SYN Flood Attack Detection Using AR Model

Rania Ghazy*, El-Sayed EL-Rabaie*, Moawad Dessouky*, Nawal El-Fishawy, Fathi Abd El-Samie***

*Dept. of Electronics and Electrical Communications, Faculty of Electronic Engineering, Menoufia University.
**Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University.
(Received: 30 Apr. 2017 – Accepted: 8 Aug. 2018)

Abstract

Due to the sophisticated characteristics of auto-regressive (AR) modeling approach, it finds applications in most anomaly detection processes. This paper extends the concept of AR modeling to create models for the estimated auto-correlation between data and control planes packet counts of the network traffic. These models are fed with the anomaly traffic containing SYN flood attack. The estimated residuals in these scenarios are used as indicators for the attacks. Simulation results revealed the success of attack detection using the proposed approach.
Keywords: auto-regressive, control planes and network traffic.

1. Introduction

Intrusion detection is a very interesting research field due to its importance to secure the performance of networks. Different types of attacks exist, such as the SYN flood attack which is one form of denial-of-service attack. In the SYN attack, a flooding process of control packets is performed on the network in an attempt to consume a large amount of resources, which prevents customers from receiving services [1,2].

In this approach, the intrusion detection process depends on estimating the correlation between the control and data planes packet counts [3]. Network traffic consists of two planes, data plane and control plane. The control plane is represented by the set, maintain, and terminate connection packets. The data plane is represented by actual transmission data packets [3]. Any suspicious activity in the correlation indicates an anomaly or

intrusion.
It is expected that data modeling techniques such as AR can be used for efficient representation of data sequences. The basic idea of data modeling is to predict the samples values of the discrete sequence available based on the sample history [4]. So, if we can create models for normal activities represented in auto-correlation sequences, the responses of the created models will reflect large variance if fed with correlation estimated with anomaly. This is the basic principle upon which the AR model can be used for anomaly or intrusion detection [4]

This paper is organized as follows. Section 2 includes cross-correlation and AR model. Section 3 covers the proposed approach. Section 4 includes simulated experiments. Section 5 includes conclusion followed by the more relevant references.

2. Cross-Correlation and AR Model

2.1 Cross-correlation

In signal processing, cross-correlation is a measure of similarity of two waveforms as a function of a time-lag applied to one of them which is similar in nature to the convolution of two functions [4,5].
For discrete functions, f and g, the cross-correlation is defined as [4]:

$$f[n] \star g[n] \stackrel{def}{\equiv} \sum_{m=-\infty}^{\infty} f^*[m]g[n+m] \quad (1)$$

In this paper, the cross-correlation is used to measure the similarity between the two traffic groups; control and data planes, whose values are swinging between 0 and 1.

2.2 AR model

Many observed time series exhibit serial autocorrelation; that is, linear association between lagging observations. This suggests that past observations might be used to predict current observations. Having such a distribution has not only become convenient, but also necessary in forecasting and monitoring of data [5,6].

A model that expresses a univariate time series y_n as a linear combination

of past observations y_{n-i} and white noise v_n is referred to as an autoregressive model (AR model) and has the form [5]:

$$y_n = \sum_{i=1}^{m} a_i y_{n-i} + v_n, \quad (2)$$

where m , a_i are the autoregressive order and the autoregressive coefficient (AR coefficient), respectively. We assume that v_n is a white noise that follows a normal distribution with mean 0 and variance σ^2 and is independent of the past time series y_{n-i} . In other words, v_n satisfies $E[v_n] = 0$, $E[v^2_n] = \sigma^2$, $E[v_n v_m] = 0$, for $n \neq m$, and $E[v_n y_m] = 0$, for $n > m$, where E denotes expectation.

Autocovariance Function

Given the time series yn, the autocovariance function ck is defined by $c_k = E[y_n y_{n-k}]$, $k = 0,\pm1, \ldots$, where k is the lag and, for simplicity, the mean of the time series is assumed to be 0. Taking the expectation after multiplying by y_{n-k} on both sides of (2) yields [5]:

$$E[y_n y_{n-k}] = \sum_{i=1}^{m} a_i E[y_{n-i} y_{n-k}] + E[v_n y_{n-k}] \quad (3)$$

Therefore, we obtain the following Yule-Walker equation [5]:

$$c_0 = \sum_{i=1}^{m} a_i c_i + \sigma^2 \quad (4)$$

$$c_k = \sum_{i=1}^{m} a_i c_{k-i} \ , \ k = 1,2, \ldots . , \quad (5)$$

A time series is said to be stationary if the mean and the auto-covariance function exist and are invariant with time. Note that for univariate time series, the auto-covariance function satisfies $C_{-k} = C_k$, Eq. (5) also holds, even if C_{k-i} is replaced by C_{k+i} . This means that the backward model satisfies the same equation, and that given the auto-covariance function, the forward and backward AR models are identical.

Estimation of the AR Model

In order to identify an AR model, it is necessary to determine the order m and estimate the AR coefficients $a_1, \ldots, a_m$ and the variance σ^2 based on the data. Given the time series $y_1, \ldots, y_N$, by computing the sample auto-covariance functions [5]:

$$\hat{c_k} = \frac{1}{N}\sum_{n=k+1}^{N} y_n y_{n-k} \quad (6)$$

and substituting them into (5), we obtain a system of linear equations for the unknown AR coefficients, $a_1, \ldots, a_m$,

$$\begin{bmatrix} \hat{c_0} & \cdots & \hat{c_{m-1}} \\ \vdots & \ddots & \vdots \\ \hat{c_{m-1}} & \cdots & \hat{c_0} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \hat{c_1} \\ \vdots \\ \hat{c_m} \end{bmatrix} \quad (7)$$

By solving this equation, the estimates $\hat{a_i}$ of the AR coefficients are obtained.

Then, from (4), an estimate of the variance σ^2 is obtained as follows [5]:

$$\hat{\sigma}^2 = \hat{c_0} - \sum_{i=1}^{m} \hat{a_i}\, \hat{c_i} \quad (8)$$

The estimates $\hat{a_1}, \ldots, \hat{a_m}$, and $\hat{\sigma}^2$ obtained by this method are referred to as theYule-Walker estimates.

3. PROPOSED APPROACH

The proposed approach is based on using the AR model with the cross–correlation between control and data packet counts for anomaly detection as follows:

1) Estimating the AR model for the correlation sequence extracted from normal traffic.
2) Evaluating the AR model output for the correlation sequence extracted from abnormal traffic.
3) Estimating the difference between the AR model input and output.
4) Applying a threshold for anomaly detection on the difference signal

If the difference is higher than threshold, there is an attack, otherwise there is no attack.

4. EXPERIMENTAL RESULTS

The simulation scenario in this paper is based on estimating the cross-correlation between recorded control and data planes packets count, from a network on a certain period. The cross-correlation is estimated in the presence and absence of attacks. It is observed that in the presence of attacks, peaks are created in the cross-correlation representation as a result of the difference between the control and data planes packets count as

illustrated in figures 1 to 3. The AR model is used to detect these narrow peaks which occur as a result of SYN flood attack presence. Our objective is to detect these peaks. So, we generated an AR model for the normal case and then we used this model for attack detection. The difference between the input and output for diffrenet orders of the AR model are estimated and compared with a threshold as illustrated in figures 4 to 6.

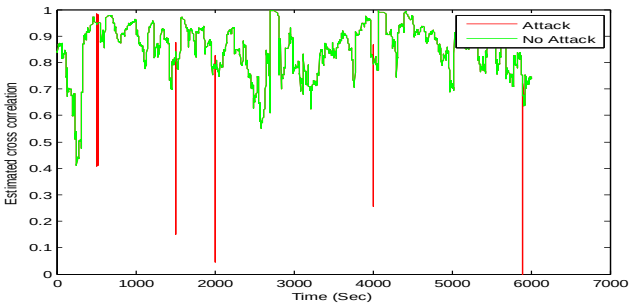

Fig. 1: The estimated correlation coefficients in the absence and presence of SYN flooding attack after applying AR model of second order.

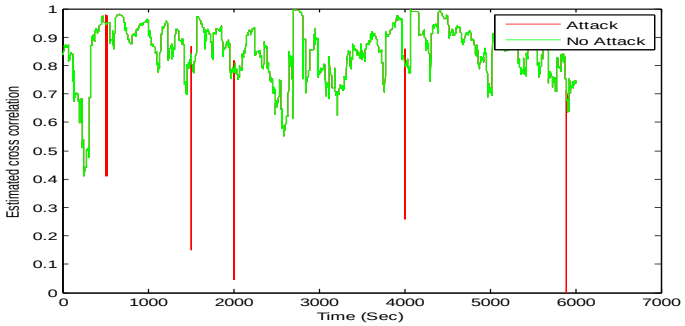

Fig. 2: The estimated correlation coefficients in the absence and presence of SYN flooding attack after applying AR model of third order.

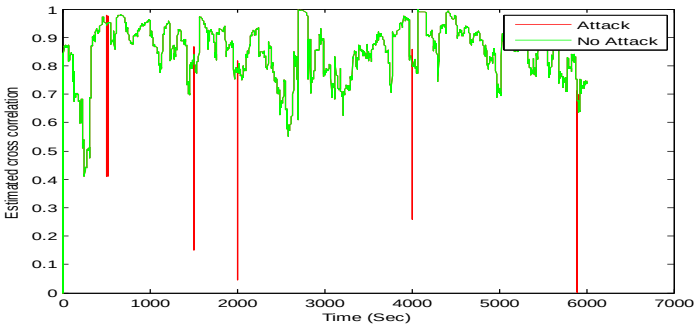

Fig. 3: The estimated correlation coefficients in the absence and presence of SYN flooding attack after applying AR model of fourth order.

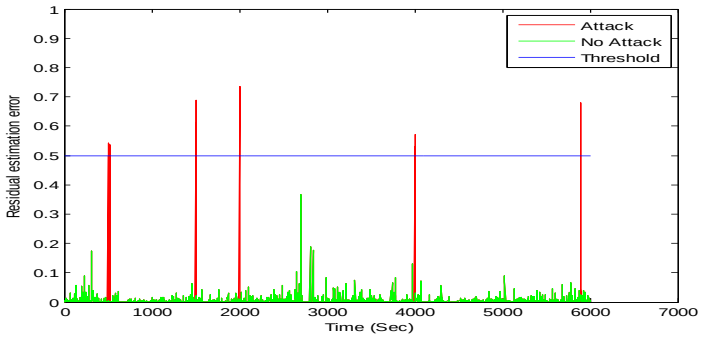

Fig. 4 The residual error for detecting the SYN flooding attack using the AR model of second order

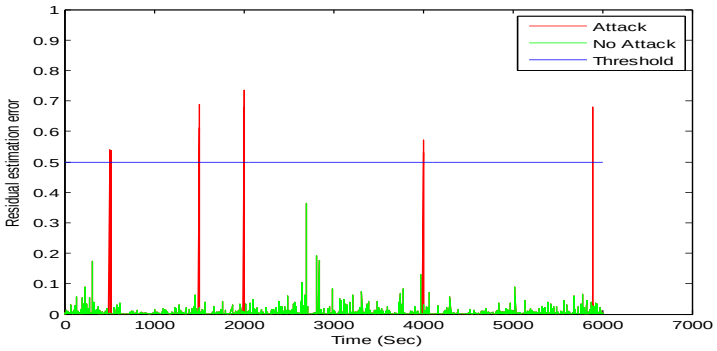

Fig. 5: The residual error in the absence and presence of SYN flooding attack after applying AR model of third order.

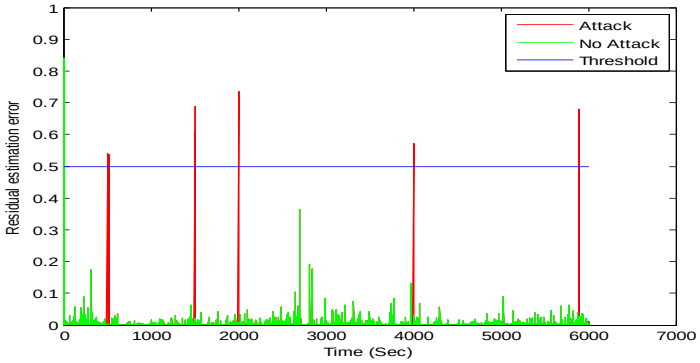

Fig. 6: The residual error in the absence and presence of SYN flooding attack after applying AR model of forth order.

The experimental results show that in the absence of attacks, the residual error values are lower than 0.5, while in the presence of an attack it is higher, and using different orders of AR does not improve the results, but it makes the calculations more complex. The residual error represents the difference between the original and the estimated correlation sequence through the AR modeling. According to these results, decreasing the threshold could increase false alarms, while increasing the threshold could decrease attack detection.

5. Conclusion

This paper presented an approach for detecting SYN flood attack based on applying the AR model on the cross-correlation between control and data packets. This approach adopts a strategy of AR model estimation for the cross-correlation in normal traffic cases, and using the estimated models to detect any abrupt changes that occur in abnormal cases. Variation of the AR model order has been investigated. Simulation results have revealed the success of the AR model approach in the detection of attacks and revealed also that the process of detection can be achieved with orders ranging from second to fourth.

References

[1] M. Manna and A. Amphawan "Review Of Syn-Flooding Attack Detection Mechanism," International Journal of Distributed and Parallel Systems (IJDPS), Vol.3, No.1, January 2012.

[2] https://en.wikipedia.org/wiki/SYN_flood (Access date Dec. 3, 2017)

[3] B. AsSadhan, H. Kim, J. Moura, and X. Wang, "Network Traffic Behavior Analysis by Decomposition into Control and Data Planes," InternationalWorkshop on Security in Systems and Networks (SSN) with conjunction of IEEE International Parallel and Distributed Processing Symposium (IPDPS), Miami, FL, USA, Apr. 18, 2008.

[4] http://en.wikipedia.org/wiki/Cross-correlation (Access date Dec. 3, 2017.

[5] A. Aibinu and J. Salami, A. Shafie , A. Najeeb "Comparing Autoregressive Moving Average (ARMA) coefficients determination using Artificial Neural Networks with other techniques" World Academy of Science, Engineering and Technology, 18, 2008.

[6] http://en.wikipedia.org/wiki/Autoregressive_model (Access date Dec. 3, 2017).

الملخص باللغة العربية

الكشف عن وجود هجوم الفيضان المتزامن عن طريق استخدام نموذج الانحدار الذاتى

بسبب الخصائص المتطورة أو نهج النمذجة التلقائية (AR) ، فإنه يجد التطبيق في معظم عمليات الكشف عن الشذوذ. توسع هذه الورقة من مفهوم نماذج AR لإنشاء نماذج للارتباط التلقائي التقديري بين عدد حزم البيانات ومستويات التحكم لحركة الشبكة. يتم تغذية هذه النماذج مع حركة المرور الشاذة التي تحتوي على هجوم فيضان SYN. يتم استخدام المخلفات المقدرة في هذه السيناريوهات كمؤشرات للهجمات. أظهرت نتائج المحاكاة نجاح الكشف عن الهجوم باستخدام المنهج المقترح.