Exposing Deviations in Information Processes using Multifractal Analysis

Yevhen Ivanichenko[1], Valerii Kozachok[1], Yurii Dreis[2], Olena Nesterova[1,3], and Kate Dmytriienko[1]

[1] *Borys Grinchenko Kyiv University, 18/2 Bulvarno-Kudriavska str., Kyiv, 04053, Ukraine*
[2] *Polissia National University, 7 Staryi ave., Zhytomyr, 10008, Ukraine*
[3] *National Pedagogical Dragomanov University, 9 Pyrohova str., Kyiv, 01601, Ukraine*

Abstract

The main requirement for modern systems of intrusion detection is the possibility of identifying deviations in information processes in order to detect unknown attack types. An overview of existing approaches to identifying network deviations based on multifractal analysis methods is given. The results of the calculation of the Hurst exponent for the time series of CPU usage for different types of user activity are presented.

Keywords

Fractal analysis, the Hurst exponent, network deviation detection.

1. Introduction

Signature methods of analysis used in modern intrusion detection systems aimed at identifying known and more specific methods of attacks, appear not to be able to detect their modifications or new types, which makes the use of such systems ineffective. Existing solutions to individual cases of detection of network deviations to this time do not allow to develop a single universal mechanism for detecting previously unknown attack types.

The current task at the moment is to find more effective universal methods for detecting network deviations that are a consequence of technical failures or unauthorized impacts. The main requirement for these methods is the possibility of identifying arbitrary types of intruders, including distributed in time. Statistical studies of network traffic indicate that it has the properties of fractality or self-similarity, as well as the variability of these characteristics in the event of deviations in the network, which allows the use of fractal analysis to detect attacks [1, 2].

The purpose of this study is an overview of modern existing approaches to identifying network deviations based on the method of fractal analysis.

2. Methods of Detecting Attacks

Attacks are deliberate actions of the offender, leading to violation of confidentiality, integrity or accessibility of the system.

Methods of detection of attacks are divided into:

1. Signature.
2. Behavioral.

Signature methods are intended to detect known and clearly described attacks and founded on a reference verification of symbol sequences and events with the database of attack signatures. Advantages of signature methods include low demands for computing power and probability of detection of attacks. Disadvantages of signature methods are the impossibility of identifying new types

CPITS-II-2021: Cybersecurity Providing in Information and Telecommunication Systems, October 26, 2021, Kyiv, Ukraine
EMAIL: y.ivanichenko@kubg.edu.ua (Y. Ivanichenko); v.kozachok@kubg.edu.ua (V. Kozachok); dreisyuri@gmail.com (Y. Dreis); o.d.nesterova@npu.edu.ua (O. Nesterova); k.dmytriienko.asp@kubg.edu.ua (K. Dmytriienko)
ORCID: 0000-0002-6408-443X (Y. Ivanichenko); 0000-0003-0072-2567 (V. Kozachok); 0000-0003-2699-1597 (Y. Dreis); 0000-0002-0402-0370 (O. Nesterova); 0000-0001-7984-7279 (K. Dmytriienko)

of attacks and modifications that exist without a clear formalization of keywords of network traffic and updating the signature database.

Behavioral methods are intended to identify unknown attacks and are based on detection of deviations from normal operation mode. Advantages of behavioral methods comprise the possibility of analyzing the dynamics of processes and the possibility of identifying new types of attacks. Disadvantages of behavioral methods include higher requirements for computing resources and capacities and lower probability of detection.

3. Fractal Analysis

The time series is a sequence of values of the studied magnitude measured at regular intervals.

The central concepts of fractal analysis are fractal dimension (D) and the Hurst exponent (H).

The fractal dimension of the set (according to Hausdorff) is determined by:

$$D = -\lim_{\varepsilon \to \infty} \frac{\lg[N(\varepsilon)]}{\lg(\varepsilon)}, \quad (1)$$

where $N(\varepsilon)$ – the minimum number of non-empty cells ε that cover a given set.

The Hurst exponent characterizes the degree of similarity of the process:

1. $0 < H < 0.5$ – random process, which does not have self-similarity and is characterized by a tendency for average value;
2. $H = 0.5$ – a completely random process without a pronounced tendency;
3. $H > 0.5$ – a trend-resistant process that has a long memory and is self-similar.

The fractal dimension is directly related to the Hurst exponent: $D = 2 - H$.

This ratio is fair when the structure of the curve that describes the fractal function is investigated with high resolution, that is, in the local limits. One of the popular methods of finding fractal dimension is R / S analysis [2]:

$$M\left[\frac{R(n)}{S(n)}\right] \sim cn^H, n \to \infty \quad (2)$$

where n – high resolution; c – a positive finite constant that does not depend on n; H – the Hurst exponent; R(n) – the scope of the time series.

$$R(n) = \max_{1 \le j \le n} \Delta_j - \min_{1 \le j \le n} \Delta_j \quad (3)$$

$$\Delta_j = \sum_{i=n}^{n} x_i - k\bar{x}, k = \overrightarrow{1, n} \quad (4)$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (5)$$

$$S(n) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (6)$$

4. Review of Existing Methods and Approaches

In [1], a method of maximum modules of wavelet transform (MMWT) is used to detect traffic deviations, which allows us to detect the singularity of the signal. The network traffic collected on the boundary router of the university network was taken as the analyzed data. Each sequence is about 24 hours long with a sampling step of 1 second. Samples of "pure" traffic without attacks, and also with various deviations are presented: at DDoS-attacks of different types of scanning. The algorithm for estimating the parameters of the multifractal spectrum is as follows:

The output signal f (t) is decomposed by means of a wavelet transform by the mother wavelet Ψ (t) into the corresponding coefficients:

$$W_f(u, i) = \left(f(t), \Psi_{u,s}(t)\right) = 2^{-\frac{j}{2}} \int \frac{t-u}{2^j} dt \quad (7)$$

The partition function is calculated:

$$S(q, j) = \sum_p \left|W_f(u_p, j\right|^q \quad (8)$$

For each value of $q \in R$ it is necessary to calculate the scale indicator:

$$\tau(q, j) = \log_{j \to 0} \inf \frac{\ln S(q,j)}{\ln 2^j} \quad (9)$$

Then the multifractal spectrum f_l (a), using the Legendre transformation is calculated:

$$f_l(a) = \min_{q \in R} \left[q\left(a + \frac{1}{2}\right) - \tau(q) \right] \quad (10)$$

For each octave j, the multifractal dimension of the order q is calculated:

$$D_{q,j} = \frac{1}{q-1}[q(a(q,j) - f(a(q),j))] \quad (11)$$

When q <0, the value of S (q, j) depends on small maxima of the amplitude | W_f (u_p, j) |, as a result, the calculations may not be stable.

In order to avoid the emergence of false maxima of modules created by computing errors in areas where f is almost constant, wavelet-maxima are combined in a chain to form a maximum curve depending on the scale.

If $\Psi = (-1)^P \theta^{(p)}$, where $\theta = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$ – Gaussian function, then all lines of maxima $u_p(j)$ are determined by curves that are limited by j = 0. Therefore, all maximum lines that do not extend to the smallest scale are deleted when calculating S (q, j).

Formalizing the difference in traffic spectra with certain deviations and without them, it is possible to compare the fractal dimensions D1, correlation dimensions D2 and intervals characterizing the "width" of the Lezhandr spectrum for each of the implementations for each octave of decomposition of j.

Information dimensions of the comparative implementations D1 are distinguished by a small stable value and practically do not depend on the number of levels of sampling. This allows us to conclude that the presence of long-term attacks in the signal and non-predicted activity changes the self-similar nature of traffic, and this property can be used in the future to detect attacks.

In [3] it is proposed to determine the deviations, based on their identity and distribution of "heavy tails". Network deviations may occur as a result of overloads, errors by network devices, DDOS attacks, attempts of unauthorized access. To reduce the impact of the periodicity of network traffic on the estimation of the Hurst exponent, the time series is divided into 24 sets of values.

For each set, a histogram is built for 24 equal intervals. For each group there are a packet number and the average package length for the same time intervals. At the next stage, the Hurst exponent is calculated by the method of periodogram that use the slope of the power spectrum. The Hurst exponent is calculated from the ratio:

$$\beta - 1 = 1 - 2H, \quad (12)$$

where β -the slope of the line on a logarithmic scale.

In practice, you must first analyze traffic in the regular network operation mode during the day. When the deviation detection mode is turned on, at first the Hurst exponent is calculated and compared with the corresponding reference value calculated in normal mode, for each parameter separately.

In the paper [4], an algorithm for detecting deviations based on a discrete stationary wavelet transform and fractal dimension is used. As the first step, a time series filtering with a discrete stationary wavelet transform is performed. This preliminary processing is necessary to increase the accuracy of the proposed method: the main components are allocated, the details are filtered.

The main advantage of the discrete stationary wavelet transform over a classic one is the preservation of the time information of the output signal at each level. In the second step, the time series is bypassed by two adjacent windows R and S. For each window, a fractal dimension is calculated according to the algorithm:

$$FD = \frac{\lg\left(\frac{L}{a}\right)}{\lg\left(\frac{d}{a}\right)} \quad (13)$$

where L – the length of the time series, d – the distance between the first point of the series and the farthest from it, a - the average distance between two adjacent points of the series.

Changes in the statistical parameters of the signal are reflected in the fractal dimension, to account for which the following function is introduced:

$$G_k = |FD_{k+1} - FD_k|, k = 1, \dots, n \quad (14)$$

where n = number of points G.

The third step is the search for local maxima G that exceed a given threshold, which are considered as deviations from normal behavior. The accuracy of the method is significantly affected by the length of the window. For the analyzed window of length l, the energy of the function G_l is calculated as:

$$G_l = \frac{\sum k \left| G_{l_k} \right|^2}{N} \quad (15)$$

The window length is calculated as the minimum of the standard energy function E_{G_l}.

In [5], a method for detecting DDOS attacks is offered based on the estimation of the Hurst exponent using the Fourier fractional transformation, which makes the transition to the frequency-time area.

For the signal x (t) the fractional Fourier transform is determined as:

$$X_a(u) = F_a(u) = \int_{-\infty}^{\infty} x(t)K_a(t,u)dt \quad (16)$$

$$K_a(t,u) = \sqrt{1 - i \cdot \cot(\alpha)} \cdot \exp[i\pi(t^2\cot(\alpha) - 2ut \cdot \csc(\alpha) + u^2\cot(\alpha))], \alpha \neq n\pi \quad (17)$$

$$K_a(t,u) = \delta(t-u), a = (2 \mp 1)\pi \quad (18)$$

$$n \in Z, \alpha = \frac{a\pi}{2} \quad (19)$$

where a – the order of fractional Fourier transformation, provided that a=1, then the formula changes to the usual Fourier transformation.

Using a discrete wavelet transform and a multi-scale method of analysis, we can calculate the Hurst exponent if we analyze the expression:

$G(j) \leftrightarrow (2H + 1)j + \text{constant}$, where j – scale.

Next, the optimal selection of the range of scale intervals is made, using the method of one-dimensional weighted estimation of least squares [8].

Experimental verification of the proposed method showed its high accuracy, which reduced the number of false positives and omissions during the detection of the attack.

It was stated that network traffic is divided into several disjoint segments. The Hurst exponent for each segment is estimated. When the threshold values are exceeded, the traffic loses the property of self-similarity, which is regarded as a DDoS attack. But the intensity of the DDoS attack can change, which leads to a change in the Hurst exponent, so detection methods based on a fixed threshold require flexibility and adaptability.

This article proposes a method consisting of two stages:

1. Statistical analysis of the time series of network traffic using discrete wavelet transform and the Schwartz information criterion to find the change point of the Hurst exponent, which signals the start of a DDoS attack.

2. Adaptive regulation of the intensity of a DDoS-attack on the basis of fuzzy logic, by analyzing the Hurst exponent and the rate of its change. The Schwartz information criterion is based on the maximum likelihood function for the model and can be used to detect the presence of a threshold point by comparing the probability of a null hypothesis (no point) and an alternative (point of presence).

The Hurst exponent is estimated using a discrete wavelet transform, because in practice this method is one of the most reliable, as it is more resistant to gentle polynomial trends and noise.

5. The Evaluation is Performed According to the Following Algorithm

For the time series of network traffic X in real time, the wavelet coefficients d (j, k) are calculated for each scale j and position k. Next, it is necessary to perform a detailed assessment of the dispersion at each scale j:

$$S_j = \sum_{k=1}^{n_j} d^2(j,k) \quad (20)$$

where n_j – the number of wavelet coefficients that are available in scale j

We assume that a new sample of traffic is received, then the amount will be updated as follows:

$$n_j \neq n_j + 1 \quad (21)$$

$$S_j = S_j + d^2(j, n_j) \quad (22)$$

Estimation of variance in scale j:

$$\varepsilon_j = \frac{S_j}{n_j} \quad (23)$$

Next, the dependence of $\log_2(\varepsilon_j)$ on scaling j is constructed and a weighted linear regression is performed for the linear section, α is calculated.You do not need to build this dependency every time you receive a new segment of traffic, this action is performed only when necessary.

Then the Hurst exponent is calculated

$$H = \frac{\alpha+1}{2} \quad (24)$$

6. The Principle of Detecting Attacks

Let X be a time series of normal traffic, Y is a time series of traffic with deviations, Z - a time series of deviations, i.e. the relation Y = X + Z holds. Based on the theorems, we can conclude that regardless of the presence of self-similarity in Z, if X is a stationary self-similar process of the second order, then Y will still be a self-similar process. But the degree of self-similarity may change.

Let r_y, r_z be autocorrelation functions X, Y, Z, respectively. Then a $\|r_y - r_x\|$, is of interest during the attack, and $r_y = r_x + r_z$. For each value of $H \in (0.5,1]$ there is only one autocorrelation function on self-similarity. Thus, we examine $\|H_y - H_x\|$, where H_x and H_x is an average values of the Hurst exponents Y and X, respectively.

The disadvantage of this method is that the wavelet transform coefficients and statistics based on the Schwartz information criterion are updated at the moment when new traffic values arrive, and the detection of the traffic self-similarity threshold will be restarted for each scale. Thus, a signal of the change in the point of self-similarity will be given, even if this change occurred on a different scale at the same moment.

After an attack is detected, close to the detection time the traffic is divided into parts. By analyzing the Hurst exponent and the speed of its change (the difference between the Hurst exponents of traffic parts before and after the moment of detection), we can determine the intensity of DDoS-attack, using the rules of fuzzy logic.

Determining the point of change of traffic self-similarity by the Schwartz information criterion which is based on the assumption that the entropy of a sequence with a variable self-similarity boundary point is greater than the entropy of the sequence in which this point is fixed. Suppose there is a sequence of length M. It is assumed that there is only one point of the self-similarity boundary at position 1 <g <M. In order to calculate the presence and location of this point, you need to calculate the entropy of the whole sequence, as well as parts $f_1 = (1, \dots, g)$ and $f_2 = (g + 1, \dots, M)$, compare their values and conclude whether the point g is marginal. If the entropy of the individual parts is much less than the entropy of the whole sequence, the point g is considered to be marginal.

General scheme of attack detection is presented in Fig.1.

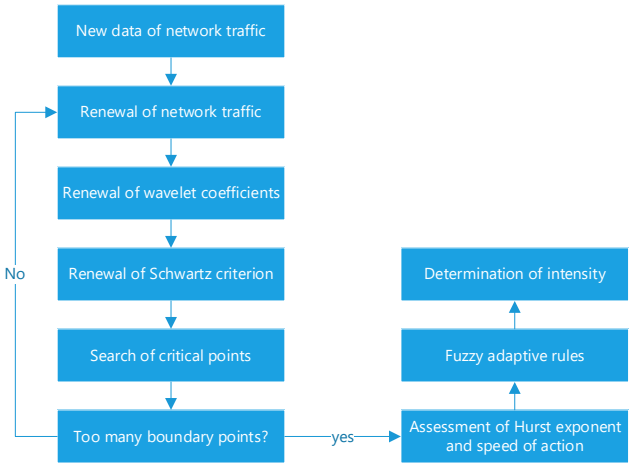

Figure 1: General scheme of attack detection

In the paper the Hurst exponents for four metrics of traffic by the iterative method in real time are calculated. Next, the collection and normalization of the results of anomaly detection to assess the security of network traffic is carried out.

The following scheme of traffic network security assessment is proposed (Fig. 2).

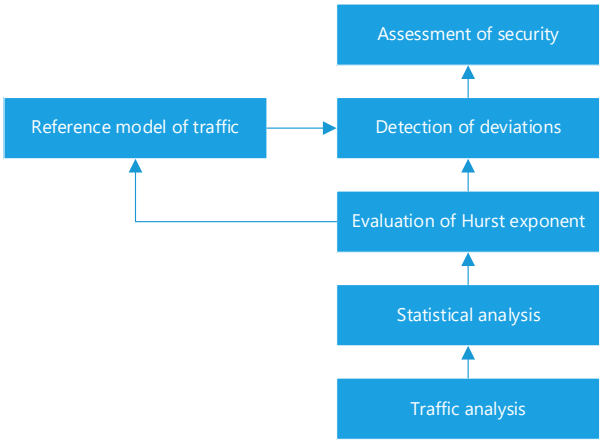

Figure 2: Assessment of local network traffic security

The algorithm for assessing traffic safety is divided into five stages:

1. Traffic collection.
2. Statistical analysis.
3. Assessment of the Hurst exponent Hearst index.
4. Detection of anomalies.
5. Security assessment.

To reduce the impact on the normal functioning of the network, traffic is duplicated on a special server that collects traffic. The software for collecting traffic on the server includes a hardware and technica complex, which has excellent performance when collecting network packets.

From the packets received from the router, information about the packet type is extracted, as well as four traffic metrics: the total number of packets, the number of TCP packets, UDP packets, ARP packets per unit of time. The Hurst exponent Hearst indices for four traffic metrics are calculated by an iterative real-time estimation method. These values are used to detect deviations and update the normal traffic model.

The current calculated value of the Hurst exponent is compared with the value from the normal model of traffic behavior. If the value is outside the allowable range, the traffic is considered abnormal. A normal traffic model is built by analyzing the normal operation of the network over a period of time.

The model includes a normal value of the Hurst exponent Hearst index and a confidence interval, and can be updated when a deviation is detected.

The criterion for assessing safety is the level of risk, which is calculated by the method of weighted averages, which takes into account the results of detection of deviations from the four traffic metrics. The level of risk provides administrators with the current state of data transmission in the network in terms of security.

Let $X_n(n = 1,2,3, \dots)$ - discrete stochastic process, and it is performed as follows:

$$X_i^{(m)} = \frac{1}{m}\sum_{k=(i-1)m+1}^{im} X_k \quad (25)$$

Then $X_i^{(m)}$ is called aggregated processes X_n of the order m with autocorrelation function $p^m(k)$ of the order m.

The stationary in a broad sense stochastic process $X_n(n = 1,2, \dots)$ is called self-similar, provided X_n and its aggregated processes $X_n^{(m)}$, of the order m which have the same autocorrelation functions $p^m(k) = p(k)(m = 1,2, \dots)$.

7. Algorithm for Iterative Evaluation of the Hurst Exponent

If the stationary mode, in a broad sense, is the time series X_i of network traffic acquires self-similarity function during the i^{th} period of time, and its autocorrelation function satisfies:

$$p_K = H(2H - 1)K^{2H-2}, K \to \infty \quad (26)$$

where H(0,5<H<1) - the Hurst exponent, which increases with increasing degree of self-similarity of the process.

As $\sum k\, p_k \to \infty$ self-similar process is often called long-scale correlation. The greater is K, the more relevance a time series has . The iterative formula for calculating H:

$$H_{i+1} = \sqrt{(p_k k^{2-2H_i}) \cdot 0{,}5}, k \to \infty \quad (27)$$

For a given time series $X_1, \dots, X^n$ it is calculated as:

1) expected value:

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad (28)$$

2) co-variance:

$$\widehat{y_k} = \frac{1}{n-k}\sum_{i=1}^{n-k}(X_i - \overline{X})(X_{i+k} - \overline{X}) \quad (29)$$

3) the autocorrelation function:

$$\widehat{p_k} = \frac{\widehat{y_k}}{\widehat{y_0}}, k = 0,1, \dots \quad (30)$$

The estimate of the autocorrelation function (p_k) serves as a replacement for p_k, then the iterative formula for calculating H takes the form:

$$\widehat{H_{i+1}} = \sqrt{(\widehat{p_k} k^{2-2\widehat{H_i}} + \widehat{H_i}) \cdot 0.5}, \quad k \to \infty \quad (31)$$

The results of the experiment showed that the iterative estimation of the Hurst exponent has a high speed and accuracy and also smaller confidence intervals for normal values compared to the methods of VarianceTime Plot. For a long-term large-scale correlation process is considered as $\widehat{H_0} = 0{,}5$.

An important condition for the execution of the iterative formula for $\widehat{H_{i+1}}$ is that $k \to \infty$, but the results of the experiment show that at k = 1 using this formula you can get the Hurst exponent with sufficient accuracy, reducing a significant number of calculations. In addition, the result is imperfect, even if k is large enough, so we take k = 1, and the formula takes the simplified form:

$$\widehat{H_{i+1}} = \sqrt{(\widehat{p_1} + H_i) \cdot 0{,}5} \quad (32)$$

In normal operating mode, network traffic satisfies the pattern of daytime use. To reduce the impact of network traffic periodicity on the Hurst exponent, it is necessary to process traffic at different periods of time.

In practice, the four above-mentioned normal traffic metrics are at first calculated during the week. Then the average weekly normal values of the Hurst exponents for four traffic metrics are calculated for each day.

After that it is necessary to use the effective method of Ketani and Gubner to calculate 98% of the confidence intervals of the Hurst exponent ($0{,}5 \le H \le 0{,}95$).

It is necessary to establish the initial state of the normal traffic model. When detected in real time mode, the value of the current calculated Hurst exponent is checked to fall into the confidence interval of the normal traffic model for each metric. If the value falls within the confidence interval, the traffic is considered normal, the detection result is 0, otherwise the traffic is considered with a deviation, and the detection result is 1. In the first case it is necessary to update the Hurst exponent and the confidence interval in the normal traffic model

The method of normalized security assessment is based on the weighted average method for accounting of all four traffic metrics. The level of risk is calculated as follows:

$$F_{traffic} = \sum_{i=1}^{4} w(i) \cdot F_{obs}(i), \quad \sum_{i=1}^{4} w(i) = 1 \quad (33)$$

$$obs = \{1 - \text{all packets}, 2 - \text{TCP packets}, 3 - \text{udp packets}, 4\ \text{arp packets}\}$$

In order to conduct the study, software was used for fractal analysis of time series (Fig. 3).

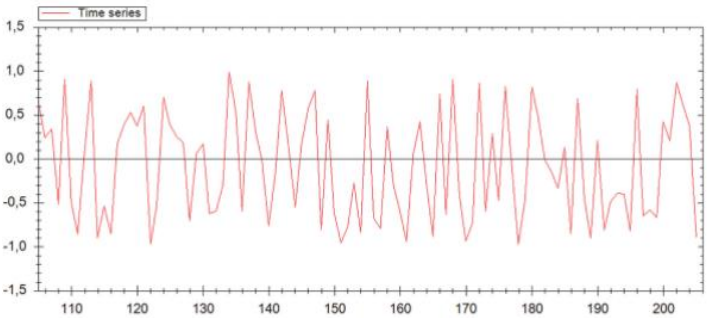

Figure 3: General view of the investigated program

Also, a graphical image of the Hurst exponent for the time series of processor's load and user expectations were modelled as you can see it in Fig. 4.

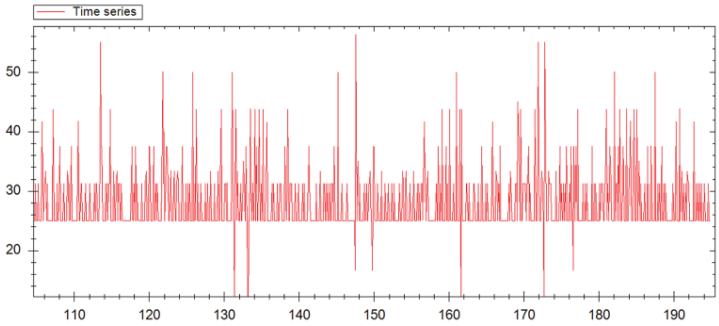

Figure 4: Time series of processor's load

8. Conclusions

In this review we are talking mainly about network traffic, for which numerous studies have shown that it has the property of self-similarity, which allows us to use this fact to create a model of normal behavior.

In this article, an experiment was performed for a time series of CPU usage, the fractal properties of which are unknown.

The results show that the Hurst exponent of the time series of this parameter changes when changing the type of user activity in a wide range, which does not allow to make a conclusion about the presence or absence of self-similarity and makes it impossible to detect anomalies using only this method for this parameter.

9. References

[1] Shelukhin, O. I. Multifractals: infocommunication applications / O. I. Shelukhin. Hotline-Telecom, 2014. 579 p.

[2] Porshnev, S. V. Mathematical models of information flows in high-speed backbone Internet channels: a tutorial / S. V. Porshnev. Hotline-Telecom, 2016. 233 p.

[3] Modeling and analysis of security and risk in complex systems: Proceedings of the International Scientific School IABR - 2016 (St. Petersburg, October 25 - 28, 2016.

[4] D.P. Zegzhda. Cybersecurity of the digital industry. Theory and practice of functional resistance to cyber attacks. E.B. Alexandrova, M.O. Kalinin, A.S. Markov, I.Yu. Zhukov, D.V. Ivanov, A.S. Konoplev, D.S. Lavrova, D.A. Moskvin, E.Yu. Pavlenko, M.A. Poltavtseva, N.N. Shenets, A.D. Dakhnovich, V.M. Krundyshev. - Moscow. Hotline - Telecom, 2020. - 560s.

[5] O.Y. Ruslyachenko, K.O. Osadchuk. Discrete data transfer technologies [Text] Module 3: Teaching manual for laboratory works and practical seminars. Information security and data transmission department. - Odessa: ONAT by the name of A.S. Popov, 2013. 60 p

[6] Bogachuk, I., Sokolov, V., Buriachok, V. (2018). Monitoring Subsystem for Wireless Systems Based on Miniature Spectrum Analyzers. 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). DOI: 10.1109/infocommst.2018.8632151.

[7] Pereverzev, A., Ageyev, D. Design method access network radio over fiber (2013). 2013 12th International Conference: The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2013, art. no. 6543268, pp. 288-292

[8] Kostrov, D. Intrusion detection systems (2002) Byte, 8 (49), pp. 14-21

[9] Radivilova, T., Hassan, H.A. Test for penetration in Wi-Fi network: Attacks on WPA2-PSK and WPA2-enterprise (2017) 2nd International Conference on Information and Telecommunication Technologies and Radio Electronics, UkrMiCo 2017 - Proceedings, art. no. 8095429

[10] Bondarenko, S., Liliya, B., Krynytska, O., Inna, G. Modelling instruments in risk management (2019) International Journal of Civil Engineering and Technology, 10 (1), pp. 1561-1568
[11] Dobrynin, I., Radivilova, T., Maltseva, N., Ageyev, D. Use of Approaches to the Methodology of Factor Analysis of Information Risks for the Quantitative Assessment of Information Risks Based on the Formation of Cause-And-Effect Links (2019) 2018 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2018 - Proceedings, art. no. 8632022, pp. 229–232.