International Journal of Security and Its Applications
Vol. 11, No. 5 (2017), pp.1-14
http://dx.doi.org/10.14257/ijsia.2017.11.5.01

A Coalesce of SNE-Wavelet-SVM Technique for Network Intrusion Detection

Yasir Hamid[1], Ludovic Journax[2], Firdous A. Shah[3] and M.Sugumaran[4]

[1]*Dept. of Computer Science, Pondicherry Engineering College, India*
[2]*Lab. LE2I UMR CNRS 6306, Dijon, France*
[3]*Dept. of Mathematics, University of Kashmir, India*
[4]*Dept. of Computer Science, Pondicherry Engineering College, India*
[1]*bhatyasirhamid@pec.edu*, [2]*l.journaux@agrosupdijon.fr*, [3]*fashah79@gmail.com*, [4]*sugu@pec.edu*

Abstract

Recognizing intrusions quickly and precisely is vital to the proficient operation of computer networks. Precisely describing critical classes of intrusions extraordinarily encourages their recognizable proof; be that as it may, the nuances and complexities of anomalous activities can without much of a stretch complicate the procedure. Due to the inherent capability of the signal processing to discover the novel and obscure attacks, they have been pretty popular for Network Intrusion Detection, and the nearness of the self-comparability in the system activity propels the appropriateness for the application Wavelets. In this work we first subject the network data to dimension reduction using Stochastic Neighbor Embedding (SNE) and then preform the wavelet decomposition of the data. The classification results of the pre-processed data using Gaussian SVM over different bandwidths uphold the claim that the proposed system has appreciably improved detection coverage for all the attack groups and the normal data as well, and at the same time minimized the false alarms.

Keywords: *Dimensionality Reduction; Feature Extraction; Intrusion Detection; Stochastic Neighbor Embedding; Wavelet Transforms*

1. Introduction

As more and more countries are pushing for a cashless economy, plastic currency and e-governance becoming a new norm, the need to secure the networks is more than ever before has become necessary [1]. The fundamental principles of information security i.e., Confidentiality, Integrity and Availability (CIA) guarantee that solely authenticated and exclusive validated entities are able to reliably access secure information [2]. The only way to retain the user-trust and attract more and more institutions to put their resources online is by assuring adequate security and privacy to their valuable assets. Internet being heterogeneous and distributed commodity, lags the centralized security mechanism [3] and is in constant threat of the malicious activities of the attackers. The motives of the attacker may be diverse greed, fun, and hatred, military and economic espionage to mention a few [4]. In order to provide adequate security to the user, a layer of citadel has been pushed in over the years [5]. The preventive measures like firewalls, filters, anti-viruses and proxies have been pretty popular now for a long time [6]. No matter how securely the network is laid, how much preventive measures are put, the attacker will find a way to compromise the system and possibly pose a threat to retain the user base [7]. So, this leads us to the other avenue, relatively newer on scene i.e., continuously monitoring the system for security compromises, and one such tool that has been pretty popular is Intrusion Detection System (IDS) [8].

ISSN: 1738-9976 IJSIA

An IDS may be a hardware, software or the combination of both, against whom the responsibility of unbolting conceivably adverse connections from the network flow is laid [9]. An IDS may be categorized as Host-based or Misuse-based depending on its position of placement within the network and Misuse-based or Anomaly-based depending on the detection technique that is at its coronary heart. Some Hybrid systems have also been proposed and are gaining popularity in accordance with the complementary nature of various system components [10]. Over the years the IDS problem has been visited by the researchers in varied and widespread environments, most popular of them all, being the application of Machine Learning (ML) techniques [11] [12]. Both supervised as well as unsupervised ML techniques have been equally popular. Even though exhaustive research has been carried out in the implementation of ML there has been only minimal acceptance in the systems for the real environments. There are many reasons, some of them may be, labour intensive labelling process, high false alarm rate, the lesser detection rate for few attack groups like User to Root (U2R) and Remote to Local(R2L) [13]. Most of the problems that hamper the application of the system in the real network are caused mainly due to following reasons i.e., imbalanced class distribution, the curse of dimensionality, sparsity of data, and noise.

We, in this work, have attempted to visit most of the shortcoming of intrusion detection. As for imbalanced class distribution is considered, we have used SMOTE [14] based sampling of minority classes and random under-sampling of dominant groups of the data. In order to minimize the effects of the curse of dimensionality we applied SNE [15], an unsupervised Non-Linear Dimension Reduction (NLDR) to project the data onto lower dimensions without sacrificing the inherent information of the data. Till now the only dimension reduction (DR) technique that has been applied to IDS is linear Principal Components Analysis (PCA) [16]. The field of DR has moved a long way forward after the inception of PCA, over the years a lots of powerful NLDR methods have been devised whose effectiveness for ID is yet to be tested, SNE is one such NLDR technique. And finally, the noise is minimized by performing the Discrete Wavelet Transformation (DWT) of the data, the data signals were decomposed over five levels using Coiflet wavelet family. This transformed data is subjected to classification using non-linear SVM with Gaussian kernel on five different bandwidths. The results show that the proposed system with SNE and Wavelet has improved detection rate and at the same time has the reduced false alarm rate, which makes the system highly competent of being applied in the real networks. A comparison of the proposed system with the prominent works in the area has proved that the proposed system has better detection rate for all the attack groups than all the works.

The remaining of the paper is organized as follows. In Section 2 we present a brief review of the related literature. A brief discussion about methods and materials used in this work is given in Section 3. The experimental methodology of the work is presented in Section 4, results and discussions is given in Section 5. Finally, the paper concludes in Section 6.

2. Motivation and Literature Review

Off all the supervised ML techniques, SVM has been most popular [17,18,19] and authors in having applied SVM mostly complimented by some pre-processing techniques [20,21,22] for Network Intrusion Detection. Most popular of the pre-processing technique is feature extraction and dimension reduction [23]. The only dimension reduction technique that has been extensively used in IDS is PCA [24,25]. But the world of dimension reduction has come a long way forward after the inception of PCA, many Non-Linear techniques have surfaced up and have been extensively used in other classification and pattern matching [26,27]. As for we know, only one NLDR technique, Locally Linear Embedding (LLE) [28] has been applied for network intrusion detection by Dash et.al in

[29]. As for wavelets are considered there has been pretty much of work towards its application for IDS mainly to denoise the signal [30] [31] [32]. Both Neural Networks [33] and SVM's [34] have been equally popular as a classifier for wavelet decomposed data. Most of them take into consideration only a particular group of attacks, this work provides a holistic approach by taking into consideration all the twenty three attack types being categorized into four different groups. There have been only a few works that have actually pre-processed the data before decomposing it using wavelets.

3. Materials and Methods

In the next few subsections we present an exhaustive discussion of various materials and methods applied in this work.

3.1 Stochastic Neighbor Embedding

Stochastic Neighbor Embedding (SNE) is unsupervised Non Linear Dimension Reduction technique. It has been pretty popular in ML community for visualizing high dimensional data into low dimensional (typically 2 or 3) spaces. Given a vector $x \epsilon \mathbb{R}^n$, dimension reduction finds an embedding $y \epsilon \mathbb{R}^m$, for x such that $n \gg m$. The advantages of representing x by y are many, few of them are saving storage, minimizing the computational time and enhancing the detection rate of the classification model. In order to transform the data from feature space to object space, SNE minimizes the divergence of the conditional probabilities of the two objects x_i and x_j conditional probability of the two points x_i and x_j being the neighbors is HD space P will be neighbors in LD space Q. The similarity of the two objects x_i and x_j is HD space P is given by the conditional probability $P_{i|j}$ that the two objects are neighbors as given by the below equation.

$$P_{j|i} = \frac{\exp(-\left||x_i - x_j|\right|^2 / 2\sigma^2)}{\sum_{k \neq i} \exp(\left||x_i - x_k|\right|^2 / 2\sigma^2)} \tag{1}$$

Where $|x_i - x_j|$ is usually taken as Euclidean distance between the two objects, and σ is the bandwidth of Gaussian. The similarity between two objects in low dimensional space is calculated as the conditional probability

Replacing σ by $\frac{1}{2}$ the Equation 2 can be rewritten as

$$Q_{j|i} = \frac{\exp(-\left||y_i - y_j|\right|^2)}{\sum_{k \neq i} \exp(\left||y_i - y_k|\right|^2)} \tag{2}$$

Where y_i is the embedding of x_i. For an ideal embedding $P_{i|j}$ and $Q_{i|j}$ should be almost equal. A cost function of SNE is aimed at attaining a proper balance between $P_{i|j}$ and$Q_{i|j}$. A natural measure is Kullback Leiber divergence of the form

$$KLD(P_i||Q_i) = \sum_i \sum_j P_{j|i} \, log \frac{P_{j|i}}{Q_{j|i}} \tag{3}$$

The cost function of the SNE is derived by summing up the conditional Kullback Leibler divergence over all the variables x_i given by,

$$W = \sum_i KLD(P_i|Q_i) \tag{4}$$

A gradient of the form

$$\frac{\delta W}{\delta y_i} = 2 \sum_j \left(P_{j|i} - Q_{j|i} + P_{i|j} - Q_{i|j}\right)(y_i - y_j) \tag{5}$$

is used to adjust the cost function over the iterations.

3.2 Wavelet Transform Method

Wavelet transform serves as an important and powerful analysing tool for time-frequency analysis and has been applied in a number of fields including signal processing, computer graphics, neuro-sciences, sampling theory, quantum mechanics and medicine. Wavelet investigation imparts a few components in like manner to Fourier analysis yet has the benefit of conquering elements in the fundamental arrangement that fluctuate crosswise over both time and frequency. Mathematically, a wavelet is a function $\psi(t) \epsilon\, L^2(\mathbb{R})$ which satisfies the case:

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \tag{6}$$

where $\hat{\psi}(\omega)$ is the Fourier transform of $\psi(t)$. Equation (1) is called the admissibility condition which guarantees the existence of the inversion formula for the continuous wavelet transform. Wavelets constitute a group of operations developed from the dilation and translation of a solitary operation called the mother wavelet $\psi(t)$. At the point when the dilation parameter a and the translation parameter b change consistently, we have the accompanying group of continuous wavelets as

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), a \neq 0, b \epsilon \mathbb{R} \tag{7}$$

As far as frequency component of the signal is considered, lesser values show cases high frequency and greater values show cases low frequency of the signal. Moreover, when the parameters a and b are restricted to discrete values as $a = a_0^{-j}$; $b = k b_0 a_0^{-j}$; where a_0 and b_0 are fixed positive constants, we have the following family of discrete wavelets:

$$\psi_{j,k}(t) = a_0^{j/2} \psi(a_0^j t - k b_0), \quad j,k \in \mathrm{Z} \tag{8}$$

For computational efficiency, $a_0 = 2$ and $b_0 = 1$ are commonly used so that Equation (3) takes the form

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \quad j,k \in \mathrm{Z} \tag{9}$$

In inclusion to that admissibility criteria (1) there are different properties that might be helpful specifically applications. For example, confinements on the support of ψ and of $\hat{\psi}$ might be necessary to have a specific number of vanishing moments that serves as the regularity of the wavelet functions and capacity of a wavelet transform to confiscate localized information. A wavelet ψ(t) has N-vanishing moments if the following criteria is satisfied:

$$\int_{-\infty}^{\infty} t^k \psi(t) dt = 0, \quad k = 0,1,2,\ldots,N \tag{10}$$

The quantity of vanishing moments is specifically identified with the regularity of the wavelet. Hence, a greater regular wavelet has more number of vanishing moments. In the last few decades, different types of orthonormal wavelet families have been constructed such as Haar wavelets, Battle-Lemari´e wavelets, Daubechies wavelets, Coifman wavelets

(Coiflets), Biorthogonal wavelets, Harmonic wavelets, Legendre wavelets, M-band wavelets and Composite wavelets.

Considerable attention has been stimulated for the construction and design of wavelets with compact support because of the two main reasons: to have fast numerical algorithms and good time or space localization properties. Among all the compactly supported wavelets, the Daubechies wavelets have scored prominence among researchers because of their valuable properties, for example, regularity, orthogonality and impermeable support yet the real disadvantage of these wavelets is their symmetry and the corresponding scaling function is not likewise of compact support. Further to overcome these disadvantages, coiflet bases were first introduced by Daubechies [35] in the context of wavelet theory, which consider both the scaling and the wavelet capacities to have a high number of vanishing moments.

An orthonormal wavelet system with compact support, is called Coiflet system of degree N if both the scaling and wavelet functions ϕ and ψ satisfy

$$\int_{-\infty}^{\infty} t^k \psi(t)dt = 0, \quad \int_{-\infty}^{\infty} t^k \phi(t)dt = \delta_{0,k}, \quad k = 0,1,2,\ldots,N \tag{11}$$

3.3 Support Vector Machine

Support Vector Machine (SVM) is one of the most popularly applied supervised machine learning technique for classification and regression. In its basic form SVM is a two class classifier, based on the concept of hyper parameter, SVM finds a hyper plane SVM is entrusted to find a hyper plane that maximizes the margin between the nearest points on either side of hyper plane. SVM is one of the best kind of Computational Learning theory which can be applied both to classification and regression problems. The working principle of SVM is to figure out a hyper-plane that separates well the different set of sample data i.e, finds the greatest minimal separation between the training samples. SVM can be applied both to linear and non-linear data. Linear separability is that a hyper-plane exists that can clearly find the serration amidst the samples, when such a separation doesn't exist, there needs to apply a kernel trick to sort out a perfect hyper-plane separating data.

Figure 1 shows one of the cases chosen as hyper plane which has minimum margin compared to other hyper plane. Although it's a hyper-plane that separates the data items well, but there is always a chance of data items slipping the other side i.e., being misclassified.

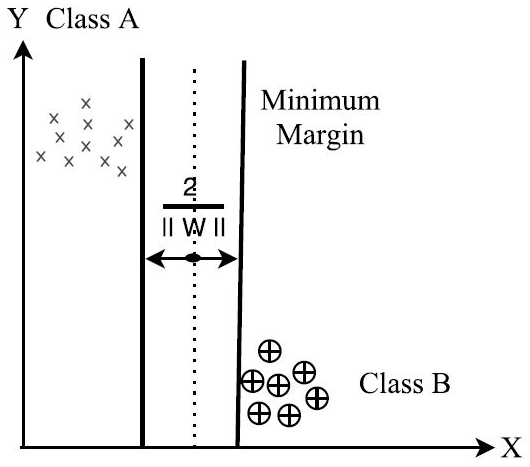


Figure 1. Hyperplane with Minimum Margin

The objective is to find a hyper-parameter that maximizes the margin and hence has lesser chances of miss classifications. Figure 2 shows the hyperplane model which has the maximum margin separating the classes certainly.

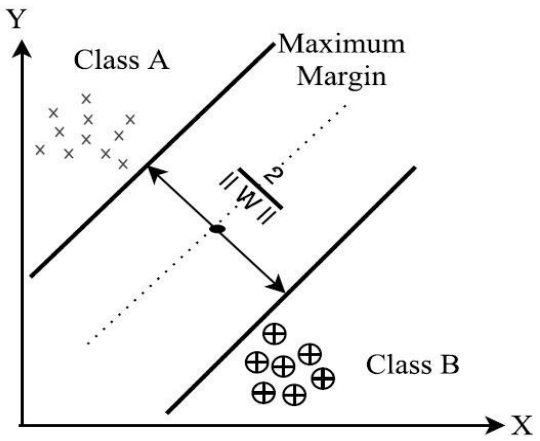


Figure 2. Hyperplane with Maximum Margin

As shown in the figure the margin is the distance between two closest points on either side of hyper plane is given by $\frac{2}{||W||}$. SVM aims at maximizing this margin by solving the objective function.

$$\begin{aligned} &\underset{W,b}{\text{Maximize}} && \frac{2}{||W||}, \\ &\text{subject to} && \beta_i(W \cdot \alpha_i + b) \geq 1, \forall i = 1..N \end{aligned} \tag{12}$$

Separable Case:

According to Vapnik's original formulation, for an available training set $\{\alpha_i, \beta_i\}_{i=1}^{N}$ where $\alpha_i \epsilon \mathbb{R}^n$ is input and $\beta_i \epsilon \{-1, +1\}$ is the classification label, there exist two conditions which have to be satisfied.

$$W.\alpha_i + b \begin{cases} \geq 1, \forall \alpha_i \epsilon \beta_i = +1 \\ \leq -1, \forall \alpha_i \epsilon \beta_i = -1 \end{cases} \tag{13}$$

These two conditions can be aggregated as

$$\beta_i(W \cdot \alpha_i + b) \geq 1, \forall i = 1..N \tag{14}$$

Where W is the weight and b is the bias

or alternatively the above equation can be rewritten as

$$1 - \beta_i(W \cdot \alpha_i + b) \leq 0, \forall i = 1..N \tag{15}$$

Inseparable Case

We need to introduce a slack variable ζ_i in the conditions where a linear function can't clearly define the separation.

$$s_i(W \cdot \alpha_i + b) \geq 1 - \zeta_i, \zeta_i \geq 0, i = 1,2..N \tag{16}$$

Now the goal is to find geometric boundary of the classifier which has maximum band width in the midst of different classes. This can be solved by reconstructing optimization problem as

$$\begin{aligned} &\underset{W,b}{\text{Min}} && \frac{1}{2}\|W\|^2 \\ &\text{subject to} && 1-\beta_i(W\cdot\alpha_i+b)-\zeta_i\le 0, \qquad \forall i=1\ldots N \end{aligned} \tag{17}$$

This optimization problem can be solved by formulating Lagrangian function as,

$$L(W,b,\zeta,\mu,\nu)=\frac{1}{2}\|W\|^2+C\sum_{i=1}^{n}\zeta_i$$

$$+\sum_{i=1}^{n}\mu\{s_i[W\cdot\alpha_i+b]-1+\zeta_i\}+\sum_{i=1}^{n}\nu_i\zeta_i \tag{18}$$

We will get a quadratic programming problem by solving Lagrangian function,where $K(r_i,r_j)$ is the kernel function, ζ is Lagrangian multiplier and ν is mean.

Thus with kernel deceptions we were able to solve inextricable data samples of the training set.

3.4 Radial Basis Function

One such kernel radial basis function also known as Gaussian function is very popular nonlinear kernel in ML community to measure the similarity or distance (whatever the case may be) between two samples. It has predominantly used in the past for support vector classification [36].

The RBF kernel on two samples $x\ and\ y$, represented as feature vectors is calculated

$$K(x,y)=exp(\frac{\|x-y\|^2}{2\sigma^2}) \tag{19}$$

as

Where, $\left|\|x-y\|\right|^2$ may be perceived as squared Euclidean distance between the two feature vectors $x\ and\ y$. A simpler definition involves a parameter $\gamma=\frac{1}{2\sigma^2}$ .Rewriting the Equation 19 , by including γ

$$k(x,y)=exp(-\gamma\|x-y\|^2) \tag{20}$$

The value of RBF kernel ranges between the interval [0-1] and decreases with the distances. The RBF has value 1 when (x = y), hence it can be interpreted as a similarity measure.

3.5 KDD'99 Dataset

KDD'99 is the most prominently used data set for the evaluation of misuse and anomaly detection methods for network intrusion detection. Based on DARPA'98 [38] intrusion detection evolution program the KDD'99 was prepared by [37]. Being around 4 gigabytes of packed crude (double) tcp dump information of 7 weeks of system activity, DARPA'98 can be handled into about 5 million connection records of 100 bytes each. In addition to 5 million connections for training around 2 million connection records are spared for testing in KDD test set. KDD training dataset comprises of roughly 4.9 million

connection vectors each of which contains a total of 41 features (some being nominal and others numeric) and is labelled as either normal or an attack, with precisely one particular attack type. A total of 23 attack groups are present in the dataset which are grouped into four broad categories DOS, PROBE, R2L and U2R. Mostly due to the computational constraints researchers select a subset of the full dataset and use 10 fold cross validation for the testing purpose.

4. Experimental Methodology

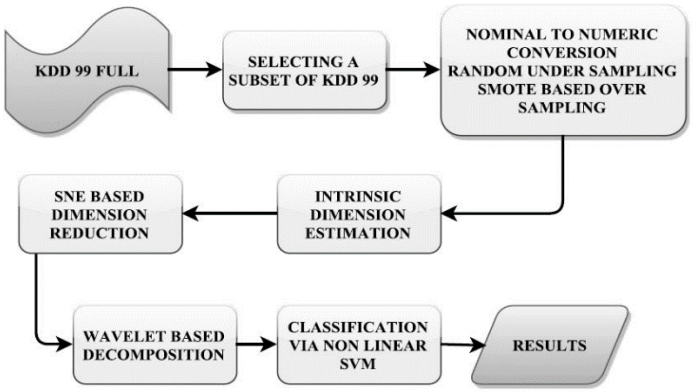


Figure 3. Block Diagram

A block diagram of the proposed model is given in Figure 3. A random subset of the instances is selected from the Full KDD99 dataset, the purpose of selecting a subset is to make the application of SNE and Wavelet practically possible.

Since, there are few attributes that are nominal in nature we proceed by converting the nominal to numeric using arbitrary coding of the nominal values. Added to eliminated class discrepancy if any, we used SMOTE based over sampling and Random under sampling to attain a balance of instances across the datasets. This balanced dataset was subjected to fractal analysis for inherent dimensionality estimation which gave the value of 3. Later SNE was implemented for reducing the dimensionality of the dataset and to project it on three dimensions. We did an experimental survey of different wavelet families on the data, and results showed that the Coiflet wavelet had superior performance than all its counterparts for ID. The reduced dataset was subjected to denoising using wavelet analysis up to level 5. On the denoised dataset a non-linear SVM with Gaussian kernel over 5 different bandwidths was applied. The results show that the proposed model of complementing a non-linear dimension reduction with signal denoising greatly improves the detection rate and at the same time reduces the false alarms.

5. Results and Discussions

The experiments were run over five bandwidths of Gaussian, Table 1 given below presents the best results that we got on the data. In addition to the best results presented in the Table 1 we also present all the results at all different values of gammas in the figures following. As we are already well acquainted with the fact that the data set for the IDS is basically a multiclass classification problem, and hence the detection rate for all the classes may not be uniform. Just to make the data easily comprehensible for the reader we have reported the weighted average of the model over all the attack groups including the normal data as well. As can be seen from the table that complimenting the SNE by wavelets has improved the results appreciably. First eliminating the curse of dimensionality using SNE improved results, then applying the wavelets on the reduced

data sets for removing any noise in the data boosted the results more. SNE and Wavelet has a highest detection rate of 99.402 followed by SNE reduced and RAW data set. Not just has been the accuracy better, rather SNE Wavelet has reduced false positive rate, and RMSE considerably. The ROC of SVM on SNE + Wavelet is 0.997 which is much better than the other two. Precision and Recall have also improved appreciably.

Table 1. Best Results

Technique	Accuracy	TP Rate	FP Rate	Precision	Recall	ROC	RMSE
RAW	89.174	0.821	0.021	0.893	0.891	0.988	0.186
SNE	97.268	0.963	0.003	0.963	0.963	0.987	0.196
SNE+Wavelet	**99.402**	**0.994**	**0.000**	**0.994**	**0.994**	**0.997**	**0.023**

In the next few figures we provide a line graph of various SVM performance on all the three datasets over all the $Gamma$ values. The purpose of putting up the graphs is to give a clear idea that the proposed system performs efficiently on all the $Gamma$ values and is not a mere coincidence.

Figure 4 given represents the accuracy of SVM on different $Gamma$ values over the dataset. The Accuracy is calculated as $ACC = (TP+TN)/(TP+FP+FN+TN)$. As is pretty clear from the figure that SVM has better Accuracy on SVM and Wavelet data. A highest accuracy of 99.4 is reported on $Gamma = 1$. For both simple SNE and SNE followed by wavelet they Accuracy at $Gamma = 0.1$ starts somewhere around 92.

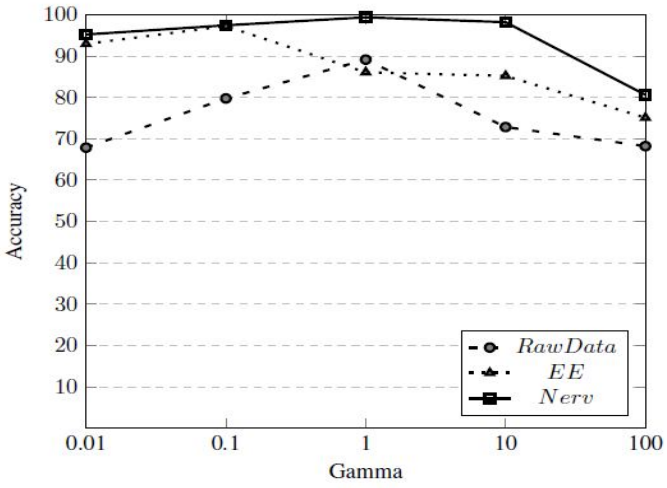


Figure 4. Accuracy

Figure 5 given represents the False Positive Rate of SVM on different $Gamma$ values over the dataset. FP Rate is the ratio of false positives to the sum of FP and TN i.e., $FPR = \frac{FP}{FP+TN}$. Lesser the FP rate better is the model. As can be seen from the Figure 5, SNE+Wavelet has least False Positive Rate over all $Gamma$ values.

Figure 6 given below represents the Precision of SVM on different $Gamma$ values over the dataset ratio of correctly classified positive instances to the total number of instanced classified as being positive i.e, $P = \frac{TP}{TP+FP}$. A higher value of precision is

preferable. As is pretty evident from the Figure 6 that SVM reports better Precision on SNE+Wavelet over all the bandwidths of the Gaussian. For both Raw and SNE reduced dataset the SVM reports an oscillating precision, and for SNE+Wavelet alone, the reported precision is somewhat stable.

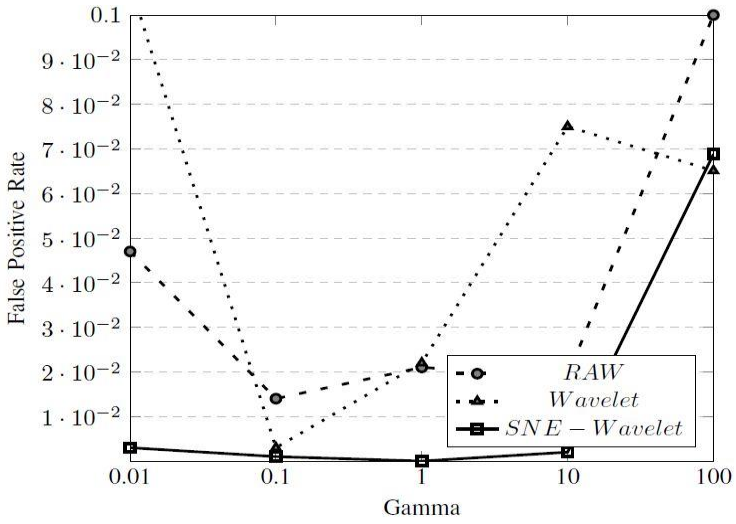


Figure 5. False Positive Rate

Figure 7 represents the Recall of SVM on different $Gamma$ values over the dataset. Recall also called True Positive Rate or specificity, is the portion of positive class instances that are correctly classified as being positive is calculated as $TPR = \frac{TP}{TP + FN}$.

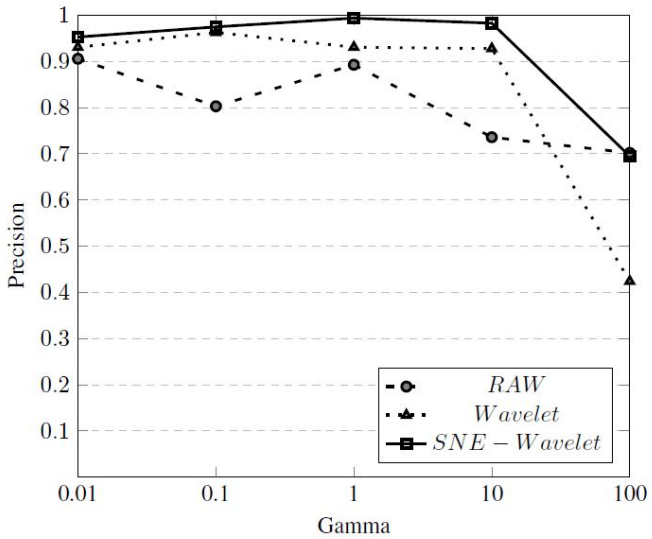


Figure 6. Precision

As can be seen from the Figure 7, that the non-linear SVM returns better recall on SNE + Wavelet model on all the gamma values. As for Raw data is considered it returns the worst performance of SVM in terms of recall. For the raw data SVM has a recall of .68 at $Gamma = 0.01$ and has the best recall at $Gamma = 1$ and after that Recall again falls for the Raw Data. Same is the case with SNE reduced data, for SNE+Wavelet alone the

Recall remains constant till the $Gamma$ is 10, at $Gamma = 100$, even though the recall falls for SNE+Wavelet but this is still better than all other configurations.

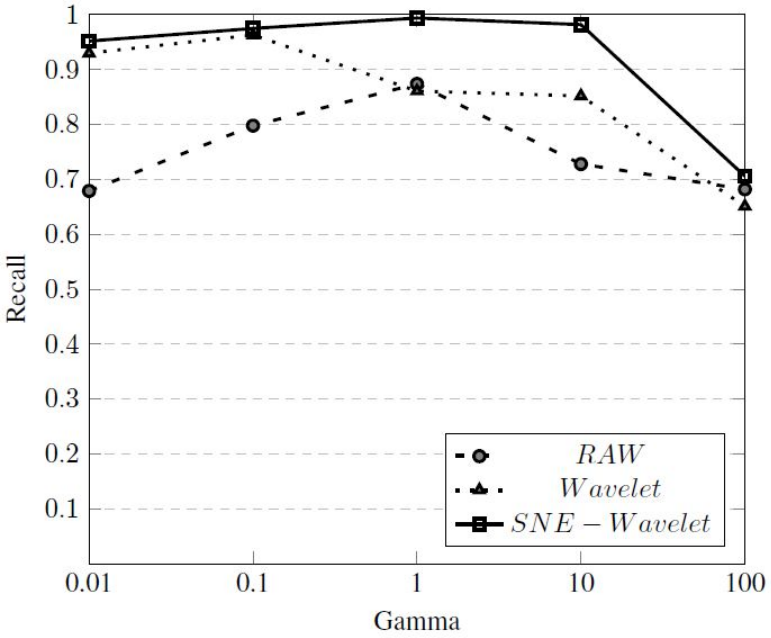


Figure 7. Recall

Figure 8 represents the ROC of SVM on different $Gamma$ values over the dataset. ROC is the area under the curve acquired by plotting true positive rate against false positive rate at different thresholds of an indicative test. Since, ROC is a curve we usually take area under that curve as the metric. The purpose of ROC is to diagrammatically show the trade-off between sensitivity and specificity.

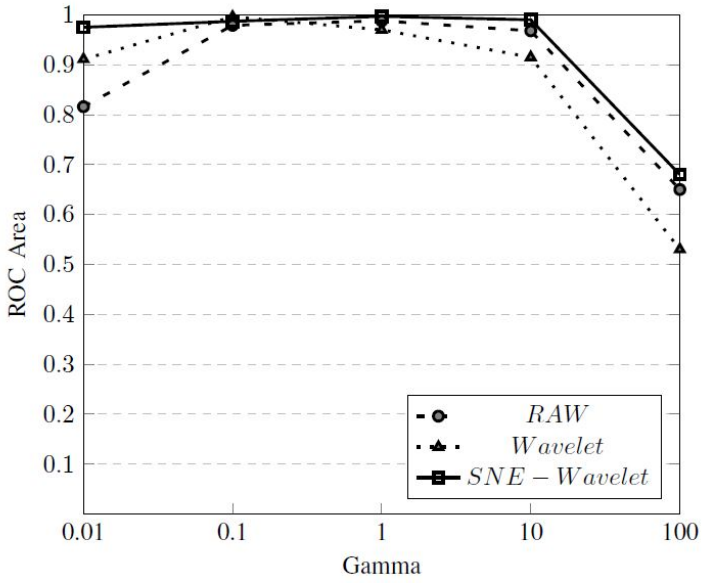


Figure 8. ROC Area

Table 2 gives a comparison of the proposed model of SNE + wavelet along with other pioneers of this field. As from the table given it can be seen that almost all the works have had lesser detection rate for two minority classes i.e., $U2R$ and $R2L$. The lesser detection

rate for these groups of attacks as predicted by the authors is due to the class skewness, sparsity of data and the noisy training data. We in this work have tried to visit these problems and the results are fascinating. As can be seen from the table the proposed model has very high detection rate for all the major as well as less dominant classes of the data and also for the normal data.

Table 2. Comparison with Related Works

MODEL	NORMAL	DOS	PROBE	R2L	U2R
Toosi et al. [39]	98.20	99.50	84.10	31.50	14.12
Xiang et al. [40]	96.80	98.66	93.40	71.43	46.97
Wang et al. [41]	97.94	97.50	76.38	15.38	09.77
Raw Data	98.29	96.51	93.56	91.20	81.80
SNE Reduced Data	98.90	98.50	91.54	93.44	81.80
SNE Wavelet	**99.87**	**99.67**	**97.76**	**96.28**	**92.71**

6. Conclusions

In this work a hybrid model for network intrusion detection was proposed. The age old problem or network intrusion detection i.e., too many false alarms, insignificant detection estimate for U2R and R2L attacks were particularly visited. SNE, a non-linear dimensionality reduction technique was applied to minimize the effects of curse of dimensionality. There on wards wavelet based decomposition was applied to denoise the signal and improve the detection capability of non-linear SVM. The experiments were run on five Gaussian bandwidths and in all the configurations the proposed system returned the better detection rate. As for detection rate of U2R attacks is considered it is 92.7 which still has the scope to be improved and will be tried in the future works. The only drawback with this work is that SNE is not scalable for very large datasets, thus becoming computationally unfeasible. In future works we would like to try different wavelet families yet to be explored and also other NLDR techniques for network intrusion detection, moreover we will try to check the effectiveness of the system on neural networks as well.

References

[1] M. A. Aydin, A. H. Zaim, and K. G. Ceylan, “A hybrid intrusion detection system design for computer network security,” Computers & Electrical Engineering, vol. 35, no. 3, **(2009)**, pp. 517–526.
[2] P. Sangkatsanee, N. Wattanapongsakorn, and C. Charnsripinyo, “Practical real-time intrusion detection using machine learning approaches,” Computer Communications, vol. 34, no. 18, **(2011)**, pp. 2227–2235.
[3] W. Stallings, Network security essentials: applications and standards. Pearson Education India, **(2007)**.
[4] S. Smith, Trusted computing platforms: design and applications. Springer, **(2013)**.
[5] Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, “An efficient intrusion detection system based on support vector machines and gradually feature removal method,” Expert Systems with Applications, vol. 39, no. 1, **(2012)**, pp. 424–430.
[6] A. Shostack, Threat modeling: Designing for security. John Wiley & Sons, **(2014)**.
[7] J. J. Davis and A. J. Clark, “Data preprocessing for anomaly based network intrusion detection: A review,” Computers & Security, vol. 30, no. 6, **(2011)**, pp.353–375.
[8] C. H. Rowland, “Intrusion detection system,”uS Patent 6,405,318, **(2002)**, Jun.11.
[9] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, “A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems,” Expert Systems with Applications, vol. 42, no. 5, **(2015)**, pp. 2670–2679.

[10] S. Axelsson, "Intrusion detection systems: A survey and taxonomy," Technical report, Tech. Rep., **(2000)**.
[11] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," Journal of Network and Computer Applications, vol. 36, no. 1, **(2013)**, pp. 16–24.
[12] Y. Hamid, M. Sug umaran, and V. Balasaraswathi, "Ids using machine learning-current state of art and future directions," British Journal of Applied Science & Technology, vol. 15, no. 3, **(2016)**.
[13] W.-C. Lin, S.-W. Ke, and C.-F. Tsai, "Cann: An intrusion detection system based on combining cluster centers and nearest neighbors," Knowledgebased systems, vol. 78, **(2015)**,pp. 13–21.
[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, **(2002)**, pp. 321–357.
[15] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," in Advances in neural information processing systems, **(2002)**, pp. 833–840.
[16] H. Hotelling, "Analysis of a complex of statistical variables into principal components." Journal of educational psychology, vol. 24, no. 6, **(1933)**, p. 417.
[17] Z. Zhang and H. Shen, "Application of online-training svms for real-time intrusion detection with different considerations," Computer Communications, vol. 28, no. 12, **(2005)**, pp. 1428–1442.
[18] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," Expert Systems with Applications, vol. 41, no. 4, **(2014)**, pp. 1690–1700.
[19] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "Highdimensional and large-scale anomaly detection using a linear one-class svm with deep learning," Pattern Recognition, vol. 58, **(2016)**, pp. 121–134.
[20] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid kpca and svm with ga model for intrusion detection," Applied Soft Computing, vol. 18, **(2014)**, pp. 178– 184.
[21] W. Feng, Q. Zhang, G. Hu, and J. X. Huang, "Mining network data for intrusion detection through combining svms with ant colony networks," Future Generation Computer Systems, vol. 37, **(2014)**, pp. 127–140.
[22] X.-S. Gan, J.-S. Duanmu, J.-F. Wang, and W. Cong, "Anomaly intrusion detection based on pls feature extraction and core vector machine," Knowledge-Based Systems, vol. 40, **(2013)**, pp. 1–6.
[23] Y. Hamid, M. Sugumaran, and L. Journaux, "A fusion of feature extraction and feature selection technique for network intrusion detection," International Journal of Security and Its Applications, vol. 10, no. 8, **(2016)**, pp. 151–158.
[24] X. Xu and X. Wang, "An adaptive network intrusion detection method based on pca and support vector machines," in International Conference on Advanced Data Mining and Applications. Springer, **(2005)**, pp. 696–703.
[25] S. Chebrolu, A. Abraham, and J. P. Thomas, "Feature deduction and ensemble design of intrusion detection systems," Computers & Security, vol. 24, no. 4, **(2005)**, pp. 295–307.
[26] S. Ginsburg, S. Ali, G. Lee, A. Basavanhally, and A. Madabhushi, "Variable importance in nonlinear kernels (vink): Classification of digitized histopathology," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, **(2013)**, pp. 238–245.
[27] Y. J. Fan and C. Kamath, "On the selection of dimension reduction techniques for scientific applications," in Real World Data Mining Applications. Springer, **(2015)**, pp. 91–121.
[28] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol. 290, no. 5500, **(2000)**, pp. 2323–2326.
[29] S. K. Dash, S. Rawat, and A. K. Pujari, "Lle on system calls for host based intrusion detection," in 2006 International Conference on Computational Intelligence and Security, vol. 1. IEEE, **(2006)**, pp. 609–612.
[30] R. Garcia, M. Sadiku, and J. Cannady, "Waid: wavelet analysis intrusion detection," in Circuits and Systems, 2002. MWSCAS-2002. The 2002 45th Midwest Symposium on, vol. 3. IEEE, **(2002)**, pp. III–688.
[31] W. Lu and A. A. Ghorbani, "Network anomaly detection based on wavelet analysis," EURASIP Journal on Advances in Signal Processing, vol. **(2009)**, p. 4.
[32] M. Hamdi and N. Boudriga, "Detecting denial-of-service attacks using the wavelet transform," Computer Communications, vol. 30, no. 16, **(2007)**, pp. 3203–3213.
[33] J. Sun, H. Yang, J. Tian, and F. Wu, "Intrusion detection method based on wavelet neural network," in Knowledge Discovery and Data Mining, 2009. WKDD 2009. Second International Workshop on. IEEE, **(2009)**, pp. 851–854.
[34] Y. Zhan, "A wavelet kernel-based support vector machine for communication network intrusion detection," in Advanced Materials Research, vol. 989. Trans Tech Publ, **(2014)**, pp. 4474–4477.
[35] I. Daubechies, "Ten lectures on wavelets, vol. 61 of cbms-nsf regional conference series in applied mathematics," **(1992)**.
[36] Y.-W. Chang, C.-J. Hsieh, K.-W. Chang, M. Ringgaard, and C.-J. Lin, "Training and testing low-degree polynomial data mappings via linear svm," Journal of Machine Learning Research, vol. 11, no. Apr, **(2010)**, pp. 1471–1490.

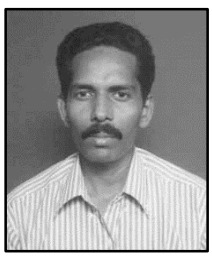
[37] S. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. Chan, "Costbased modeling for fraud and intrusion detection: Results from the jam project, in'darpa information survivability conference and exposition, 2000. discex'00," in Proceedings, **(2000)**, pp. 130–144.
[38] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. R. Kendall, D. Mc-Clung, D. Weber, S. E. Webster, D. Wyschogrod, R. K. Cunningham et al., "Evaluating intrusion detection systems: The 1998 darpa off-line intrusion detection evaluation," in DARPA Information Survivability Conference and Exposition, 2000. DISCEX'00. Proceedings, vol. 2. IEEE, **(2000)**, pp. 12–26.
[39] A. N. Toosi and M. Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuro-fuzzy classifiers," Computer communications, vol. 30, no. 10, **(2007)**, pp. 2201–2212.
[41] S.-S. Wang, K.-Q. Yan, S.-C. Wang, and C.-W. Liu, "An integrated intrusion detection system for cluster-based wireless sensor networks," Expert Systems with Applications, vol. 38, no. 12, **(2011)**, pp. 15 234–15 243.
[40] C. Xiang, P. C. Yong, and L. S. Meng, "Design of multiple-level hybrid classifier for intrusion detection system using bayesian clustering and decision trees," Pattern Recognition Letters, vol. 29, no. 7, **(2008)**,pp. 918–924.

Authors

Yasir Hamid, received his Master's degree in Computer Applications from University of Kashmir in the year 2014. He is currently as a Ph.D Scholar in Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry. His areas of interests are Machine Learning, Network Security, Non Linear Dimension Reduction and data visualization.

Ludovic Journaux, received his PhD in image processing and computer sciences from the University of Burgundy (France) in 2006. He is currently working as associate professor at Agrosup Dijon and is a member of LE2I laboratory (UMR 6306): Laboratory of Electronics, Computer Sciences, and Images. His research interests include image processing, data mining, statistical analysis, artificial intelligence and classification.

Firdous A. Shah is Assistant Professor, University of Kashmir, India, He received his B. Sc. and M. Sc. degrees from University of Kashmir in 2000 and 2002, respectively. He received his Doctorate degree in Mathematics from the Department of Mathematics, Jamia Millia Islamia, New Delhi, India in 2007. His primary research interests include basic theory of wavelets, wavelet packets and application of wavelets in financial time series.



M. Sugumaran, received his M.Sc degree in mathematics from University of Madras in 1986 and M.Tech degree in computer science and data processing from Indian Institute of Technology, Kharagpur, India in 1991, and obtained his Ph.D from Anna University, Chennai in 2008. He is currently working as Professor and Head of Computer Science and Engineering at Pondicherry Engineering College, India. His areas of interests are theoretical computer science, analysis of algorithms, parallel and distributed computing, and spatial-temporal data.

