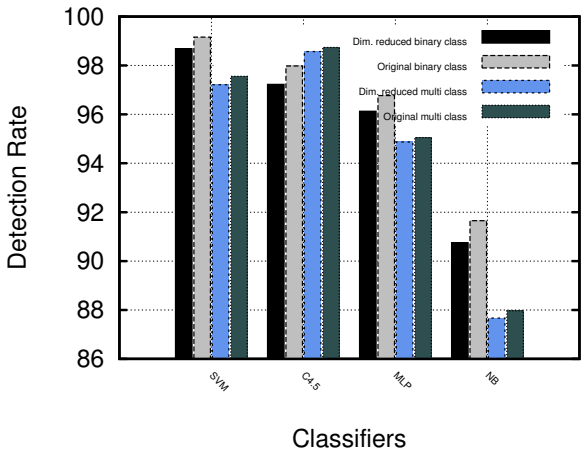

Fig. 3: Detection rate of various classifiers on the original and the dimensionality reduced NSL-KDD dataset

VI. Conclusion and Future Work

In this paper, we proposed an intrusion detection framework to enhance the performance of anomaly based IDSs through data dimensionality reduction technique called Principal Component Analysis (PCA). Experimental results show that the performance of various anomaly based IDS models on the dimensionally reduced dataset obtained after applying PCA is comparable to that obtained using the entire feature set of the original dataset. However, since the number of features in the dimensionality reduced dataset is much smaller compared to that of the original dataset, the computational overhead of anomaly based IDSs are significantly reduced which makes them viable for real time deployment in high speed networks. For our future work, we aim to fine tune various parameters of the proposed framework to further enhance its performance.

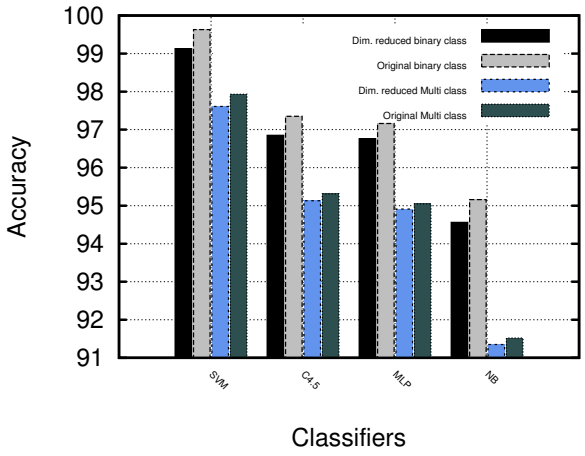

Fig. 4: Accuracy of various classifiers on the original and the dimensionality reduced NSL-KDD dataset

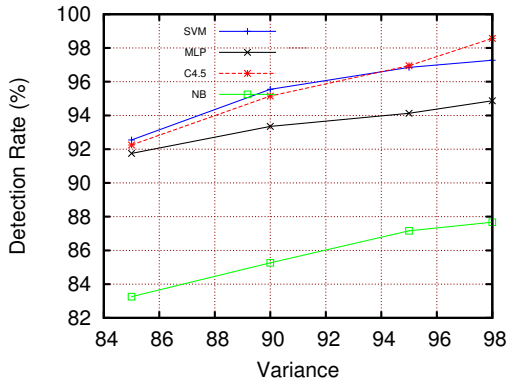

Fig. 5: Detection rate of various anomaly based IDS models on the dimensionality reduced multi-class NSL-KDD dataset with varying degree of variance retention

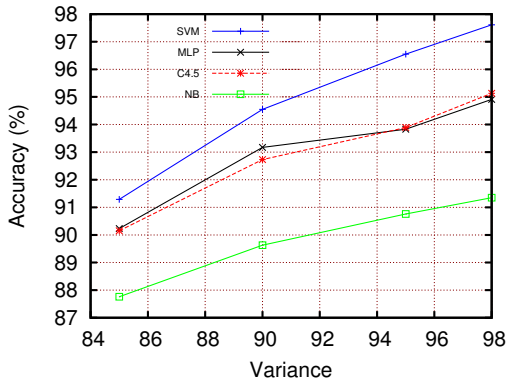

Fig. 6: Accuracy of various anomaly based IDS models on the dimensionality reduced multi-class NSL-KDD dataset with varying degree of variance retention

REFERENCES

[1] E. Karapistoli, P. Sarigiannidis, and A. A. Economides, *Visual-Assisted Wormhole Attack Detection for Wireless Sensor Networks*. Springer International Publishing, 2015, pp. 222–238.

[2] ——, "SRNET: A Real-time, Cross-based Anomaly Detection and Visualization System for Wireless Sensor Networks," in *Proceedings of the Tenth Workshop on Visualization for Cyber Security*. ACM, 2013, pp. 49–56.

[3] S. Ranjitha Kumari and P. Krishna Kumari, "A comparative analysis of stream data classifiers and conventional classifiers for anomaly intrusion detection," *Advanced Science Letters*, vol. 21, no. 10.

[4] M. Roesch, "Snort - Lightweight Intrusion Detection for Networks," in *Proceedings of the 13th USENIX Conference on System Administration*, 1999, pp. 229–238.

[5] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-time," *Computer Networks*, vol. 31, no. 23-24, pp. 2435–2463, Dec. 1999.

[6] D. Moore, C. Shannon, and k. claffy, "Code-Red: A Case Study on the Spread and Victims of an Internet Worm," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurment*, 2002, pp. 273–284.

[7] Y. Yu and H. Wu, "Anomaly intrusion detection based upon data mining techniques and fuzzy logic," in *IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2012, pp. 514–517.

[8] L. Coppolino, S. D'Antonio, A. Garofalo, and L. Romano, "Applying Data Mining Techniques to Intrusion Detection in Wireless Sensor Networks," in *Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, Oct 2013, pp. 247–254.

[9] J. Ryan, M. Lin, and R. Miikkulainen, "Intrusion Detection with Neural Networks," in *Advances in Neural Information Processing Systems 10, [NIPS Conference*, 1997, pp. 943–949.

[10] A. K. Ghosh and A. Schwartzbard, "A Study in Using Neural Networks for Anomaly and Misuse Detection," in *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, 1999, pp. 12–12.

[11] J. Cannady, "Artificial Neural Networks for Misuse Detection," in *NATIONAL INFORMATION SYSTEMS SECURITY CONFERENCE*, 1998, pp. 443–456.

[12] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection using neural networks and support vector machines," in *Proceedings of the 2002 International Joint Conference on Neural Networks, 2002. IJCNN '02.*, vol. 2, 2002, pp. 1702–1707.

[13] M. Nassar, R. State, and O. Festor, "Monitoring SIP Traffic Using Support Vector Machines," in *Proceedings of the 11th International Symposium on Recent Advances in Intrusion Detection*, 2008, pp. 311–330.

[14] P. L. Nur, A. N. Zincir-heywood, and M. I. Heywood, "Host-Based Intrusion Detection Using Self-Organizing Maps," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, 2002, pp. 1714–1719.

[15] K. Labib and R. Vemuri, "NSOM: A Real-Time Network-Based Intrusion Detection System Using Self-Organizing Maps," 2000.

[16] F. E. Heba, A. Darwish, A. E. Hassanien, and A. Abraham, "Principle components analysis and Support Vector Machine based Intrusion Detection System." in *ISDA*, 2010, pp. 363–367.

[17] M. Aharon, M. Elad, and A. Bruckstein, "SVDD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[18] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *IEEE Symposium on Computational Intelligence for Security and Defense Applications*, July 2009, pp. 1–6.

[19] http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[20] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems)*, 2005.