ASSESSMENT OF THE STATE OF THE COMPUTER SYSTEM BASED ON THE HURST EXPONENT

Sergei Semenov, Svetlana Gavrilenko, Victor Chelak.

Abstract: The method of identifying abnormal behavior of computer systems based on the Hurst exponent is examined in this report. Results of the research suggest the possibility of using the Hurst exponent for identifying the anomalous behavior of computer systems in the overall system to detect malicious software.

Keywords: Hurst exponent, computer security, computer virus, abnormal behavior.

1. Introduction

Today cyberthreats are developing actively, and understanding the direction of this development plays a key role in providing enterprises' effective security. If viruses aren't identified at early stages, the costs of recovery after the attack increase by more than twofold. For example, the total cost of recovery after a cyber-attack lasting a week or more amounted to over $ 1 million. At the same time, the immediate response to the incident cost the company an average of 400 thousand dollars. [1].

To date, there is a large number of different antivirus programs, the basis of which is most often based on signature and heuristic analysis technologies [2]. Data on threats is collected from a variety of sources, including cloud infrastructure, search robots, botnet monitoring services, spam traps. New cyber threats are determined by checking URLs, domains, IP addresses, checksums of files, timestamps, file names, DNS data and other characteristics inherent in the programs. The received information is carefully checked, organized, cleaned and analyzed both with the help of heuristic analysis tools and by the analysts of the company developing antivirus software [3].

Heuristic analyzers, as a rule, include intellectual subsystems based on the theory of artificial intelligence, for example, on the basis of methods of fuzzy logic, cluster analysis, coordinated heuristics or the theory of neural networks [4-8]. At the same time, they are all based on the assumption that the computer systems (CS) have their own pattern of normal behavior and any significant deviations from it may be due to the impact of intruders. That is why a very important task is the choice or formation of such a template that would reproduce the functional portrait of the CS and fix its anomalous behavior with a given accuracy.

The analysis of the literature showed that methods of statistical data processing (for example, control cards, BDS testing [5-6]) are widely used to detect anomalies in production management and business processes. To justify the prediction of trends, in natural sciences, the Hurst exponent is often used to identify new characteristics of the process. [7-12]

2. Development of methods for detecting intrusions in computer systems based on the Hurst index

The Hurst exponent was first used by the outstanding British hydrologist Harold Edwin Hurst when designing a dam on the Nile in Egypt to assess the inflow and outflow of water [7]. Hurst, having studied the records of the floods of the Nile for nine centuries, found regularity in this process. He proved the possibility of distinguishing a random series from a non-random one, even if the random series is not normally distributed, linking it to the degree of self-similarity of the process. An object that has this "quality" is statistically similar in different scales - spatial or temporal, that is, it has a cyclicity.

The calculation of the Hurst exponent [12] can be performed using the following formula (1):

$$H = \frac{\log(\mathrm{R/S})}{\log(\mathrm{aN})} \qquad (1)$$

Where:

H is the Hurst exponent;

S is the standard deviation of a set of observations x;

R is the range of the accumulated deviation Zu;

N is the number of observation periods;

A is a given constant, a positive number.

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - X_a)^2} \qquad (2)$$

where X_a – is the arithmetical mean of a set of observations x in N periods

$$X_a = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad (3)$$

The range of the accumulated deviation R is the most important element of the formula for calculating the Hurst index. In general, it is calculated as follows:

$$R = \max\nolimits_{1\le u\le N}(Z_u) - \min\nolimits_{1\le u\le N}(Z_u) \qquad (4)$$

Z_u is the accumulated deviation of set x from the average X_a:

$$Z_u = \sum_{i=1}^{u}(x_i - X_a) \qquad (5)$$

The Hurst exponent (H) characterizes the degree of self-similarity of the process as follows [11-12]:

1) 0 <H <0.5 is a random process that does not have self-similarity and is characterized by a tendency toward an average value;

2) H = 0.5 is a completely random process without a pronounced tendency;

3) H> 0.5 is a trend-based process that has a long memory and is self-similar.

From the formula for calculating the Hurst exponent, it can be seen that its growth is influenced by:

- increase in the range of the R oscillations;
- reduction of the root-mean-square error S;
- decrease in the number of observations of N.

In this paper, the possibility of fixing the abnormal behavior of a computer system (CS) based on the Hurst exponent is analyzed.

It is known that the abnormal behavior of the operation of computer systems is characterized by excessive consumption of system resources, such as RAM, traffic, CPU utilization, increases the number of received and transmitted data packets.

In the course of the study, it was decided to use the central processor (CPU) as input data.

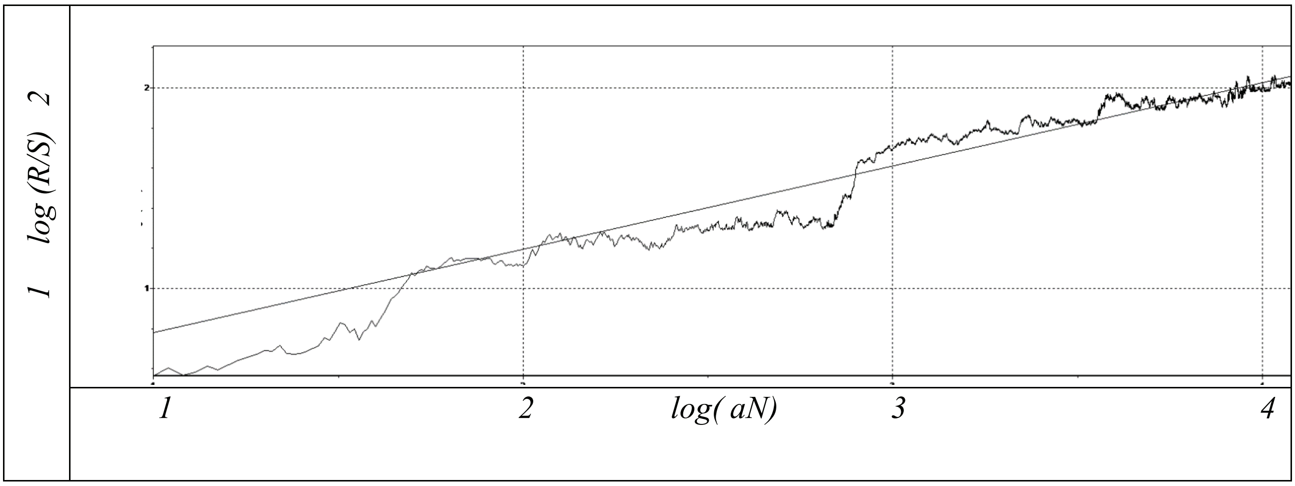
To carry out the research, a software model has been developed that provides for the variation of N - the number of time series values. The value of the CPU load is scanned every second and saved in a file.

The received input values are fed into the input of the analysis module, which processes the data and calculates the Hurst exponent.

Fig. 1 shows the curve of the relation between the set of values *log* (*R / S*) to *log* (*aN*) of the CPU load and the result of calculation of the Hurst exponent for infection by the virus type VirKP55

As can be seen from Fig. 1, the Hurst exponent indicates the randomness of the process, which does not have self-similarity and is characterized by a desire for an average value.

Similar results were obtained when the system was infected with other types of viruses. Fig. 2

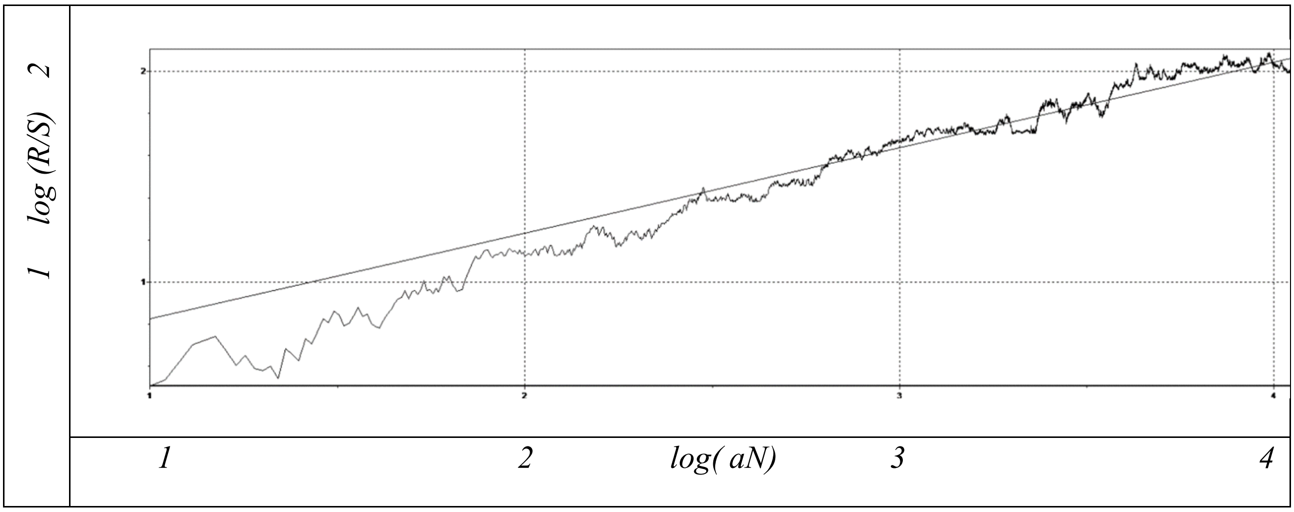


The Hurst exponent, H = 0.414775 ± 0,125261

Fig. 1. The curve of the relation between log (R/S) to log(aN) of the CPU load and the result of calculation of the Hurst exponent for infection by the virus type VirKP55

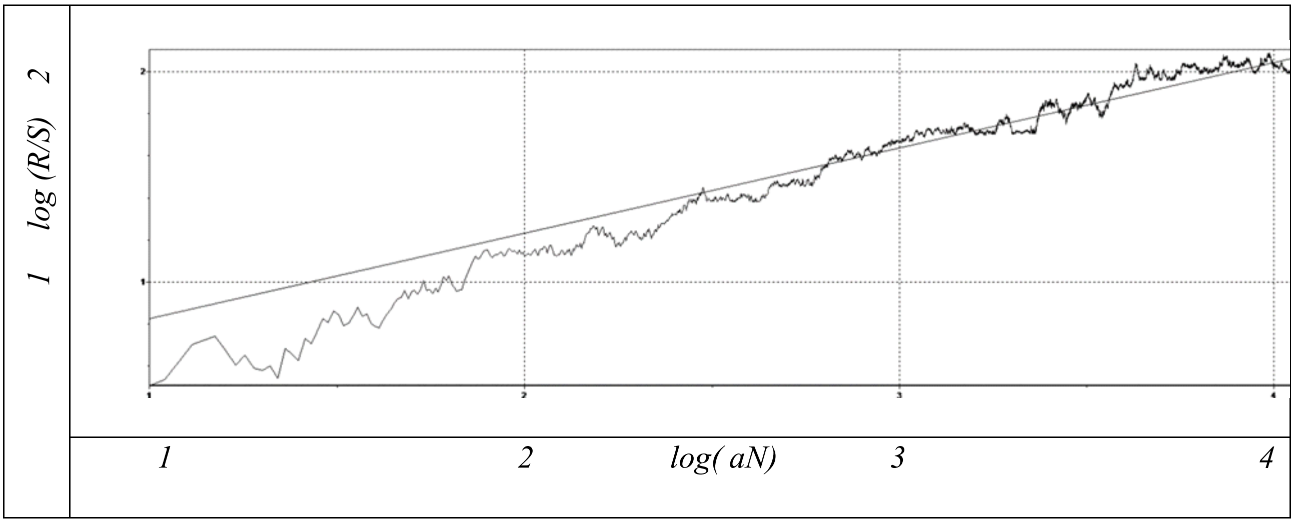
shows the result of modeling when a virus is infected with a virus type VirMask.

The results of analyzing the operation of a computer system in safe mode are of interest. Fig. 3 shows one of the results of research and calculation of the Hurst index, as well as a graph of the dependence of the numerical load range of the CPU in safe mode. Analysis of the graph (Figure 3) shows the presence of long-term dependencies in statistical data.



Hurst exponent, H = 0.406405± 0.124816

Fig. 2. The curve of the relation between log (R/S) to log(aN) of the CPU load and the result of calculation of the Hurst exponent for infection by the virus type VirMask



Hurst exponent, H = 0.812745± 0.289510

Fig. 3. The curve of the relation between log (R/S) to log(aN) of the CPU load in safe mode of a CS

3. Conclusion

The results of the research showed:

1. Under the impact of malicious software on the computer system, the statistical indicators of the main parameters of the system functioning (CPU load, RAM, etc.) change.

2. The impact of a number of viruses on the computer system leads to a change in the Hurst exponent, which indicates the randomness of the process, which does not have self-similarity and is characterized by a tendency toward an average value of 0.5.

3. It should be noted that one of the shortcomings of the proposed method for identifying the state of CS is a large (in some cases up to 0.4) deviation from the average value of the fractal dimension and Hurst. Therefore, in order to improve the accuracy of the structural identification of the state, it is necessary to use other methods of state identification, (control cards, BDS testing, etc.) along with the method considered.

References:

[1] **K. Kaspersky.** *Kaspersky lab presented a new analytical service - Kaspersky Threat Lookup*. – Web: https://www.pcweek.ru/security/news-company/detail.php?ID=195249 .

[2] **O.I. Shelukhin, D.Zh. Sakalema, A.S. Filinova.** *Identifying intrusions in computer systems. Print.:* Telecom-Hotline, 2013, pg. 220

[3] **A.V. Lukatsky.** *Identifying attacks. Print.:* VHB-Petersburg, 2001, pg. 624.

[4] **S.G. Semyonov, V.V. Davydov, S.Y. Gavrilenko.** *Data security in computerized control systems (monography).* «LAP LAMBERT ACADEMIC PUBLISHING»: Cemany, 2014, 236 c.

[5] **S. Gavrilenko, V. Chelak, Hornostal O.** *Intrusion detection in computer systems.* Proceedings of the symposium “Metrology and metrology assurance”– Sozopol, Bulgaria, 2016, pp. 342-347

[6] **Gavrilenko S**. Developing parametrical criterion for registering abnormal behavior in computer and telecommunication systems on the basis of economic test// G. Semenov, S. Gavrilenko, V. Chelak//Actual problems of economics. – Kiev, 2016, Vol 4(178), pp. 451-459.

[7] **Semenov S.** Approximating computer system operation technologies under external action through the brusselator model with perturbation in the form of dynamic chaos / S. Semenov, S. Gavrilenko // Revista RECENT – Industrial Engineering Journal – Transilvania University of Brasov – Romania, Vol. 16 (2015), No. 1 (44).

[8] **Eskin E.** Anomaly detection over noisy data using learned probability distributions. In Proc. 17th International Conf. on Machine Learning, pages 255-262. Morgan Kaufmann, San Francisco, CA, 2010

[9] **Hurst H. E.,** *Long-term Storage of Reservoirs.* Transactions of the American Society of Civil Engineers 116, 1951.

[10] **Piskaryov Dmitriy.** *Calculating the Hurst exponent:* – Web: https://www. mql5.com/ru/articles/2930

[11] **D.S. Kirillov** *Distributions of the Hurst exponent of the time-varying labeled time series.* IPM Preprint named after M.V. Keldish. 2013. № 11. Pg. 16. URL: http://librarykeldysh.ru /preprint.asp?id=2013-11

[12] **Eric Nayman.** *Calculating the Hurst exponent with the purpose of finding trends (persistence) in fincncial markets and macroeconomic indicators:* – Web: http://wealth-lab.net/ Data/Sites/1/SharedFiles /doc/forindicators/articles/04_erik_naiman_herst.pdf

Authors:

Serhiy Hennadiyovich Semyonov. Kharkiv Military University (1989), Doctor of Engineering (2014), senior staff scientist (2012). Place of work – Professor of the "Computer Science and Programming" department of the National Technical University "KhPI". Scientific interests: security in technical systems. Ukraine, 61002, Kharkov, ul. Kirpicheva, 21, *s_semenov@ukr.net.*

Svetlana Yuryivna Gavrylenko, Kharkiv Polytechnical Institute (1986), Doctor of Science (2001), associate professor (2005). Place of work – Professor of the "Computer Science and Programming" department of the National Technical University "KhPI". Scientific interests: security in technical systems, digital machines. 61002, Ukraine, Kharkiv, 21 Kirpichova str., *gavrilenko08@gmail.com.*

Victor Vladimirovich Chelak. Place of education – "Computer Science and Programming" department of the National Technical University "KhPI". Scientific interests: information security in systems and networks. 61002, Ukraine, Kharkiv, 21 Kirpichova str., *victor.chelak@gmail.com.*