DDoS Attacks Detection by Means of Statistical Models

Tomasz Andrysiak and Łukasz Saganowski

Abstract In this article we present a network traffic DDoS attacks detection method based on modeling the variability with the use of conditional average and variance in examined time series. Variability predictions of the analyzed network traffic are realized by estimated statistical models ARFIMA and FIGARCH. We propose simple parameter estimation models with the use of maximum likelihood function. The choice of sparingly parameterized form of the models is realized by means of information criteria representing a compromise between brevity of representation and the size of the prediction error. In the described method we propose using statistical relations between predicted and analyzed network traffic in order to detect abnormal behavior possibly being a result of a network attack. Performed experiments confirmed effectiveness of the analyzed method and cogency of the statistical models. *abstract* environment.

Keywords DDoS attacks · Anomaly detection · Statistical models

1 Introduction

At present, the biggest challenge for information systems is providing proper protection against threats. Growing number of attacks, their spreading range, and complexity enforce a dynamic development of network protection systems. This is realized by mechanisms of supervising and monitoring security of computer networks. They are implemented as IDS/IPS Intrusion Detection/Prevention Systems.

T. Andrysiak (✉) · Ł. Saganowski
Institute of Telecommunications, UTP University of Science and Technology,
ul. Kaliskiego 7, 85-789 Bydgoszcz, Poland
e-mail: tomasz.andrysiak@utp.edu.pl

Ł. Saganowski
e-mail: lukasz.saganowski@utp.edu.pl

R. Burduk et al. (eds.), *Proceedings of the 9th International Conference on Computer Recognition Systems CORES 2015*, Advances in Intelligent Systems and Computing 403, DOI 10.1007/978-3-319-26227-7_75

They detect attacks directed onto widely understood network resources of information systems [1]. The techniques used in IDS systems based on statistical methods can be divided into two groups. The first one consists of methods using threshold analysis examining the frequency of events and exceeding their limits in the described time unit. The information about an attack is achieved when the examined units exceed certain threshold values. A crucial drawback of those methods is their susceptibility to errors connected with temporary violent rise in legal network traffic and problems connected with setting reference levels causing an alarm [2]. The second group consists of methods detecting statistical anomalies on the basis of estimated specific parameter profiles of a network traffic. The profiles are characterized by average quantity values, i.e., the number of IP packages, the number of newly dialed connections per time unit, ratio of packages of individual network protocols, etc. It can also be observed that there are some statistical dependences resulting from the part of the day (for instance a greater network traffic strictly after starting work). It is also possible to keep statistics for individual network protocols (for example, quantity ratio of SYN and FIN packages of TCP protocol). IDS systems based on those methods are able to learn a typical network profile—this process lasts from few to several weeks and then compare the current network activity with the memorized profile. The comparison of these two profiles will provide a basis for determining whether there is something disturbance occurring in the network (for instance an attack) [3]. The primary advantage of methods based on anomaly detection is their ability to identify unknown attack types, because they do not depend on information how a particular attack looks like, but on what does not correspond to regular norms of the network traffic. Therefore, IDS/IPS systems detecting anomalies are more effective than systems using signatures in case of identifying new, unknown attack types. Anomaly detection methods have been a topic of numerous surveys and review articles [4]. In works describing the methods there were used techniques consisting in machine learning, neural networks, and expert systems. At present, anomaly detection methods that are particularly intensively developed are those based on statistical models describing the analyzed network traffic. The most often used models are autoregressive ARMA or ARIMA, and Conditional Heteroscedastic Models ARCH and GARCH, which allow to estimate profiles of a normal network traffic [4, 5]. In the present article, we propose using estimation of statistical models ARFIMA and FIGARCH for defined behavior profiles of a given network traffic. The process of anomaly detection (a network attack) is realized by comparison of parameters of a normal behavior (predicted on the basis of tested statistical models) and parameters of real network traffic. This paper is organized as follows. After the introduction, in Sect. 2, the overview of DDoS attacks is presented. In Sect. 3 the ARFIMA and FIGARCH model for data traffic prediction is described in details. Then, in Sect. 4, the anomaly detection system based on ARFIMA—FIGARCH model estimation is shown. Experimental results and conclusion are given thereafter.

2 Overview of DDoS Attacks

Currently, DoS and DDoS attacks have become an important issue of broadly defined IT infrastructure security. Victims of the attacks are often single personal computers as well as supercomputers and vast networks. The outcomes of such activities are experienced by regular Internet users, biggest companies dealing in new technologies that often provide mass services, and powerful governmental organizations of many countries. Despite substantial effort and funds directed to enhancing IT security procedures, at present, we are not able to protect effectively against such attacks.

Attacks such as distributed denial of service (DDoS) use already known techniques of denial of service (DoS) realized with new technology. DoS attack has two crucial restrictions. First, it is performed from a single computer whose Internet connection bandwidth is too low compared to the bandwidth of the victim. Second, while performing the attack from one computer, the attacker may be subjected to a faster detection. Therefore, DoS attack is often conducted on smaller servers containing WWW sites. Attacks on bigger objects, for instance a portal or DNS server, require using a more sophisticated method DDoS, i.e., Distributed Denial of Service, which was created as a response to DoS limitations. The main difference between both methods concerns quantity factor. In DDoS, an attack is performed not from a single computer, but simultaneously from numerous overtaken machines. The sole idea of DDoS attack is therefore simple. However, what constitutes a challenge is its preparation which sometimes lasts many months. The reason is obvious, it is necessary to take over so many computers that will make the attack successful. The period of preparations is the longer, the more powerful are the victims system resources.

Why are the DDoS attacks so dangerous? Most of all, they are difficult to deter due to the fact that their source is greatly distributed. What is worse, the hosts administrators most often do not realize that they are actively participating in the attacks. The statistics are appalling a survey carried out by University of California, San Diego, point that monthly there are performed approximately 15,000 DDoS attacks.

There are a number of methods for conducting a DDoS attack. First, every operational system requires free memory space. If the attacker succeeds in allocating the whole available memory, theoretically, the system will stop functioning, or at least its performance will fall drastically. Such a brutal attack is able to block normal work of even the most efficient IT systems. The second method is based on the use of restrictions of file systems. The third means consists in using malfunctioning network applications or the kernel or errors in the operating system configuration. It is much easier to protect against the above-mentioned kind of attack by proper configuration of such a system. Most of all, it is characteristic for DoS method, which in contrast with DDoS, usually is not based on sending a great number of requests. Errors in TCP/IP stacks of different operational systems constitute an example here. In extreme cases, sending a few packages will be enough to remotely hang the server. The last method is generating a sufficiently big network traffic so that routers or servers cannot handle it [6].

Attacks of this kind are becoming a more and more serious problem. According to quarterly reports published by Prolexic company, within the last 12 months the number of DDoS attacks has risen by 22 %. Campaigns last longer-not 28.5 h as previously, but 34.5 h (a rise by 21 %). The average traffic generated during the attack is approximately 2 GB/s and is more or less 25 % greater than in 2013. The record so far was an attack on the Spamhaus, an organization dedicated to the fight against spam. In March 2013, a hostile network traffic was directed toward servers of that organization with the speed of 300 GB/s. However, according to Arbor company, most of attacks (over 60 %) still do not exceed 1 GB/s. Nevertheless, they still constitute a serious threat [7].

The reason for DDoS attacks being so problematic is that nowadays there are no effective means and methods allowing to protect the IT systems from them. It is only possible to limit the outcomes of those attacks by early identification. One of such solutions is detection of network traffic anomalies that are aftermath of a DDoS attack.

3 Statistical Models for Network Traffic Prediction

Most research on statistical analysis of time series concerns processes characterized by lack of or poor connection between variables which are separated by some time period. Nevertheless, in numerous uses there is a need for modeling processes whose autocorrelation function is slowly decreasing, and the relation between distant observations—even though it is not big—is essential. An interesting approach toward properties of long-memory time series was applying the autoregression with moving average in the process of fractional diversification. As a result, ARFIMA model (Fractional Differenced Noise and Auto Regressive Moving Average) was obtained and implemented by Grange, Joyeux, and Hosking [8, 9]. ARFIMA is a generalization of ARMA and ARIMA models. Another approach describing time series was taking into account the dependence of the conditional variance of the process on its previous values with the use of ARCH model (Autoregressive Conditional Heteroskedastic Model) introduced by Engel [10]. Generalization of this approach was FIGARCH model (Fractionally Integrated GARCH) introduced by Baillie et al. [11].

3.1 ARFIMA Model

The autoregressive fractional integrated moving average model called ARFIMA (p, d, q) is a combination of fractional differenced noise and auto regressive moving average which is proposed by Grange, Joyeux, and Hosking, in order to analyze the long-memory property [8, 9]. The ARFIMA (p, d, q) model for time series y_t is written as

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t, \quad t = 1, 2, \ldots, \Omega, \tag{1}$$

where y_t is the time series, $\varepsilon_t \sim (0, \sigma^2)$ is the white noise process with zero-mean and variances σ^2, $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ is the autoregressive polynomial and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ is the moving average polynomial, L is the backward shift operator, and $(1 - L)^d$ is the fractional differencing operator given by the following binomial expansion:

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k \tag{2}$$

and

$$\binom{d}{k}(-1)^k = \frac{\Gamma(d+1)(-1)^k}{\Gamma(d-k+1)\Gamma(k+1)} = \frac{\Gamma(-d+k)}{\Gamma(-d)\Gamma(k+1)}, \tag{3}$$

where $\Gamma(*)$ denotes the gamma function and d is the number of differences required to give a stationary series and $(1 - L)^d$ is the dth power of the differencing operator. When $d \in (-0, 5\ , 0, 5)$, the $ARFIMA(p, d, q)$ process is stationary, and if $d \in (0\ , 0, 5)$ the process presents long-memory behavior. Forecasting ARFIMA processes are usually carried out using an infinite autoregressive representation of (1), written as $\Pi(L)\, y_t = \varepsilon_t$,

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + \varepsilon_t, \tag{4}$$

where $\Pi(L) = 1 - \pi_1 L - \pi_2 L^2 - \cdots = \Phi\ (L)\ (1 - L)^d\ \Theta\ (L)^{-1}$. In terms of practical implementation, this form needs truncation after k lags, but there is no obvious way of doing it. This truncation problem will also be related to the forecast horizon considered in predictions (see [12]). From (4), it is clear that the forecasting rule will pick up the influence of distant lags, thus capturing their persistent influence. However, if a shift in the process occurs, this means that pre-shift lags will also have some weight on the prediction, which may cause some biases for post-shift horizons [13].

3.2 FIGARCH Model

The model enabling description of long-memory in variance series is FIGARCH (p, d, q) (fractionally integrated GARCH) introduced by Baillie, Bollerslev, and Mikkelsen et al. [11]. The FIGARCH (p, d, q) model for time series y_t can be written as

$$y_t = \mu + \varepsilon_t, \quad t = 1, 2, \ldots, \Omega, \tag{5}$$

$$\varepsilon_t = z_t\sqrt{h_t}, \quad \varepsilon_t|\Theta_{t-1} \sim N\left(0, h_t\right), \tag{6}$$

$$h_t = \alpha_0 + \beta\left(L\right)h_t + \left[1 - \beta\left(L\right) - \left[1 - \phi\left(L\right)\right]\left(1 - L\right)^d\right]\varepsilon_t^2, \tag{7}$$

where z_t is a zero-mean and unit variance process, h_t is a positive time-dependent conditional variance defined as $h_t = E\left(\varepsilon_t^2|\Theta_{t-1}\right)$ and Θ_{t-1} is the information set up to time $t-1$. The FIGARCH (p, d, q) model of the conditional variance can be motivated as ARFIMA model applied to the squared innovations

$$\varphi\left(L\right)\left(1 - L\right)^d\varepsilon_t^2 = \alpha_0 + \left(1 - \beta\left(L\right)\right)\vartheta_t, \quad \vartheta_t = \varepsilon_t^2 - h_t, \tag{8}$$

where $\varphi\left(L\right) = \varphi_1 L - \varphi_2 L^2 - \cdots - \varphi_p L^p$ and $\beta\left(L\right) = \beta_1 L + \beta_2 L^2 + \cdots + \beta_q L^q$ and $\left(1 - \beta\left(L\right)\right)$ have all their roots outside the unit circle, L is the lag operator and $0 < d < 1$ is the fractional integration parameter. If $d = 0$, then FIGARCH model is reduced to GARCH; for $d = 1$ though, it becomes IGARCH model. However, FIGARCH model does not always reduce to GARCH model. If GARCH process is stationary in broader sense, then the influence of current variance on its forecasting values decreases to zero in exponential pace. In IGARCH case, the current variance has indefinite influence on the forecast of conditional variance. For FIGARCH process, the mentioned influence decreases to zero far more slowly than in GARCH process, i.e., according to the hyperbolic function [11, 14]. Rearranging the terms in (8), an alternative representation for the FIGARCH (p, d, q) model may be obtained as

$$\left[1 - \beta\left(L\right)\right]h_t = \alpha_0 + \left[1 - \beta\left(L\right) - \varphi\left(L\right)\right]\left(1 - L\right)^d\varepsilon_t^2. \tag{9}$$

From (10), the conditional variance h_t of y_t is given by

$$h_t = \alpha_0\left[1 - \beta\left(1\right)\right]^{-1} + \lambda\left(L\right)\varepsilon_t^2, \tag{10}$$

where $\lambda\left(L\right) = \lambda_1 L + \lambda_2 L^2 + \cdots$ Of course, for the FIGARCH (p, d, q), for (8) to be well-defined, the conditional variance in the $ARH\left(\infty\right)$ representation in (10) must be non-negative, *i.e.*, $\lambda_k = 0$ for $k = 1, 2, \ldots$ Solving the problem of forecasting using Eq. (10) may be obtained as

$$h_{t+1} = \alpha_0\left[1 - \beta\left(1\right)\right]^{-1} + \lambda_1\varepsilon_t^2 + \lambda_2\varepsilon_{t-1}^2 + \cdots \tag{11}$$

The one-step ahead forecast of h_t is given by

$$h_t\left(1\right) = \alpha_0\left[1 - \beta\left(1\right)\right]^{-1} + \lambda_1\varepsilon_t^2 + \lambda_2\varepsilon_{t-1}^2 + \cdots \tag{12}$$

By analogy, the two-step ahead forecast is given by

$$h_t\left(2\right) = \alpha_0\left[1 - \beta\left(1\right)\right]^{-1} + \lambda_1 h_t\left(1\right) + \lambda_2\varepsilon_t^2 + \cdots \tag{13}$$

In general, the n-step ahead forecast is can be written as

$$h_t(n) = \alpha_0 [1 - \beta(1)]^{-1} + \lambda_1 h_t(n-1) + \cdots + \lambda_{n-1} h_t(1) + \lambda_n \varepsilon_t^2 + \lambda_{n+1} \varepsilon_{t-1}^2 + \cdots \tag{14}$$

In practical application, we stop at a large N and this leads to the forecasting equation

$$h_t(n) \approx \alpha_0 [1 - \beta(1)]^{-1} + \sum_{i=1}^{n-1} \lambda_i h_t(n-i) + \sum_{j=0}^{N} \lambda_{n+j} \varepsilon_{t-j}^2. \tag{15}$$

The parameters will have to be replaced by their corresponding estimates [14].

4 Parameters Estimation and the Choice of Model

The most often used methods of estimation of autoregressive models parameters are: maximum likelihood method (MLE) and quasi-maximum likelihood method (QMLE). This is due to the fact that estimation of the parameters by means of both methods is relatively simple and effective. The basic problem of computing with MLE method is finding a solution to the equation

$$\frac{\partial \ln(L_\Omega(\rho))}{\partial \theta} = 0, \tag{16}$$

where θ is the estimated set of parameters, $L_\Omega(\rho)$ is the likelihood function, and Ω is a number of observations. Mostly, in general case the analytic solution to the Eq. (16) is impossible and then numerical estimation is employed. The basic problem occurring while using the maximum likelihood method is necessity to define the whole model, and consequently the sensitivity of the resulting estimator for any errors in the specification of the AR and MA polynomials responsible for the dynamics of the process [15, 16]. There is no universal criterion for the choice of the model. Usually, the case is as follows: the more complex model, the greater is the value of the likelihood function. As a result, adjusting the model to the data is more effective. However, estimation of a higher number of parameters is connected with bigger errors. Therefore, it is crucial to find a compromise between the quantity of parameters occurring in the model and the value of likelihood function. The choice of the economic form of the model is often based on information criteria such as Akaike (AIC) or Schwarz (SIC). Values of the mentioned criteria can be estimated on the basis of the following formulas:

$$AIC(\rho) = -2 \ln(L_\Omega(\rho)) + 2\rho, \tag{17}$$

$$SIC(\rho) = -2 \ln(L_\Omega(\rho)) + \rho \ln(\Omega), \tag{18}$$

where ρ is the number of the model's parameters. From different forms of the model, the one that is chosen has the smallest information criterion value [12, 17]. In our article, we proposed the maximum likelihood method for parameters estimation and the choice of the form of the model. The method was chosen due to its relative simplicity and computational efficiency. For ARFIMA model, we used HR estimator (described in Haslett and Raftery [18]) and automatic model selection algorithm based on the information criteria (see Hyndman and Khandakar [19]). For FIGARCH model estimation, we used methodology described in the present article [14].

5 Experimental Results

In this section, we presented some results in case of ARFIMA and FIGARH statistical model usage for DDoS attack detection. We simulated real-world DDoS and application specific DDoS attacks for single LAN test network. As a network sensor we used SNORT IDS [20]. SNORT in our case is responsible for traffic capture and extracting network traffic features (see Table 1). Additionally we also used traffic testbed that contains DDoS attacks [21]. Twelve traffic features were used for evaluation of presented ARFIMA and FIGARH statistical models. Obviously, not all traffic features were sufficient for detecting all simulated attacks because they

Table 1 Network traffic features used for experiments

Traffic feature	Traffic feature description
f_1	Number of TCP packets
f_2	In TCP packets
f_3	Out TCP packets
f_4	Number of TCP packets in LAN
f_5	Number of UDP datagrams
f_6	Number of UDP datagrams in LAN
f_7	Number of ICMP packets
f_8	Out ICMP packets
f_9	Number of ICMP packets in LAN
f_{10}	Number of TCP packets with SYN and ACK flags
f_{11}	Out TCP packets (port 80)
f_{12}	In TCP packets (port 80)

Table 2 Detection Rate DR (%) and False Positive FP (%) for a given network traffic features

Traffic feature	FIGARH	ARFIMA	Traffic Feature	FIGARH	ARFIMA
f_1	5.26	8.26	f_1	5.46	4.23
f_2	5.26	12.52	f_2	5.17	4.84
f_3	0.00	12.52	f_3	5.45	4.22
f_4	15.78	10.52	f_4	5.44	4.02
f_5	10.52	14.52	f_5	5.64	4.23
f_6	25.22	35.24	f_6	5.24	4.24
f_7	90.73	98.43	f_7	7.68	6.12
f_8	83.68	96.43	f_8	1.22	0.32
f_9	80.42	85.95	f_9	6.34	4.20
f_{10}	10.52	14.22	f_{10}	5.23	4.56
f_{11}	0.00	8.26	f_{11}	4.58	3.26
f_{12}	0.00	14.22	f_{12}	4.86	3.52

Table 3 Evaluation of proposed method with the use of real world network traffic testbed [21] for 4 days of traffic

Trace date	2008-05-21	2008-08-20	2008-11-15	2009-01-15
ARFIMA DR (%)	85	80	95	82
FIGARH DR (%)	80	70	85	75

Table 4 Evaluation of proposed method with the use of real world network traffic testbed [21] for 4 days of traffic

Trace date	2008-02-20	2008-02-21	2009-05-21	2009-02-15
ARFIMA DR (%)	82	81	92	81
FIGARH DR (%)	79	75	84	77

have not got impact on entire set of traffic features presented in Table 1. In Table 2 we presented detection rate DR and false positive FP values for 12 traffic features. Additionally in Tables 3 and 4, there are results for external testbed for 4 days of network traffic, respectively. We can conclude that ARFIMA model gives us better results in case of DR and FP for the used testbed in our experiments.

6 Conclusion

Cybersecurity of information systems is contemporarily a key research factor. The growing number of DDoS attacks, their expending reach, and the level of complexity stimulate the dynamic development of network defensive systems. The tech-

niques of statistical anomaly detections are recently the most commonly used for monitoring as well as detecting the attacks. In the present article, the construction of statistical autoregressive models, ARFIMA and FIFARCH, has been described. The above-mentioned models present the statistic variability of modeled parameters by means of the average or conditional variance. For estimation of parameters and identification of models the maximum likelihood method together with information criteria were used. As a result of their work the satisfying statistic measurements for researched signals of network traffic were obtained. The process of anomaly (attacks) detecting consist in comparison of estimated behavior parameters with real network traffic factors. The obtained results outstandingly signify that the anomalies included in the network traffic signal can be detected by suggested methods.

References

1. Jackson, K.: Intrusion Detection Systems (IDS). Product Survey. Los Alamos National Library, LA-UR-99-3883 (1999)
2. Esposito, M., Mazzariello, C., Oliviero, F., Romano, S.P., Sansone C.: Evaluating pattern recognition techniques in intrusion detection systems. In: PRIS, pp. 144–153 (2005)
3. Lakhina, A., Crovella, M., Diot, C.H.: Characterization of network-wide anomalies in traffic flows. In: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement, pp. 201–206 (2004)
4. Chondola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**(3), 1–72 (2009)
5. Rodriguez, A., Mozos, M.: Improving network security through traffic log anomaly detection using time series analysis. In: Computational Intelligence in Security for Information Systems, pp. 125–133 (2010)
6. Liang H., Xiaoming B.: Research of DDoS attack mechanism and its defense frame, computer research and development (ICCRD). In: 3rd International Conference, pp. 440–442 (2011)
7. Atak i Obrona 2013 Raport, Ataki i metody obrony w internecie w Polsce (2013)
8. Granger, C.W.J., Joyeux, R.: An introduction to long-memory time series models andfractional differencing. J. Time Ser. Anal. **1**, 15–29 (1980)
9. Hosking, J.: Fractional differencing. Biometrika **68**, 165–176 (1981)
10. Engle, R.: Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica **50**, 987–1008 (1982)
11. Baillie, R., Bollerslev, T., Mikkelsen, H.: Fractionally integrated generalized autoregressive conditional heteroskedasticity. J. Econom. **74**, 3–30 (1996)
12. Crato, N., Ray, B.K.: Model selection and forecasting for long-range dependent processes. J. Forecast. **15**, 107–125 (1996)
13. Gabriel, V.J., Martins, L.F.: On the forecasting ability of ARFIMA models when infrequent breaks occur. Econom. J. **7**, 455–475 (2004)
14. Tayefi, M., Ramanathan, T.V.: An overview of FIGARCH and related time series models. AUSTRIAN J. Stat. **41**(3), 175–196 (2012)
15. Box, G., Jenkins, G., Reinsel, G.: Time Series Analysis. Holden-day, San Francisco (1970)
16. Brockwell, P., Davis, R.: Introduction to Time Series and Forecasting. Springer, Berlin (2002)
17. Beran, J.A.: Statistics for Long-Memory Processes. Chapman and Hall, New York (1994)
18. Haslett, J.: Raftery AE space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). Appl. Stat. **38**(1), 1–50 (1989)
19. Hyndman, R.J., Khandakar, Y.: Automatic time series forecasting: the forecast Package for R. J. Stat. Softw. **27**(3), 1–22 (2008)
20. SNORT—Intrusion Detection System, https://www.snort.org/
21. The CAIDA Dataset, http://www.caida.org/data (2006–2009)