Network Attack Detection Method Based on Their Time - Frequency Decomposition

Dimitris Sklavounos[1], Spiros Chountasis[2], Aloysius Edoh[3]
[1]Metropolitan College, 74 Sorou St, Amarousio 11525 Athens, Greece
[2]Independent Power Transmission Operator, Asklipiou 22, Krioneri, 14568 Athens, Greece
[3]University of East London, 4-6 University Way, London E16 2RD
dsklavounos@mitropolitiko.edu.gr, schountasis@admie.gr, edoh@uel.ac.uk

ABSTRACT

In the last few years, attack detection has become a powerful methodology for network protection and network security measures. The present work presents a new detection scheme for data recorded over the network, applicable on the broad scientific field of information security, including detection and prevention.

The proposed method employs bi-dimensional (time-frequency) data representations of the forms of the Short Time Fourier transform (STFT) as well as the Wigner Distribution (WD). Moreover, the method applies factorization and Singular Value Decomposition (SVD) of these two-dimensional matrices in order to achieve intrusion detection. The current scheme was performed and evaluated for the case of a dataset KDD-NSL and the efficiency and robustness of the procedure is proven experimentally.

KEYWORDS

Root Local (R2L) intrusion detection, Short Time Fourier Transform, Wigner Distribution, Singular Value Decomposition, NSL-KDD Dataset

1 INTRODUCTION

1.1 Attack Mechanism for intrusion types

Nowadays, the increase and complexity of the cyber intrusions is rather obvious as they are causing major problems to the computerized systems. Well known intrusion attack types like DoS or R2L still remain a real threat as their sophisticated techniques have increased the difficulty of being detected. Some of the new approaches include the application of statistical methods, Artificial Intelligence (AI), network simulation tools and big data predictive analytics techniques.

The aforementioned approaches aim at creating effective IDS systems such as misuse, anomaly detection and hybrid systems, that monitor and analyze traffic in computer systems and networks, in order to detect security threats [1].

Although many IDS research works have been carried out, the complexity of current cyber attacks requires hybrid detection approach that has anomaly system with diagnosis capability and misuse systems with prognosis ability.

The proposed scheme is based on the consideration that data can be fully expressed in a bi-dimensional time and frequency (T&F) domain. By examining the “source bytes” attribute of the classified NSL-KDD dataset that way, the method achieved R2L intrusion detection. This pictorial representation of the data contains all the required information for analysis and data handling proceedings. Furthermore, the T&F data decomposition takes place by applying the SVD as a promising matrix factorization technique and the subtraction of the diagonal matrices came out with the satisfactory results of detection.

The proposed scheme is experimentally evaluated, leading to the conclusion that the joint T&F representation of the recorded data together with the SVD may provide a promising technique for network attack detection under various attacks such as remote to local (R2L).

The rest of the paper is organized as follows: In Section 2 the basic theoretical concepts behind our proposed method are briefly discussed. Sub section 2.1 describes the utilized dataset. The detection method based on the T&F representations and matrix factorization and decomposition properties of these interpretations are presented in sub section 2.2. The numerical evaluation and related results of the proposed technique are illustrated in Section 3. Finally, in Section 4 the conclusions and possible extensions of the current work are given.

1.2 Related work

In the works [2], [3] of the same research direction, the NSL-KDD dataset was utilized for the detection of DoS intrusions by focusing on the “source bytes” of UDP and ICMP protocols. The applied control mechanisms (CUSUM and EWMA) successfully achieved detection of these type attacks [2]. Moreover, the source bytes of the TCP protocol were examined for the detection of R2L attacks. The two above detection methods utilized, with the EWMA to be more efficient in immediacy and accuracy [3]. In both works the mean value was calculated using normal network operation (where no attacks involved). The superiority of the EWMA chart was the motivation for its application on the detection of all types of intrusions of the present work. The proposed scheme is based on the consideration that data can be fully expressed in a bi-dimensional i.e., T&F domain. This pictorial representation of the data contains all the recent information for the required for the network analysis and data handling proceedings. Afterwards, it is decomposed using the SVD a well-known matrix factorization. The proposed scheme is experimentally evaluated, leading to the conclusion that the joint T&F representation of the recorded data together with the SVD provide a suitable technique for network attack detection under various attacks.

2 METHODOLOGY

2.1. Description of the Utilized Dataset

The dataset utilized in the present work is the NBL-KDD which consists of 42 attributes. The current version is the newer version of the KDD’99 dataset and it has been filtered so all data duplication of the previous version is removed. The 20% of the entire data set is earmarked as training data and branded as “KDDTrain+_20Percent” with 25192 instances while the reminding 22544 instances are reserved as “KDDtest+”. There are 42 attributes in each version and the 42nd attribute is categorized as ‘class’ which specifies whether the given instance is a normal connection or an attack. The dataset files were downloaded from [6]. This work used the "KDDtrain20percent" dataset for evaluation purpose. For clarity the 42nd attribute in the dataset was named “xattack” which contains a numbering that specifies the type of the attack in numbering form as: (1) stands for DoS, (2) is for inside attackers aka User to Root (U2R), (3) is for Remote to User (R2L), (4) is for Probe and (5) is the normal operation packets.

2.2. Detection Method

Based on the idea of our previous works [2], [3], this method examines the source byte values of the NSL-KDD dataset and aims to achieve detection of R2L attacks. Two statistical techniques were applied: the Cumulative sum as well as the exponential weighted moving average control charts with the aim to detect intrusions based on the same attribute (source bytes) of the NSL-KDD dataset. Both techniques were efficient in cases where the attacking source byte values caused shifts in the mean value of the normal traffic distribution. Most of the R2L attacking source byte values were difficult to be detected, as they were close to the mean value (334 bytes) and thus, they didn't cause any significant diversion of the mean value.

The present work is an attempt of representing information of source bytes in a T & F form by applying STFT and WD methods aiming to examine if there would be a possibility of achieving R2L intrusion detection in the SVD. Hence, SVD has been applied in both WD illustrations [4],[5] of the source bytes of the assumed normal traffic and the source bytes under R2L attacks. Moreover, when subtracting the diagonal matrices Σ of both the above SVDs, a significant difference in the singular values was observed, leading to the conclusion that detection may be achieved under strictly defined conditions related to the substance of the type of traffic (normal and under attack).

2.2.1 Time – Frequency Representation

For the past decades, there has been an alternative development for the study of time-varying spectra. The basic idea was to devise a joint function of time and frequency, as a representation that will describe the energy density or intensity of a signal simultaneously in T&F. The motivation for devising such a time–frequency representation is to (a) find and illustrate the fraction of the energy in a certain frequency and time range, (b) calculate the distribution of frequency at a particular time, and (c) calculate the global and local moments of the distribution such as the mean frequency and its local spread.

The T&F plane corresponds to two orthogonal axes for time and frequency, respectively. For a signal where $s(t)$ is represented along the time axis and $S(\omega)$ is represented along the frequency axis, the Fourier Transform (FT) is an operator that produces a change in the representation of the signal corresponding to a counter clockwise axis rotation equal to $\pi/2$.

The STFT is the result of applying the FT at different points in time on finite length (i.e., short time) sections of a signal. This description is fundamental to signal analysis because it introduces a time dependency, which the FT of thewhole signal does not have. This important T&F analysis tool is frequently used for speech signal processing. The STFT of a signal $s(t)$ is defined as [8]:

$$\mathrm{STFT}(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}s(t')h\,(t'-t)e^{-i\omega t'}dt' \quad (1)$$

where *h(t)* is a suitably chosen analysis window.

The STFT can also be computed from its FTs $S(\omega)$ and $H(\omega)$ as shown in equation (2).

$$\mathrm{STFT}(t,\omega)=\frac{1}{2\pi}e^{-i\omega t}\int_{-\infty}^{\infty}s(\omega')H\,(\omega'-\omega)e^{-i\omega' t}d\omega' \quad (2)$$

The WD is the most widely studied and applied bi-dimensionalsignal representation. It was first introduced in the academic field of quantum mechanics. The WD of the signal *s(t)* and its FT *S(ω)* is defined by Cohen [7] Eq. (3) and Eq. (4)

$$\mathrm{WD}(t,\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}s\left(t+\frac{1}{2}\tau\right)s^{*}\ \left(t-\frac{1}{2}\tau\right)e^{-i\omega\tau}d\tau \quad (3)$$

$$\mathrm{WD}(t,\omega)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}S\left(\omega+\frac{1}{2}v\right)S^{*}\ \left(\omega-\frac{1}{2}v\right)e^{-itv}dv \quad (4)$$

where τ and v are the time and frequency lag, respectively. The asterisk will denote throughout this paper the complex conjugation.

The WD of a signal can be interpreted as a pseudo-energy density of the signal, sinceit is real, covariant to time and frequency domain translations, but is not always positive.

The signal energy in the T&F region can be determined by integrating the distribution over that region [7],[8].

2.2.2 Singular Value Decomposition (SVD)

The SVD is a matrix computation tool with various applications. The main advantages of our method based on the SVD are the fact that when a small perturbation occurs on the T&F plane, larger variation of their diagonal matrices' difference occur.

Moreover, the singular values represent intrinsic algebraic image properties.The following SVD factorization of a $m \times n$ matrix A will be used:

$$A = U\Sigma V^{T} \quad (5)$$

where U and V are orthogonal matrices which satisfy $I_m = U^T U$ and $I_n = V^T V$ where $I_{n,m}$ is the $n \times n$ identity matrix. The matrix Σ is a diagonal matrix whose entries are the singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_N \geq 0$ of A.

3. Evaluation of the Method

The method examines the source byte values as they have been recorded into the NSL-KDD dataset. The representation of the normal (no attacks involved) source byte values in time and frequency domain are depicted in Figures 1 and 2 respectively.

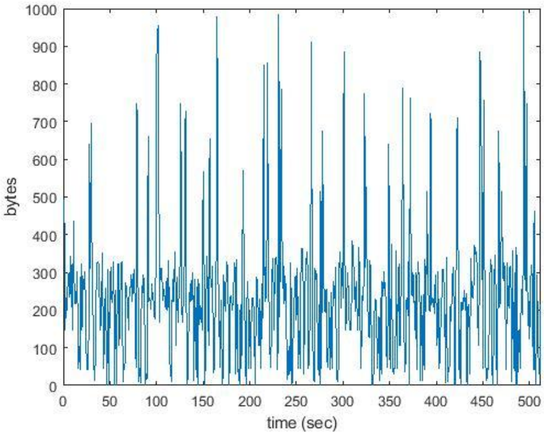

Figure 1. Normal source byte values in time domain

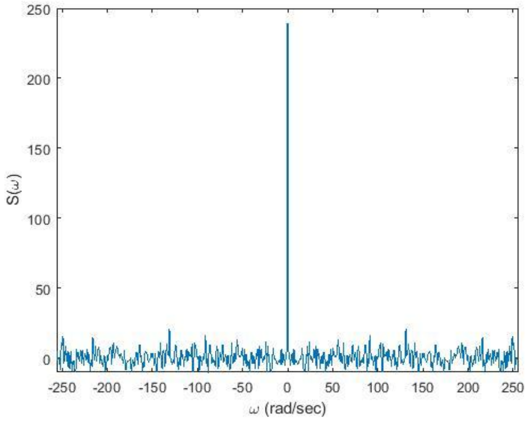

Figure 2. Fourier Transform for Figure 1

A representation of the same form as in Fig. 1 and 2 have the recorded source bytes including R2L attacks.

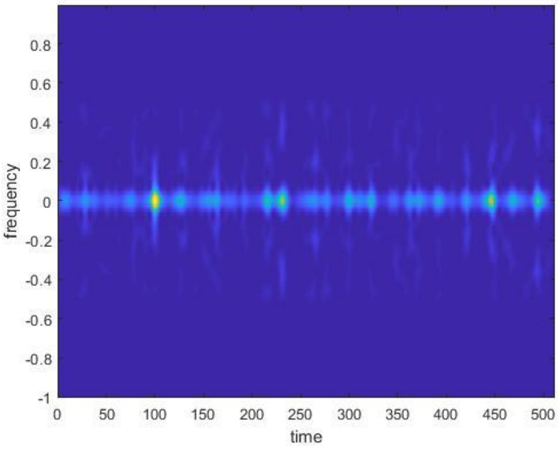

Figure 3. STFT representation for Figure 1

Figures 3 and 4 depict the STFT and the WD respectively and as shown, it is clear that the WD has higher resolution.

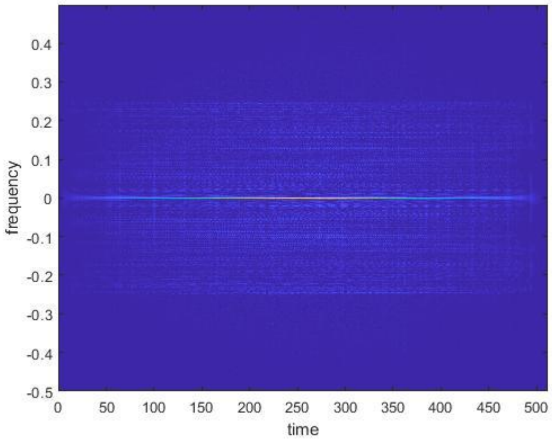

Figure 4. WD representation for Figure 1

A representation of the same form as in Fig. 3 & 4 has the recorded source bytes including R2L attacks.

Our approach was to test the difference from the T&F representations between the SVD of the matrices produced, before and after the attack. The next figures depict the 3-D plots of the normal recorded bytes, the bytes recorded under R2L attack and the difference between the SVD of these matrices i.e, R2L attack and normal bytes.

For the numerical simulations we have chosen the WD as it is a T&F representation that provides a high resolution. As input we consider the diagonal matrices of the SVD produced by the WDs. This measure is illustrated in Figure 6 as a 3D plot, where the x and y axes represent the position of each entry of the 225 x 225 matrix while the z-value is the magnitude of each matrix element. The diagonal structure of the matrix can be clearly seen in this figure. The highest values have been concentrated at the origin and they are rapidly decreased.

In Figure 5 the WD of the source bytes values under R2L attack is illustrated, while in Figure 6 the SVD diagonal matrix (Σ) of above type of source bytes is illustrated.

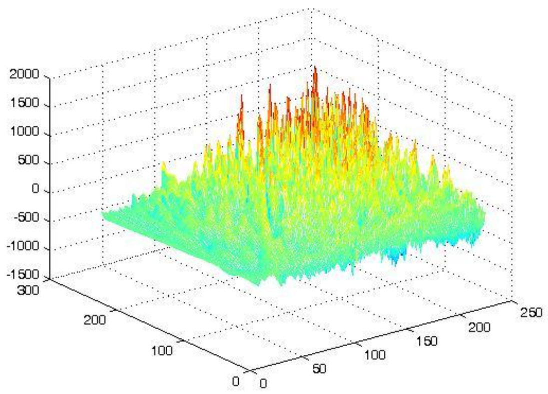

Figure 5. WD of the source bytes values under R2L attack

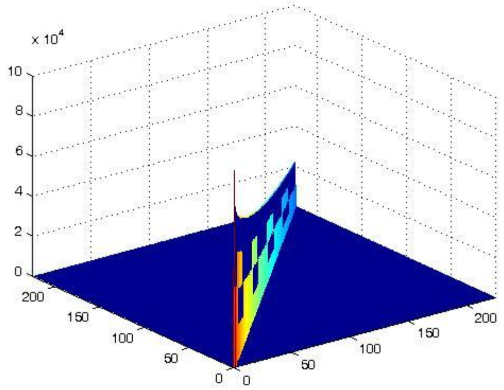

Figure 6. SVD *(Σ)* matrix of the source bytes under R2L attacks

Figure 7 depicts the 3D representation of the difference of the SVD diagonal matrices (Σ) of the normal traffic bytes and the those under R2L attacks. Figure 8 depicts the 2D form of the SVD diagonal matrices of the normal (red) as well as the R2L attacking WDs. Figure 9 depicts the SVD (Σ) matrix difference of figure 8.

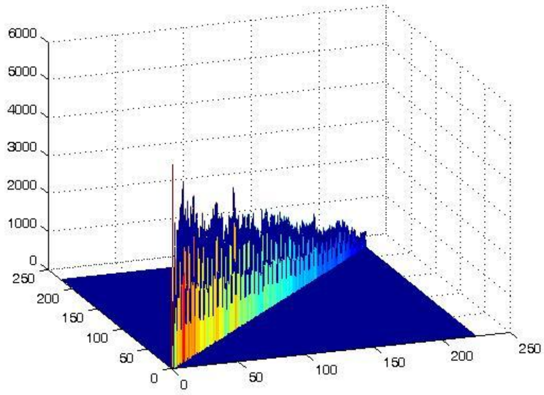

Figure 7. SVD *(Σ)* matrix difference of the normal and R2L attacking source bytes.

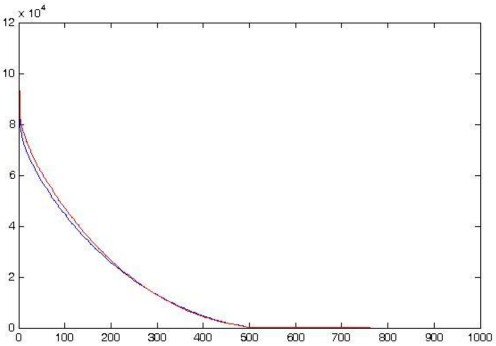

Figure 8. 2D illustration of the SVD (Σ) matrices of the normal (red) and the attacking (blue) source bytes

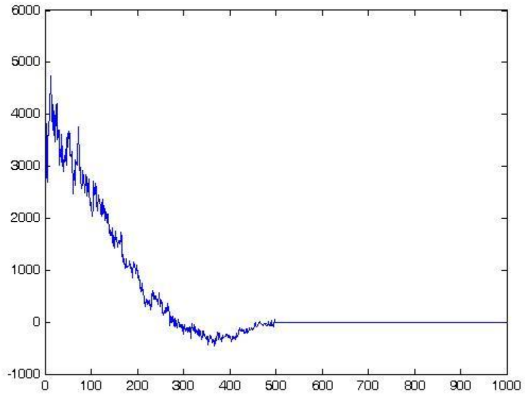

Figure 9. Difference of the (Σ) matrices of Fig.8

As depicted in figures 8 and 9, significant difference in the SVD (Σ) matrices is observed. This observation may provide a strong element for R2L intrusion detection.

4. Conclusion and future work

The intent of this paper was to derive new insights into the network detection process for protection purposes in order to analyze an effective practical networking design. That is accomplished, in general, through the development of a framework for a time–frequency representation. This work introduces a new scheme for detecting various attacks on the matrix decomposed domain that is closely related to the quadratic distributions such as the STFT and WD. The decomposition proposed in this work is the SVD. The proposed scheme maintains the advantages of time and frequency domain data representations.

The experimental evaluation on well-defined dataset shows that the detection ability of the proposed method on network attack is fairly feasible. The future prospects of this work include the more accurate determination of the normal network traffic as opposed to the traffic with attacks in the T & F manner, as well as forming a flash detection system adapted to the intensity distribution, and investigate the robustness of various network attacks.

5 REFERENCES

[1] Hovav, Anat, and John D'Arcy, "The impact of denial-of-service attack announcements on the market value of firms", Risk Management and Insurance Review 6.2 (2003):97-121.

[2] Dimitris Sklavounos, Aloysius Edoh, George Paraskevopoulos, "Utilization of Statistical Control Charts for DoS Network Intrusion Detection",International Journal of Cyber-Security and Digital Forensics (IJCSDF) 7(2): 166-174

[3] Dimitris Sklavounos, Aloysius Edoh, Markos Plytas, "A Statistical Approach Based on EWMA and CUSUM Control Charts for R2L Intrusion Detection", 978-1-5386-2143-1/17 $31.00 © 2017 IEEE, DOI 10.1109/CCC.2017.15

[4] Spiros Chountasis, Dimitrios Pappas, Vasilios N. Katsikis, "Signal watermarking in bi-dimensional representations using matrix factorizations",Comp. Appl. Math.,DOI 10.1007/s40314-015-0230-7

[5] Stamatis Mastromichalakis, Spiros Chountasis and Michalis A. Savelonas "Image representation of a signal for a fractional Fourier transform watermarking scheme," Journal of Information and Optimization Sciences

[6] https://github.com/FransHBotes/NSLKDD- Process Mag 9(2):21–67 Dataset, (10/7/2016)

[7] Cohen L. Time-frequency distributions a review. Proc IEEE 77: 941-981, 1989.

[8] Hlawatsch F, Boudreaux-Bartels GF, Linear and quadratic time-frequency signal representations, IEEE Signal Processing Magazine, 9(2), 1992.