American Journal of Applied Sciences 11 (8): 1405-1411, 2014
ISSN: 1546-9239

doi:10.3844/ajassp.2014.1405.1411 Published Online 11 (8) 2014 (http://www.thescipub.com/ajas.toc)

MULTI SCALE TIME SERIES PREDICTION FOR INTRUSION DETECTION

[1]Palanivel, G. and [2]K. Duraiswamy

[1]Department of Electronics and Communication Engineering,
University College of Engineering-Pattukottai, Rajamadam 614701, India
[2]Department of Computer Science and Engineering,
K.S.Rangasamy College of Technology, Tiruchengode-637209, India

Received 2013-11-28; Revised 2013-12-20; Accepted 2013-06-25

ABSTRACT

We propose an anomaly-based network intrusion detection system, which analyzes traffic features to detect anomalies. The proposed system can be used both in online as well as off-line mode for detecting deviations from the expected behavior. Although our approach uses network packet or flow data, it is general enough to be adaptable for use with any other network variable, which may be used as a signal for anomaly detection. It differs from most existing approaches in its use of wavelet transform for generating different time scales for a signal and using these scales as an input to a two-stage neural network predictor. The predictor predicts the expected signal value and labels considerable deviations from this value as anomalies. The primary contribution of our work would be to empirically evaluate the effectiveness of multi resolution analysis as an input to neural network prediction engine specifically for the purpose of intrusion detection. The role of Intrusion Detection Systems (IDSs), as special-purpose devices to detect anomalies and attacks in a network, is becoming more important. First, anomaly-based methods cannot achieve an outstanding performance without a comprehensive labeled and up-to-date training set with all different attack types, which is very costly and time-consuming to create if not impossible. Second, efficient and effective fusion of several detection technologies becomes a big challenge for building an operational hybrid intrusion detection system.

Keyword: Anomaly Detection, Two-Stage Neural Network Predictor, Multi-Resolution Analysis

1. INTRODUCTION

Network security incidents or intrusions reported by Computer Emergency Response Team-Coordination Center (CERT-CC) have risen at an exponential rate over the past decade. This has led to an increased need for effective intrusion detection and prevention systems to counter the threat of cyber attacks.

Most intrusion detection systems in use today are pattern matching or signature based systems, which depend on a large set of signatures to identify known attacks. Though these systems have a low false positive rate, they have limited efficacy in detection of novel attacks and simple mutations of known attacks. Moreover, writing effective signatures is an exacting task dependent on the expertise of the creator. Due to these limitations, contemporary research efforts have focused more on anomaly-based detection approaches.

An anomaly-based intrusion detection system learns the normal behavioral pattern and flags any deviations from this behavior as an anomaly. Such systems are effective against zero-day attacks, but are prone to a high rate of false positives. Over the years, researchers have turned out of various techniques, some of which are discussed below.

Corresponding Author: Palanivel, G., Department of Electronics and Communication Engineering,
University College of Engineering-Pattukottai, Rajamadam 614701, India

1.1. Rule Based

Mahoney and Chan (2003) present an algorithm for learning rules of normal behavior by mining network data to find precedent consequent pairs of general rules from randomly selected samples. Another approach presented by (Gomez and Dasgupta, 2002) (1) uses genetic algorithms for deriving fuzzy rules for classifying normal and abnormal behavior in networks.

1.2. Clustering

Application of unsupervised clustering te chnique s f or anomaly detection is discussed in (Guan *et al*., 2003; Portnoy *et al*., 2001) Most of these works use packet header data but another technique, described in (Zanero and Savaresi, 2004), uses a two stage anomaly detector in which the first stage involves packet clustering and the second stage involves time correlated anomaly detection. Though this approach appears more effective than simple packet clustering, it is unlikely to be feasible for real-time intrusion detection. Also, some of these techniques assume the availability of labeled network data for training, which is hard to obtain in practice.

1.3. Signal Processing

Several signal processing based intrusion and anomaly detection approaches have been proposed. (Cheng *et al*., 2002) present spectral analysis technique based on the premise that normal TCP flows must exhibit periodicity Inc packet transport associated with Round Trip Times (RTT).

1.4. Prediction

The work by (Brutlag, 2000) describes the use of exponential smoothing using Holt Winters Algorithm for prediction of network traffic

2. WORK STATEMENT

A method for intrusion detection, which treats unexpected observations as anomalies, is suggested. To define the expected behavior our approach uses wavelet decomposed multi-scale time series signals as input to a two stage neural network for prediction of what value the observed variable should take according to past observations. If the predicted value and observed value differ significantly, an anomaly is detected. While wavelet processing has been employed in intrusion detection (Barford *et al*., 2002; Kim *et al*., 2004; Zanero and Savaresi, 2004) and prediction (Brutlag, 2000) has also been used, a combination of these techniques for network intrusion detection is novel.

3. APPROACH AND METHODS

Time series data is a sequence of observation values ordered in time. An active field of research in time series analysis is forecasting the future values based on past observations. There are many different models used for time series prediction. They can be broadly classified as linear and non-linear. Linear models include Moving Average (MA), Auto Regression (AR) and Auto Regressive Moving Average (ARMA). These models give good predictions only for stationary time series, i.e., series in which the mean, variance and auto covariance is independent of time. Research in internet traffic measurements (Paxson and Floyd, 1995) shows that aggregate packet arrival rate is a no stationary process and therefore cannot be appropriately modeled by linear prediction approaches.

Neural network predictors are one of the most powerful and well established non-linear pre- diction models. Hence, they are a strong candidate for internet traffic prediction. The choice of multi-layer feed-forward networks was inspired by the fact that they have been proved to be universal function approximates in (Hornik *et al*., 1998) i.e., they can approximate any arbitrary function if sufficient hidden layer neurons are provided. We may also consider using other neural network architectures like Support Vector machines which are also known to perform well for such problems.

Discrete Wavelet transform is an approach for multi-resolution analysis of signals which has gained wide spread popularity due to its effectiveness as a noise reduction, data compression and hidden feature extraction model. The **Fig. 1** below illustrates the process of wavelet transform. Discrete Wavelet Transform (DWT) is a tool, which provides the time-frequency representation of a signal. The need for DWT arises from a weakness of the Discrete Fourier Transform (DFT). The DFT converts a time signal into a frequency signal but in the process all temporal information is lost. In other words, if a change occurs in the signal, the flourier transform of the signal may show that change but would not give any information about when that change happened. For stationary signals, this is not a major problem. However, most interesting signals contain numerous non stationary or transitory characteristics: Drift trends, abrupt changes and beginnings and ends of events. These characteristics are often the most important part of the signal and Fourier analysis cannot detect them. To get over this problem, the Fourier transform was adapted to analyze only a small section of the signal at a time-a process, often referred to, as windowing the signal.

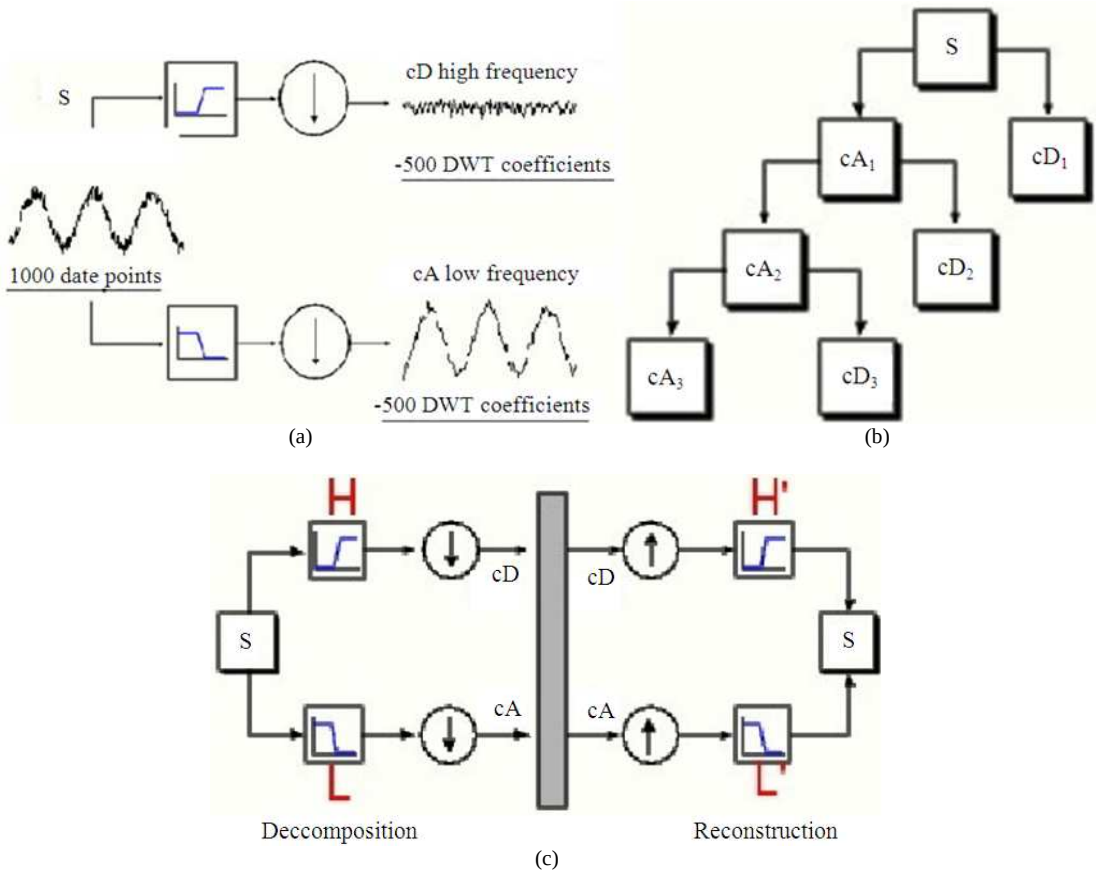


Fig. 1. (a) Wavelet Transform (b) One Level of Decomposition (c) Iterative Decomposition

This technique called Short-Time Fourier Transform (STFT), maps a signal into a two-dimensional function of time and frequency. The drawback of STFT is that once a particular size for the time window is chosen, it the same for all frequencies. Many signals require a more flexible approach-one in which, the window size can be varied to determine either time or frequency more accurately. Wavelet analysis represents this next logical step: A windowing technique with variable-sized regions. Wavelet analysis allows the use of long time intervals where more precise low-frequency information is required and shorter regions where high frequency information is sought. Wavelet analysis is capable of revealing aspects of data that other signal analysis techniques miss viz. trends, breakdown points, discontinuities in higher derivatives and self-similarity. Furthermore, wavelet analysis can often compress or de-noise a signal without appreciable degradation.

A signal gets its 'nature' from the low frequency part and its 'nuances' from high frequency part. Hence, the low-frequency parts are called approximations and high-frequency parts are called details. **Figure 1** illustrates the process of wavelet transform. It shows an input signal, which is passed through a low pass filter followed by a down-sampler to obtain its approximation. The signal is also passed through a high pass filter followed by a down-sampler to obtain its detail. The down sampling is done to remove the redundancy introduced by the filtering process. In practice, this filtering process is iteratively repeated with the approximation of the n^{th} stage becoming the input signal for the $n+1^{th}$ stage. The set of filters form, what is called filter bank and lead to a Multi-Resolution analysis of the signal. The process of inverse DWT is the exact opposite of DWT and requires up sampling and complementary filtering synthesizing the original signal. Usually, the DWT is followed by decimation (or setting to zero) of coefficients of the scale, which are not of interest, so that when the signal is synthesized, only frequencies of interest are present in the signal.

To develop a scheme which can reduce the cost of computational overhead (Begum and Purusothaman, 2011), number of messages needed during the time of key refreshing and the number of keys stored in servers and members. The cost of establishing the key and renewal is proportionate to the size of the group and subsequently fetches a bottleneck performance in achieving scalability. By using a Cluster Based Hierarchical Key Distribution Protocol, the load of key management can be shared among dummy nodes of a cluster without revealing the group messages to them.

The proposed research adapts a framework suggested in (Geva, 1998) for predicting the expected value of a signal derived from a packet or flow variable. It is shown in (Geva, 1998) that the use of wavelet decomposed signal produces superior prediction results than directly using the input signal for prediction of sun spots.

Our approach is to use a similar frame work but explicitly provide an additional input having temporal information. This input would carry information such as time of day, day of week and week of month to aid the neural network in learning cyclical patterns and modulating its prediction with respect to long term historical data not provided by the current signal.

A major difference between our application and usual applications of neural network based non- linear time series prediction is the presence of attack data, which modifies the current time series. Other applications rule out such data as noise. For our application, noise consists of the minor variations in the signal due to inherently bursty nature of network traffic and intrusions should not be considered as noise. There are two approaches t o deal with this problem. One approach is to let the predictor change its prediction due to the change induced in the time series by the intrusion. Since the intrusion is expected to last only a short time the end of intrusion will again cause another change when the behavior returns to normal. The intuition behind this approach is that an unexpected but legitimate change like introduction of new servers leading to increased network load would cause just one transition, unlike an intrusion which, would cause at least two unexpected changes and therefore, a quick succession of multiple anomalies may be used as an indication of an attack. However, this approach could lead to high false negative rate, especially in cases of automated scripts used for attacking a network on a regular and cyclical basis. In this case, the predictor would learn this attack as predictable normal behavior and will not raise alerts. The other approach is to use statistical measures like the prediction error sigma also known as the smoothed standard deviation of prediction error defined by the Equation 1. The Equation 2 describes the initial value given to σ for m observations:

$$\sigma p,t = \sqrt{\alpha \in_t^2 + (1 - \alpha)\sigma_{p,t}^2} \quad (1)$$

$$\sigma p,0 = \sqrt{\frac{\sum_{i=1}^{m} \in^2}{m}} \quad (2)$$

Where:

$\sigma p, t$ = He smoothed standard deviation of predicted error
ϵt = The one step prediction error, i.e.,
$\epsilon t = xt\text{-}x\hat{}t$, α = Error weighting factor

In case the error is much higher than prediction error sigma, the system would raise an alert and in place of the real data, the predictor would use its own prediction with a small increment or decrement to compensate for this change. Otherwise, the change would be considered as a legitimate one and the neural predictor would consider it as one of the inputs for further predictions.

The framework is illustrated in the **Fig. 2** below. It consists of multiple feed-forward neural nets, one for each level of wavelet decomposition. The primary advantage of using wavelet decomposition is to allow the neural network to learn the correct weight that each level (i.e., long term (of the order of many hours) to short term (of the order of few seconds) should be given in order to most accurately predict the expected time series signal. At the same time the input for temporal information helps the network to identify longer cyclic trends.

4. APPROPRIATE RESULTS

Network traffic shows a predictable time cycle. To justify this, shown below is data obtained from MIT Lincoln Labs 1999 dataset. The **Fig. 3** shows two days' (Week1 Monday and Tuesday) packet per minute 999counts along with their approximations, which show clear similarities in shape (13). One may argue that the similarity in this data is because the background traffic is synthetically generated. However, shown below **(Fig. 4a and b)** are two continuous days traces obtained from the NLANR Auckland data repository, which show a cyclic similarity across days like the MIT Lincoln Labs dataset.

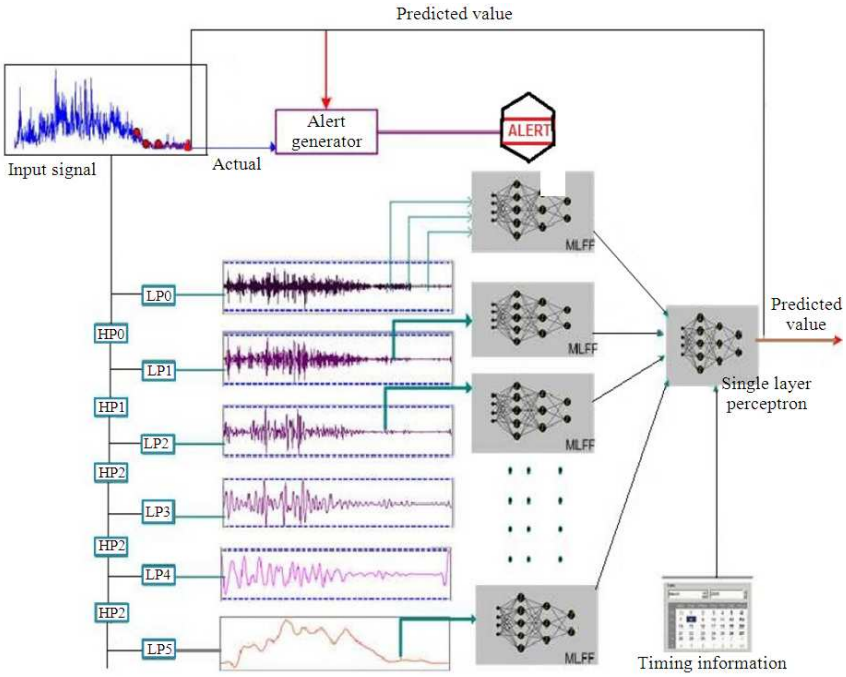


Fig. 2. System Architecture

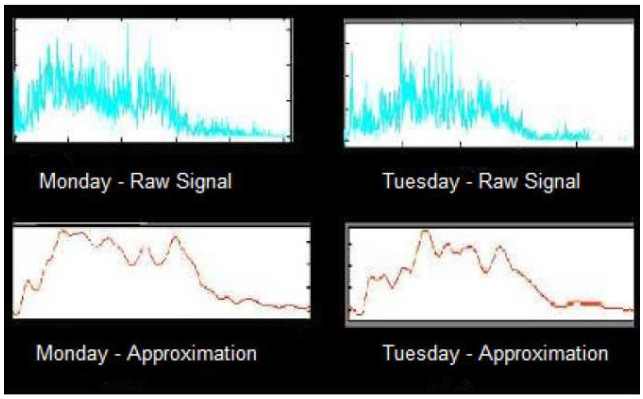
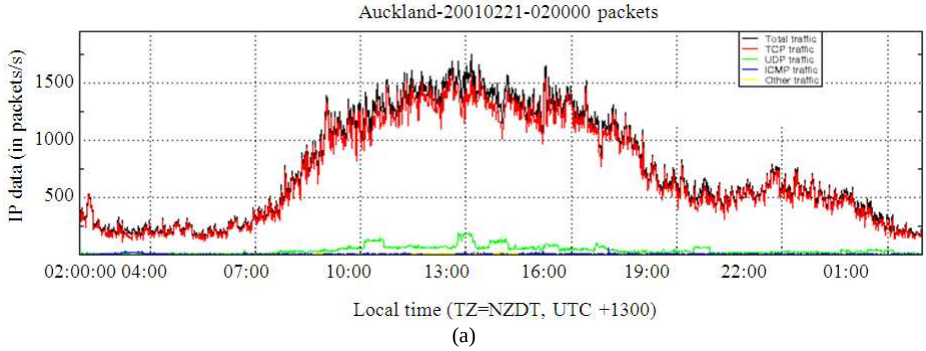


Fig. 3. Two days' Lincoln lab data and respective wavelet approximations



(a)

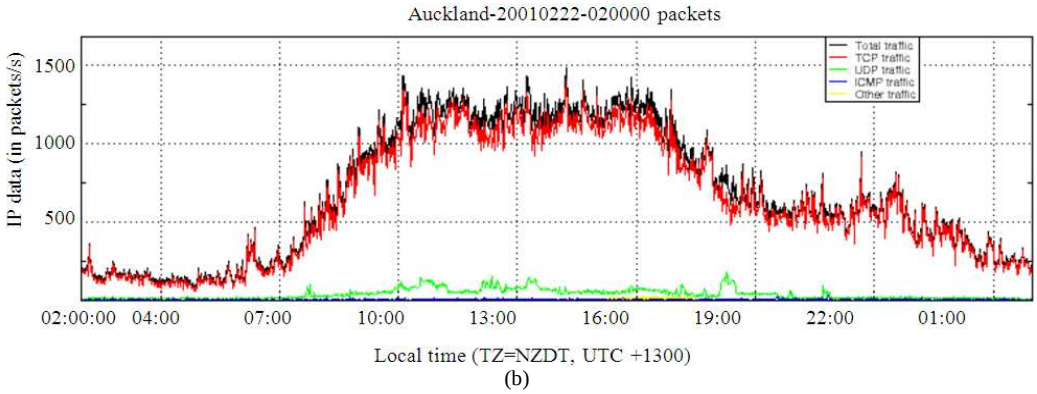


Fig. 4. NLANR Auckland Traces (a) 21st Feb 2001, (b) 22nd Feb 2001

5. CONCLUSION

The effectiveness of multi-resolution analysis as an input to neural network for the purpose of intrusion detection was evaluated. With the help of a predictor, the expected signal value and deviations from this value as anomalies had been predicted. Two approaches have been analyses for this work. First approach is to let the predictor change its prediction due to the change induced in the time series by the intrusion. Second approach is to use statistical measures like the prediction error sigma also known as the smoothed standard deviation was analyzed.

The result of the current work can be used to create a labeling technique for anomaly detection systems. This goal should be easily achieved since it is already known which features are best to detect certain types of attacks and by monitoring their values in parallel with the detection system a method can be proposed to create meaningful attack labels.

6. REFERENCE

Barford, P., J. Kline, D. Plonka and A. Ron, 2002. A signal analysis of network traffic anomalies. Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement, Nov. 06-08, ACM Press, New York, pp: 71-82. DOI: 10.1145/637201.637210

Begum, S.J. and D.T. Purusothaman, 2011. A6 new scalable and reliable cost effective key agreement protocol for secure group communication. J. Comput. Sci., 328-340. DOI: 10.3844/jcssp.2011.328.340

Brutlag, J.D., 2000. Aberrant behavior detection in time series for network monitoring. Proceedings of the Large Installation Conference on System Administration, USENIX Association Berkeley, (AAB' 00), CA, USA, pp: 139-146.

Cheng, C.M., H. Kung and K.S. Tan, 2002. Use of spectral analysis in defense against DoS attacks, Global Telecomm. Conf. GLOBECOM, 3: 2143-2148, DOI: 10.1109/GLOCOM.2002.1189011

Geva, A.B., 1998. ScaleNet-multiscale neural-network architecture for time series prediction, IEEE Trans. Neural Netw., 9: 1471-1482. DOI: 10.1109/72.728396

Gomez, J. and D. Dasgupta, 2002. Evolving fuzzy classifiers for intrusion detection. Proceedings of the IEEE Workshop on Information Assurance, IEEE Press. Jun. 2001, United States Military Academy, West Point, NY, pp: 1-8.

Guan, Y., A. Ghorbani and N. Belacel, 2003. An unsupervised clustering algorithm for intrusion detection. Proceedings of the 16th Conference of the Canadian Society for Computational Studies of Intelligence (Ontario, Canada), (IOC' 03), Springer-Verlag Berlin, Heidelber, pp: 616-117.

Hornik, K., M. Stinchcombe and H. White, 1989. Multilayer feedforward networks are universal approximators, Neural Netw., 5: 359-366. DOI: 10.1016/0893-6080(89)90020-8

Kim, S.S., A.L. Narasimha Reddy and M. Vannucci, 2004. Detecting Traffic Anomalies through Aggregate Analysis of Packet Header Data. In: Networking 04, Lecture Notes in Computer Science, Mitrou, N., K. Kontovasilis, G.N. Rouskas and I. Iliadis and L. Merakos (Eds.)., Springer Berlin Heidelberg, ISBN-10: 978-3-540-21959-0, pp: 1047-1059.

Mahoney, M.V. and P.K. Chan, 2003. Learning rules for anomaly detection of hostile network traffic. Proceedings of the 3rd IEEE International Conference on Data Mining, Nov. 19-22, IEEE Xplore Press, pp: 601-601. DOI: 10.1109/ICDM.2003.1250987

Paxson, V. and F. Sally, 1995. Wide area traffic: The failure of Poisson modeling. IEEE/ACM Trans. Netw., 3: 226-244. DOI: 10.1109/90.392383

Portnoy, L., E. Eskin and S. Stolfo, 2001. Intrusion detection with unlabeled data using clustering. Proceedings of the ACM Workshop on Data Mining Applied to Security, (DMSA '01), ACM Press, New York, pp: 5-8.

Zanero, S. and M.S. Savaresi, 2004. Unsupervised learning techniques for an intrusion detection system. Proceedings of the ACM Symposium on Applied Computing, Mar. 14-17, ACM Press, New York, pp: 412-419. DOI: 10.1145/967900.967988