A Method of Determining the Fractal Dimension of Network Traffic by Its Probabilistic Properties and Experimental Research of the Quality of This Method

Hanna Drieieva, Oleksandr Drieiev, Yelyzaveta Meleshko, Mykola Yakymenko and Volodymyr Mikhav

Central Ukrainian National Technical University, 8, Universytetskyi prosp., Kropyvnytskyi, 25006, Ukraine

Abstract

Taking into account the fractal properties of computer network traffic allows one to predict the information processes in them. The known criteria for determining the fractal dimension, such as the Hurst exponent, have significant errors and deviations for some cases, so it is advisable to develop new methods for estimating the fractal characteristics of the researched signal. The authors had proposed a method for determining the fractal dimension of network traffic by its probabilistic properties. The purpose of this paper is to research the quality of the proposed method. In this work, a binary time series is used to model fractal binary network traffic, which persistence is regulated by setting up the probability of one state change to the opposite by means of Markov chain. The generated traffic was used to investigate the quality of the proposed probabilistic method of determining the fractal dimension of network traffic and to compare it with the method based on R/S analysis. A series of experiments was conducted which showed that R/S analysis gives different values with different cumulative sums for the same data, that indicates the ambiguity of the method. The probabilistic method does not have this disadvantage and gives unambiguous results. Also, the developed method has a lower deviations from the mean value of the Hurst exponent, and therefore it is more accurate in determining the fractal dimension than R/S analysis method – R/S analysis has a deviation of 2.5%, and the developed method has 1.8%.

Keywords

time series forecasting, fractal dimension, Hurst exponent, network traffic, data analysis, computer simulation

1. Introduction

During researching and modeling information processes in telecommunication systems and computer networks, it is necessary to take into account the fractal properties of telecommunication traffic. In particular, taking into account fractal properties significantly affects the results of simulation modeling of information exchange processes, and, accordingly, it allows predict the development of events in information processes that are associated not only with the movement of information in telecommunications systems [1-6]. All these sources are related to the problem of determining the fractal dimension of time series, and the dates of publications vary from 1994 to 2021, which indicates that the problem of rapid and accurate determination of fractal dimensions is not finally solved and needs further research.

Often, to determine the fractal characteristics of time series, used methods are based on the simulation of random walks [7-11], which depends on the standard deviation of the random variable and the length

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: gannadreeva@gmail.com (H. Drieieva); drey.sanya@gmail.com (O. Drieiev); elismeleshko@gmail.com (Ye. Meleshko); m.yakymenko@gmail.com (M. Yakymenko); mihaw.wolodymyr@gmail.com (V. Mikhav)
ORCID: 0000-0002-8557-3443 (H. Drieieva); 0000-0001-6951-2002 (O. Drieiev); 0000-0001-8791-0063 (Ye. Meleshko); 0000-0003-3290-6088 (M. Yakymenko); 0000-0003-4816-4680 (V. Mikhav)

of the series. In particular, such an indicator is R/S analysis [12-16]. It is believed that R/S analysis for searching the self-similarity has a higher accuracy for long series with large cumulative sums [6]. Also, it is noted that determining the self-similarity coefficient depends on the length of the researched series [17]. It is not always possible to obtain long time series for its analysis. For example, time series properties change before it will be able to be used the calculated coefficient, or the process itself is too short to obtain enough data for reliable analysis. Therefore, it can be seen that the flow of scientific works on methods for obtaining the self-similarity coefficient of the time series does not decrease. Eventually it demonstrates the relevance of the given problem.

1.1. Related Works and Problem Statement

So, during solving problems of forecasting and optimizing the operation of telecommunication networks, it is almost necessary to take into account the fractal nature of traffic on the Internet [18-21].

The known criteria by which fractal dimension is determined, for example, the Hurst exponent [22], have significant errors and deviations for individual implementations, so it is advisable to develop new methods for assessing the fractal characteristics of the signal under study, such researches are observed in a number of publications, for example, in [23] was proposed method using wavelet analysis.

Some review publications, which consider and compare several methods, use the definition of the fractal dimension of individual short-term implementations, based on which determine the properties of this signal [24]. But there are cases of ergodic signal with a known theoretical justification of its probability density distribution function. In such cases, short implementations have a high probability of deviations, but such implementations allow us to estimate the weights of the known probability distribution. Exactly in such cases it is useful to have high-quality means of determining the fractal dimension based on the probability distribution function, which is the reason for the constant search and development of new better methods for determining the fractal dimension. In the processes of information transmission in telecommunication systems and networks, the search for the fractal dimension of network traffic is widely used. To do this, use different methods and approaches aimed at reducing random deviations when performing calculations on relatively small implementations of numerical series [19-21, 23, 24]. Given the lack of input data, additional data on the nature of the studied time series can significantly improve the situation. For example, with the help of theoretical justifications for modeling time series, they can be realized on the basis of random processes with given probability distributions, such as Poisson process [25, p. 32], Markov chains [25, p. 89], queues based on Pareto process [26] and others. Each of the distributions has its own scope and is based on accepted hypotheses and experimental testing on long-term implementations. However, real processes have only approximate distributions to the theoretical ones, it is worth remembering that the theoretical flow of data per unit time is limited by the bandwidth of the input channel of the network. Therefore, the approximation can be improved by taking into account the real probability distributions obtained experimentally.

The authors proposed a method for determining the fractal dimension of a numerical sequence by the probability distribution of the values of its elements (Probabilistic method) in [27, 28]. In the mentioned work, the problem of obtaining a mathematical relation to obtain the expected value of the fractal dimension of network traffic based on the knowledge of the probability density distribution $p(x)$ was set and solved. The results can be used to predict changes in network traffic, as well as to improve the modeling of fractal random sequences, queuing systems [29], simulating the traffic of telecommunications networks [30].

The purpose of this work is an experimental study of the quality of the proposed method for determining the fractal dimension of network traffic by its probabilistic properties.

2. A method for determining the fractal dimension of network traffic by its probabilised properties

2.1. Theoretical substantiation and essence of the method

Modern traffic management systems take into account its fractal properties. Network traffic can be represented as a time series. In systems that use packet information exchange, it is more convenient to use a binary time series, and the representation of traffic at the level of "data packet is present", "there is no data packet". In this work, a probabilistic analysis of such a time series is performed in order to highlight the probabilistic characteristics that reveal not only the intensity of such traffic, but also its fractal properties.

The following relations are valid for the probability density distribution $p(x)$:

$$\int_{-\infty}^{+\infty} p(x)dx = 1, \qquad p(x) \geq 0, \qquad M(x) = \int_{-\infty}^{+\infty} x \cdot p(x)dx \tag{1}$$

where $M(x)$ – is the mathematical expectation of the random variable x. A numerical sequence consists of a series of implementations x_i, where i – is the serial number of the element.

By definition, the Minkowski dimension is the value of the following limit:

$$\lim_{\varepsilon \to 0} \frac{\ln(N_\varepsilon)}{\ln(\varepsilon)},$$

where ε – is the diameter of the coating element; N_ε – is their number. In practice, the problem is solved geometrically by covering the studied figure with squares (cubes), where its side is taken as the diameter. In [1, 31] there is a variant in which the coverage is replaced by rectangles of width ε and height, which is minimal to cover the area of the graphical representation of the numerical series.

When using rectangles, the number of figures of the coating is replaced by the coverage area:

$$S(\varepsilon) = \sum_{k=0}^{N_\varepsilon} h_k, \tag{2}$$

where h_k – is the height of the corresponding rectangle k.

If take $\varepsilon > 1$ for the number of discrete samples of the numerical series x_i, then h_k is searched according to the following algorithm:

$$h_k = \max(x_i) - \min(x_i), \qquad \text{where } i = k\varepsilon, \ldots, (k+1)\varepsilon - 1. \tag{3}$$

For an inclined straight line, the coating with a decrease ε in 2 times reduces the area also twice.

For a figure that covers a plane (for example, the Hilbert or Piano curve), when ε changes, the value of the area does not change and covers the entire part of the plane. This corresponds to the fractal dimension of the plane $D = 2$. Therefore, the fractal dimension is expressed through the area of coverage [2, 3] by relation (4), if the width of the rectangle is changed γ times:

$$D = 2 - \log_\gamma(S(\varepsilon \cdot \gamma)/S(\varepsilon)). \tag{4}$$

For a significant number of samples of the numerical sequence N, it is possible to replace the values of the heights of the rectangles of coverage h_k by their mathematical expectation $M(h_k)$. But the mathematical expectation of the height of the rectangle depends on the number of samples per partition ε, and this dependence is nonlinear. Therefore, we denote the mathematical expectation of the height of a rectangle depending on its width ε as $M(\varepsilon)$, where the mathematical expectation of the area of an individual rectangle will be expressed as the product of its width and the expected height (5):

$$S_k(\varepsilon) = M(\varepsilon) \cdot \varepsilon. \tag{5}$$

Then have the expression for determining the fractal dimension through the coverage of rectangles of formula (2), taking into account (5):

$$S(\varepsilon) = \sum_{k=0}^{N_\varepsilon} S_k(\varepsilon) \Rightarrow \ S(\varepsilon) = \sum_{k=0}^{N_\varepsilon} M(\varepsilon) \cdot \varepsilon \ \Rightarrow \ S(\varepsilon) = M(\varepsilon) \cdot \varepsilon N_\varepsilon \Rightarrow \ S(\varepsilon) = M(\varepsilon) \cdot N, \tag{6}$$

where N is the total number of elements of the series, which is divisible by ε (although for $N >> \varepsilon$ this requirement can be neglected and incomplete last rectangles of coverage can be accepted).

If use the obtained relation (6) to the dimension value, obtain:

$$D = 2 - \log_\gamma\left(\frac{M(\varepsilon \cdot \gamma)}{M(\varepsilon)}\right), \tag{7}$$

in which the number of numbers in the implementation of series N has been reduced and is no longer used.

Unfortunately, the mathematical expectation of the height of the rectangle (3) is not known. Therefore, we obtain from the mathematical expectation of the quantity itself. To do this, find the probability $q_3(h)$ of the height of the rectangle for $\varepsilon = 2$:

$$q_2(h) = \int_{-\infty}^{+\infty} 2p(x) \cdot p(x+h)dx, \tag{8}$$

an additional factor of 2 is used here, because the resulting probability turns out in two independent cases when the smaller number is obtained earlier and later.

The general expression for $q_3(h)$ is obtained if all subsequent realizations of the number x, by the number ε-2, will be located on the interval $[x, x+h]$. To do this, the probability of $\int_x^{x+h} p(x)dx$ falling into this interval must be increased to the power of ε-2 and multiplied by their possible combinations of the location of the limit values x and $x + h$, as which acts binomial coefficient $C_\varepsilon^{\varepsilon-2} = \varepsilon(\varepsilon-1)/2$:

$$p_{\varepsilon-2}(x, x+h) = \varepsilon(\varepsilon-1)\left(\int_x^{x+h} p(t)dt\right)^{\varepsilon-2}, \tag{9}$$

Therefore, for $q_3(h)$, the expression is written with (8) taking into account (9) as the following integral (10):

$$q_3(h) = \varepsilon(\varepsilon-1)\int_{-\infty}^{+\infty} p(x) \cdot p(x+h)\left(\int_x^{x+h} p(t)dt\right)dx. \tag{10}$$

Formula (10) in fact provides information about the probability density distribution of the height of the rectangle of coverage with a width of ε samples. The value of the expression is not valid for $\varepsilon < 2$, because for a single width of the rectangle, it should cover only one point, so its height will always be zero. In other cases, zero or less samples cannot be taken, because the number of samples ε is a natural number.

Now that there is a value for the probability distribution of the width of the rectangle (10), we can write the value of the mathematical expectation of this height (11):

$$M(\varepsilon) = \varepsilon(\varepsilon-1) \cdot \int_0^{+\infty} h\left(\int_{-\infty}^{+\infty} p(x) \cdot p(x+h)\left(\int_x^{x+h} p(t)dt\right)^{\varepsilon-2} dx\right)dh. \tag{11}$$

For the mathematical expectation $M(\varepsilon)$ the integration limit is taken on the interval [0; +∞], because the height of the rectangle cannot have a negative value.

Finally, using the obtained mathematical expectation of the height of the rectangle (11) for expression (7), we have an estimate of the fractal dimension based on the density of the distribution (1):

$$D = 2 - \log_\gamma\left(\frac{\gamma(\varepsilon \cdot \gamma - 1)\int_0^{+\infty} h\left(\int_{-\infty}^{+\infty} p(x) \cdot p(x+h)\left(\int_x^{x+h} p(t)dt\right)^{\varepsilon \cdot \gamma - 2} dx\right)dh}{(\varepsilon-1) \cdot \int_0^{+\infty} h\left(\int_{-\infty}^{+\infty} p(x) \cdot p(x+h)\left(\int_x^{x+h} p(t)dt\right)^{\varepsilon-2} dx\right)dh}\right). \tag{12}$$

Sometimes it is advantageous to use a non-multiple partition relation, so the following notation of expression (12) will be more useful:

$$D = 2 - \log_{\varepsilon_2/\varepsilon_1}\left(\frac{\varepsilon_2(\varepsilon_2 \cdot \gamma - 1)\int_0^{+\infty} h\left(\int_{-\infty}^{+\infty} p(x) \cdot p(x+h)\left(\int_x^{x+h} p(t)dt\right)^{\varepsilon_2 - 2} dx\right)dh}{\varepsilon_1(\varepsilon_1-1) \cdot \int_0^{+\infty} h\left(\int_{-\infty}^{+\infty} p(x) \cdot p(x+h)\left(\int_x^{x+h} p(t)dt\right)^{\varepsilon_1-2} dx\right)dh}\right). \tag{13}$$

A formula for obtaining a mathematical expectation of the fractal dimension of a sequence of random numbers with a known probability density distribution (13) is obtained.

The obtained formula (13) is transient for estimating the fractal dimension of a binary series, which consists only of elements that are equal to 0 or 1 (Fig. 1).

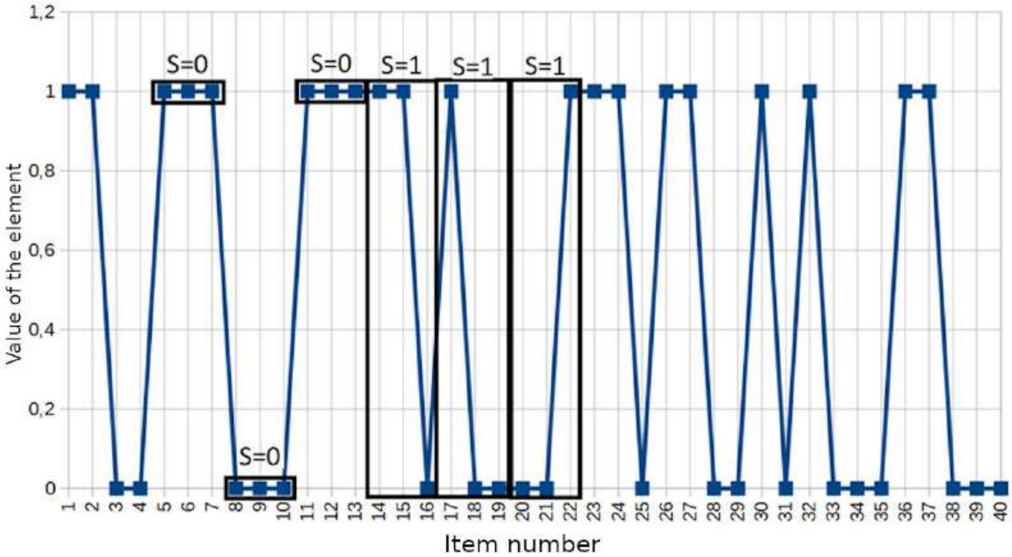

Figure 1: Visualization of zero coverings on a discrete binary number series

To estimate the fractal dimension of such a binary series, need to determine the probability of obtaining a sequence of a given length ε zeros $p_0(\varepsilon)$ and units $p_1(\varepsilon)$:

$$p_0(\varepsilon) = \lambda_0^{\varepsilon},\ p_1(\varepsilon) = \lambda_3^{\varepsilon},$$

where λ_0 is the probability of maintaining the zero state, λ_3 is the probability of maintaining the single state. According to the defined probabilities, the mathematical expectation of the degree of coverage n elements of the sequence will be expressed by the following formula:

$$M(n) = 1 - \frac{(1-\lambda_3)\lambda_0^n + (1-\lambda_0)\lambda_3^n}{(1-\lambda_0) + (1-\lambda_3)}.$$

If, for convenience, go to probability designations to change the current state in the series from zero to one and vice versa, λ_1=1-λ_0 and λ_2=1-λ_3, then in (13) can use discrete sums instead of integrals. Determining the mathematical expectation of the degree of coverage $M(n)$, allows to use the following formula to determine the fractal dimension of the binary sequence:

$$D(n_1, n_2, \lambda_1, \lambda_2) = 1 + ln\left(\frac{\lambda_1 + \lambda_2 - \lambda_2(1-\lambda_1)^{n_1} - \lambda_1(1-\lambda_2)^{n_1}}{\lambda_1 + \lambda_2 - \lambda_2(1-\lambda_1)^{n_2} - \lambda_1(1-\lambda_2)^{n_2}}\right) / ln(n_1/n_2).$$

The fractal dimension search expression contains not only the transition probabilities but also the coverage lengths. This leads to ambiguity in the resulting dimension. To get rid of the ambiguity, direct n_1 to n_2 and then direct the number of elements to 1. In result of finding the boundaries, we obtain the final formula for determining the fractal dimension of a binary sequence by transition probabilities [27]:

$$D(\lambda_1, \lambda_2) = 1 - \frac{\lambda_2(1-\lambda_1)ln(1-\lambda_1) + \lambda_1(1-\lambda_2)ln(1-\lambda_2)}{2\lambda_1\lambda_2}. \quad (14)$$

Due to the fact that the binary time series is given by the probabilities of changing or leaving the state, it was possible to express the fractal dimension analytically (14). Visually, this can be assessed in Fig. 2. Unfortunately, the coating method has an effect on the value of the fractal dimension, so to compare the results you need to make the transition from the resulting dimension to others by regression.

As a result, we obtained a method for determining the fractal dimension of network traffic by its probabilistic properties:

Stage 1. **The binary time series** for which it is necessary to define fractal properties accumulates.

Stage 2. **The probabilistic properties of the series are evaluated**. The accumulated data calculates the number of occurrences of pairs of consecutive elements n_{00} : "0, 0" and n_{01} : "0, 1". We obtain λ_1 as the probability of change of state from "0" to "1" by the definition of probability $\lambda_1 = n_{01} / (n_{01} + n_{00})$.

Similarly, the probability of changing the state from “1” to “0” is estimated as $\lambda_2 = n_{10} / (n_{10} + n_{11})$, where n_{10} is the number of pairs of consecutive elements “1, 0”; n_{11} – is the number of pairs of consecutive elements “1, 1”.

***Stage 3.* The fractal dimension of a binary sequence is determined** by formula (14).

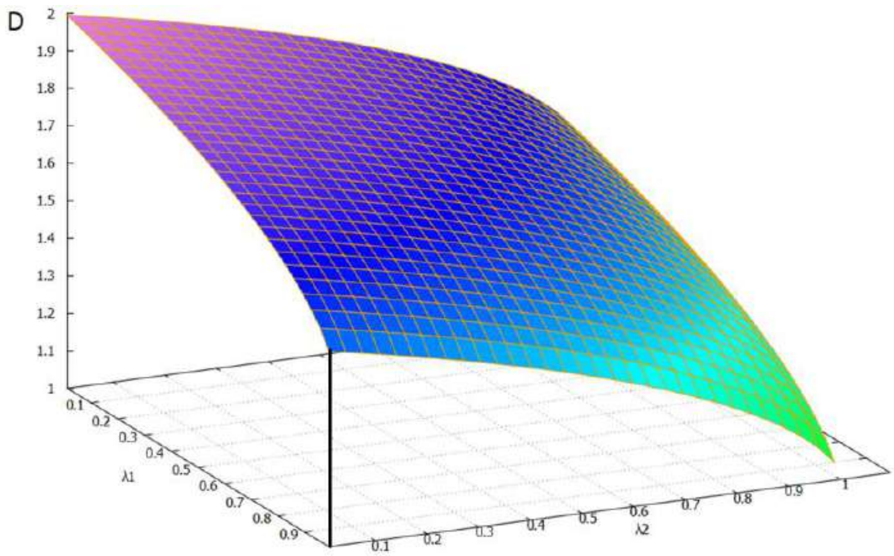

Figure 2: Dependence of the fractal dimension of a binary series on the probabilities of changes in the state of the system

For practical use of the results there is a need to create a binary sequence generator with specified probability characteristics. This question is discussed below.

2.2. Experimental study of the quality of the method

To conduct experiments, it was decided to simulate network traffic with predefined properties. It was decided to model based on the theory of Markov processes, which is often used to model the traffic of different queuing systems [32-38]. For generation fractal binary traffic used Markov chain, shown in Fig. 3.

In this work, a binary time series was created to simulate network traffic, the persistence of which is regulated by setting the probabilities of state change to the opposite λ_1, λ_2 (Fig. 3).

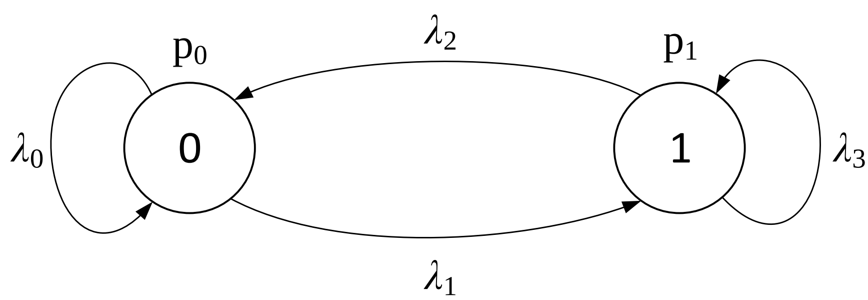

Figure 3: Markov chain for generating fractal binary traffic

This generator is characterized by states 0 or 1, and the probabilities of being in these states as $p_0 = \lambda_2/(\lambda_1 + \lambda_2)$ and $p_1 = \lambda_2/(\lambda_1 + \lambda_2)$, where λ_i – is the probabilities of the corresponding transitions [27]. The traffic intensity of such a generator will be within [0, 1] and will be equal to the probability of obtaining the output of the generator 1: p_1. The algorithm of operation of such a generator is shown in Fig. 4.

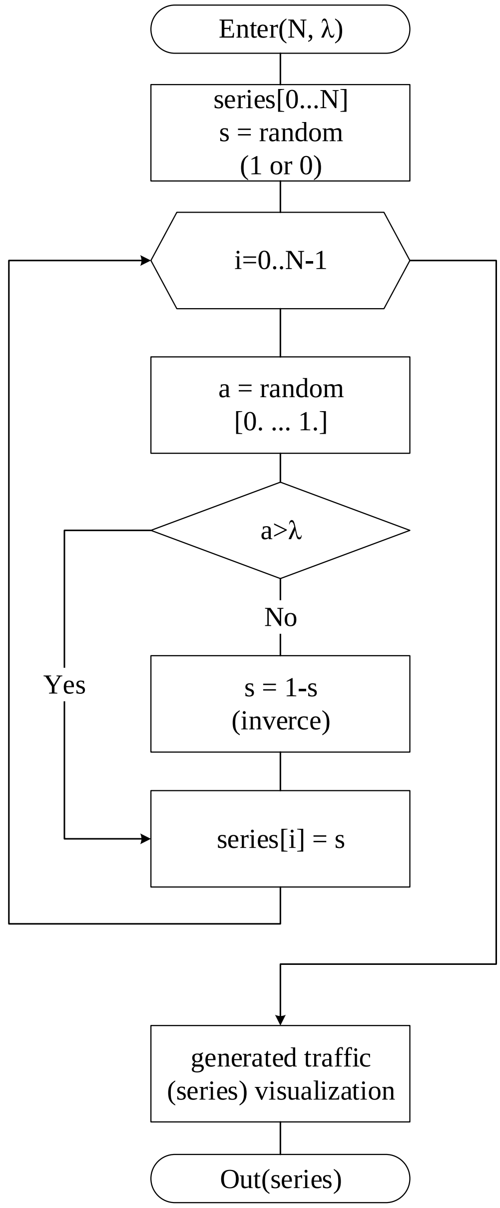

Figure 4: Algorithm for generating fractal binary traffic

Generating traffic with intensity ½ begins with setting the probability of maintaining the state λ. The algorithm contains a variable to save the previous state. The cycle repeats the generation of a pseudo-random number from the range [0; 1) with a uniform distribution, for which a comparison is made with a given probability λ. When passing the test for comparison, the state is preserved, and the output is given the value of the previous state, otherwise the state changes to the opposite. It is expected that the generated binary traffic according to this algorithm has a controlled fractal dimension according to relation (14).

In order to determine the fractal properties of the binary series obtained by the generator from Figs. 3, an experimental measurement of the Hurst exponent by R/S analysis was performed, the results of which are shown in Fig. 5.

Fig. 5 contains the results of the analysis of binary traffic with intensity $p_1 = 0.5$. To do this, it suffices to fulfill the condition of equality of transition probabilities $\lambda = \lambda_0 = \lambda_3$, or equivalent to $\lambda_1 = \lambda_2$. This indicator is responsible for the persistence of the time series and affects its fractal properties. Also, to calculate the Hurst exponent H, need to select a series of random wanderings – cumulative sums. The number of steps of the cumulative sums is shown by a separate axis L.

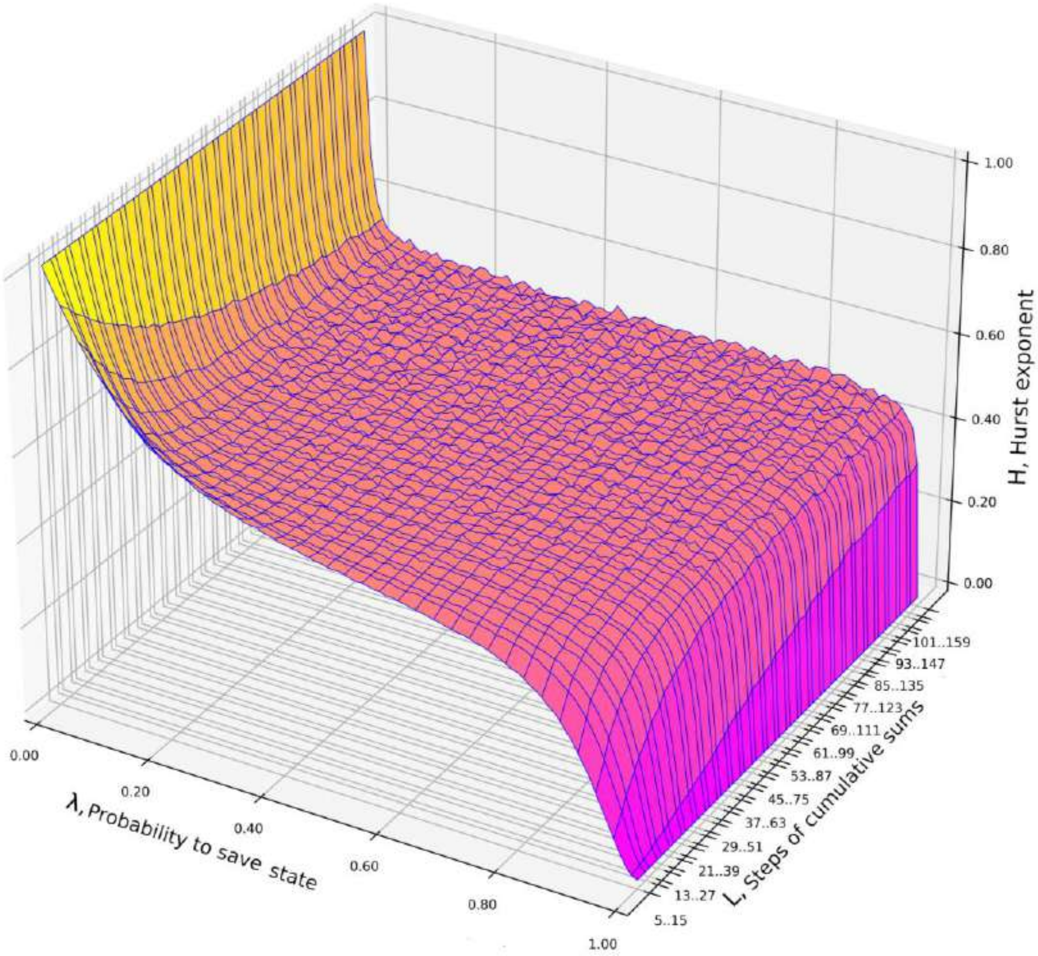

Figure 5: The results of experimental measurement of the Hurst exponent *H*

The graph shows that as the length of the cumulative sums decreases, the graph goes to a straight line that connects the unit and zero values of the Hurst exponent. Conversely, if the cumulative sums are long enough, the Hurst exponent goes to a value of 0.5 without reflecting the persistence of the time series.

The Hurst exponent was calculated according to the R/S analysis algorithm, which is implemented by the following program code [31]:

```
# Input data:
# ts - array with series data
# lasg - array with shifts in cumulative data
ts = numpy.cumsum(ts)
tau = [numpy.sqrt(numpy.subtract(ts[lag:], ts[:-lag]).std()) for lag in lags]
m = numpy.polyfit(numpy.log(lags), numpy.log(tau), 1)
hurst = m[0]*2.0
# output data: hurst - Hurst exponent
```

The algorithm used for the experiment is shown in Fig. 6. The algorithm takes at the input the sequence to be analyzed, and *lags* – is the value of shifts in the calculation of cumulative sums. For each point of the graph, which corresponds to the probability of maintaining the state and range of lengths of cumulative sums, 100 times a random series of 10000 values is generated, for which Hurst exponents are calculated with the specified lengths of cumulative sums. The graph shows the average value for these 100 measurements.

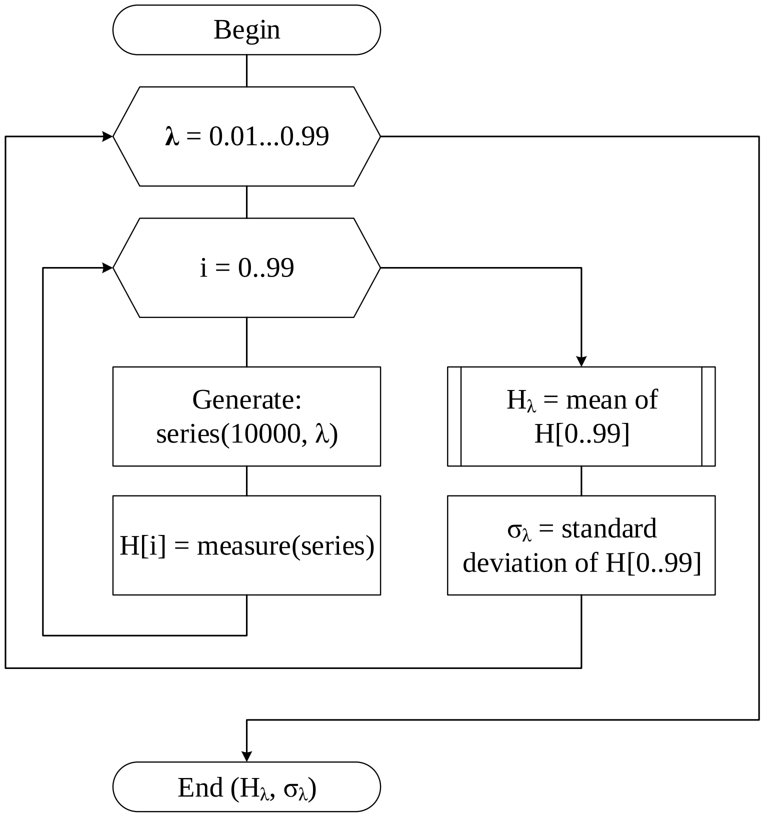

Figure 6: Algorithm for plotting changes in the Hurst exponent

Boundary results need better clarity. In Fig. 7 shows the result of using the minimum cumulative sums in 1 and 3 consecutive values:

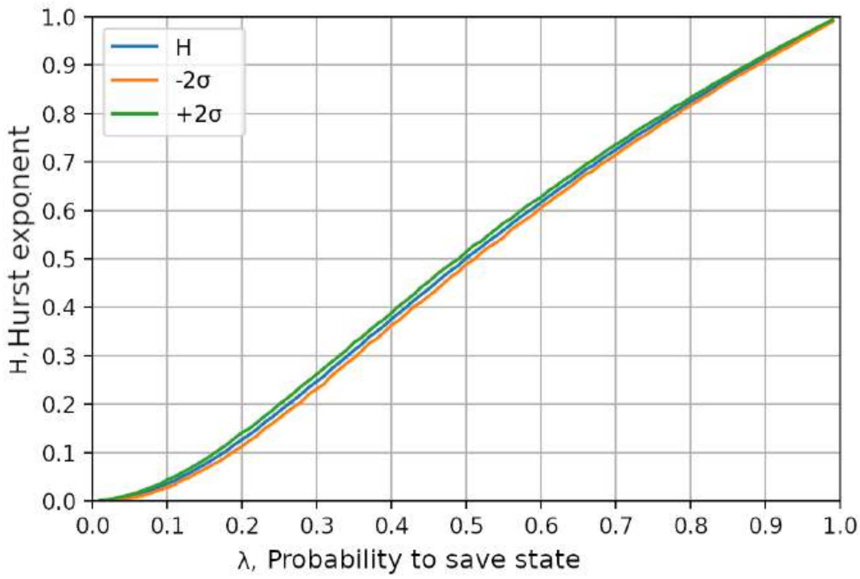

Figure 7: The results of experimental measurements of the Hurst exponent for cumulative sums in 1, 3 samples with deviations of 2σ

As a result, we have a curve that is an approximation of the linear relationship between the probability of maintaining the previous state λ and the Hurst exponent. Also, the markup of deviations shows a small discrepancy between the results from experiment to experiment. However, this is the result of a deliberately election long series of experiments in order to obtain reliable experimental results.

The following Fig. 8 is a representation of another extreme position (in Fig. 5, this trend is developing "deep"), when the cumulative sums are long:

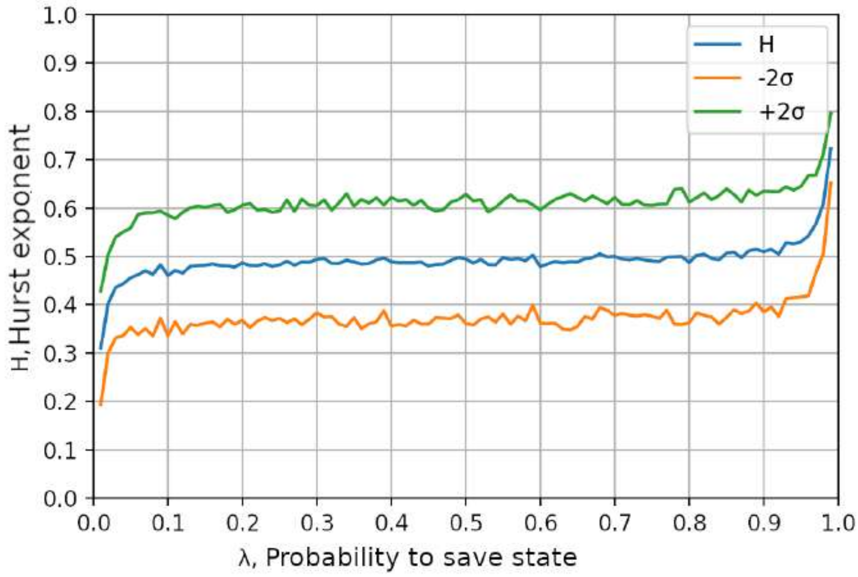

Figure 8: The results of experimental measurements of the Hurst exponent for cumulative sums in 100, 120, 150 samples with deviations of 2σ

Here it is obvious that there is a significant deviation in some measurements, which affected the bandwidth of 2σ. This indicates that the measurement of the Hurst exponent is extremely inaccurate. In this case, in a wide range of probabilities of preservation of the previous state, about [0.1; 0.9], R/S analysis showed that the series is random.

Comparison of graphs in Fig. 7 and Fig. 8, as a result of the analysis of the results of Fig. 5, leads to the conclusion that the result depends on the user's choice of parameters, in particular the length of the series or the lengths of the cumulative sums.

It is proposed to use the Probabilistic approach, which is based on direct measurement of the probability of state change from 0 to 1 (r0) and from state 1 to state 0 (r1) according to the formula for determining fractal dimension (14), which is implemented in Python 3 programming language:

```
r0 = (series[0:-2] < series[1:-1]).sum()/(series[0:-2]<1).sum()
r1 = (series[0:-2] > series[1:-1]).sum()/(series[0:-2]>0).sum()
d = 1-(r1*(1-r0)*math.log(1.0-r0)+r0*(1-r1)*math.log(1.0-r1))/(2.0*r0*r1)
H = 2.0-d
```

where, series is a one-dimensional array containing a binary time series. Using this code to determine the fractal properties of the time series is shown in Fig. 9.

The results of the research of the quality of the Probabilistic method and its comparison with the well-known method of R/S analysis are shown in Table 1.

To construct each row of the table, 100 generations of a binary sequence with a length of 10000 elements with the corresponding value of λ were performed. For each of them, the Hurst exponent H was determined by R/S analysis and by based on an estimate of the probability of transition from 1 to 0 and from 0 to 1 (Probabilistic method, PM).

Thanks to hundreds of independent tests, it is possible to determine the average Hurst exponent (15) standard deviation σ (16) and percentage of deviation (17):

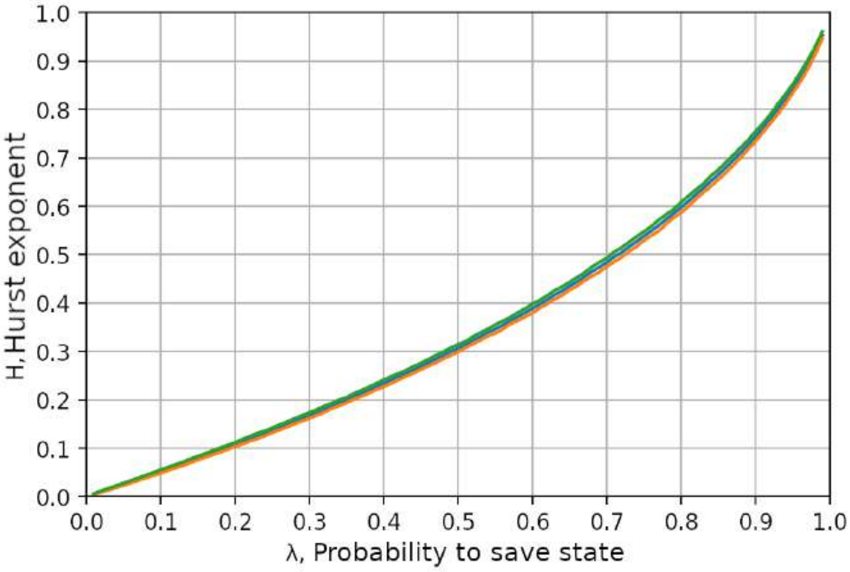

Figure 9: The results of the experimental probabilistic estimation of the Hurst exponent with deviations of 2σ

Table 1
The results of work quality comparison of methods for fractal dimension definition – the known R/S analysis method and proposed Probabilistic method (PM)

№	λ	H, R/S	H, PM	σ, R/S	σ, PM	Deviations for R/S, %	Deviations for PM, %
1	0,01	0,020862173	0,004996201	0,001825691	0,000435045	8,751204779	8,707522638
2	0,06	0,115130822	0,030678064	0,005365203	0,001102508	4,660093493	3,593801913
3	0,11	0,197424495	0,057299125	0,006720639	0,001636951	3,404157102	2,856851724
4	0,16	0,266660733	0,084636434	0,008166731	0,001946542	3,062592345	2,299887334
5	0,21	0,321224610	0,113072840	0,009107072	0,002310433	2,835110483	2,043314238
6	0,26	0,367964061	0,143068708	0,009842329	0,002886502	2,674807286	2,017564092
7	0,31	0,404989816	0,173980449	0,010944700	0,003172788	2,702463230	1,823646757
8	0,36	0,435730431	0,206981323	0,011123077	0,003080814	2,552742787	1,488450246
9	0,41	0,462658106	0,240125774	0,011055708	0,003745437	2,389606470	1,559781461
10	0,46	0,485639655	0,277501570	0,009411981	0,003754746	1,938058694	1,353054180
11	0,51	0,502866839	0,313936619	0,011826159	0,004215003	2,351747645	1,342628967
12	0,56	0,520859605	0,354596924	0,011126878	0,004295506	2,136252974	1,211377335
13	0,61	0,540472278	0,398324387	0,010182414	0,004817891	1,883984610	1,209539751
14	0,66	0,561685180	0,444444295	0,012015554	0,004776983	2,139197425	1,074821662
15	0,71	0,587349580	0,494695839	0,010646189	0,004218451	1,812581414	0,852736347
16	0,76	0,619086392	0,549212501	0,010121335	0,005005190	1,634882623	0,911339444
17	0,81	0,667516915	0,609945932	0,010737486	0,005076275	1,608571380	0,832250120
18	0,86	0,737425650	0,680389339	0,009941918	0,004865064	1,348192641	0,715041204
19	0,91	0,836641573	0,762035440	0,010570512	0,004452206	1,263445799	0,584251869
20	0,96	0,949146972	0,866270314	0,007584074	0,005016391	0,799041091	0,579079216
			Averages:	0,009416	0,003541	2,597437	1,852847

To build each row of the table, 100 binary sequence generations with a length of 10000 elements with an appropriate value of λ were performed. There are σ – is the standard deviation σ (16) and Deviations – is the percentage of deviation (17).

The constructed table allows to estimate the accuracy of the found Hurst exponents by last two columns (Table 1). According to these columns, the relative deviation for the Probabilistic method is smaller. This is because R/S analysis requires to estimate the “scatter” of R, which is the cumulative sum of several elements of the studied sequence. It is clear that such a value requires more measurements of traffic values and requires estimating the values with the appropriate accuracy, i.e. more experiments.

$$H = \frac{1}{100}\sum_{i=1}^{100} H_i, \tag{15}$$

$$\sigma = \sqrt{\frac{1}{100}\sum_{i=1}^{100} (H_i - H)^2}, \tag{16}$$

$$\text{Deviations} = \frac{\sigma}{H} \cdot 100\%. \tag{17}$$

R/S analysis gives different values for different lags of the cumulative sum and the larger the step of the cumulative sum, the more different the value of the Hurst exponents determined by it. This indicates the ambiguity of the method and the possibility of manipulating the results by choosing the size of the cumulative sums. The proposed Probabilistic method does not have this disadvantage. The developed method is unambiguous and does not depend on the period of consideration, because it does not use cumulative sums.

Also, as can be seen from Table 1, developed Probabilistic method has a lower percentage deviation from the mean value of the Hurst exponents, namely 1.8% as opposed to the R/S analysis, which has a deviation 2.5%. Therefore, the proposed method has a greater accuracy in determining the fractal dimension.

3. Conclusions

This experimental study investigates the quality of work of the previously proposed method of determining the fractal dimension of network traffic by its probabilistic properties. Also, the results of the experimental comparison of the work quality of the proposed Probabilistic method compared to the method based on R/S analysis.

A simulated binary time series using Markov process theory was used to conduct the experiments. To simulate network traffic a binary time series was created with persistence regulated by setting up the probabilities of changing states “0” and “1” to the opposite ones.

As shown by experiments on simulated network traffic, the proposed Probabilistic method has greater accuracy and unambiguity of results regardless of the length of the studied series in contrast to R/S analysis.

R/S analysis gives different values for different sizes of the cumulative sum, and the larger the step of the cumulative sum, the more diverges the value of the Hurst exponent determined by it. This indicates the ambiguity of the method and the possibility of manipulating the results by choosing the different lags of the cumulative sums. The proposed Probabilistic method does not have this disadvantage. The developed method is unambiguous and does not depend on the period of consideration, because it does not use cumulative sums.

Also, the developed method has a lower percentage of deviation from the mean value of the Hurst exponent and therefore has higher accuracy in determining the fractal dimension – R/S analysis has a deviation of 2.5%, and the developed method 1.8%.

4. References

[1] C. Ma, G. Dai, J. Zhou, Short-Term Traffic Flow Prediction for Urban Road Sections Based on Time Series Analysis and LSTM_BILSTM Method, in IEEE Transactions on Intelligent Transportation Systems (2021). URL: https://ieeexplore.ieee.org/document/9364926. doi: https://doi.org/10.1109/TITS.2021.3055258.

[2] P. Dymora, M. Mazurek, Influence of Model and Traffic Pattern on Determining the Self-Similarity in IP Networks, Applied Sciences, Vol. 11, Issue 1(190), 2021. URL: https://www.mdpi.com/2076-3417/11/1/190. doi: https://doi.org/10.3390/app11010190.

[3] A. Phinyomark, R. Larracy, E. Scheme, Fractal Analysis of Human Gait Variability via Stride Interval Time Series. Front Physiol, Vol. 11, Article 333, 2020 Apr 15. URL: https://pubmed.ncbi.nlm.nih.gov/32351405. doi: https://doi.org/10.3389/fphys.2020.00333.

[4] G. Millán, Traffic Flows Analysis in High-Speed Computer Networks Using Time Series, arXiv preprint arXiv:2103.03984 [stat.AP], 2021. URL: https://arxiv.org/abs/2103.03984. doi: https://doi.org/10.48550/arXiv.2103.03984.

[5] D. Zhukov, J. Perova, V. Kalinin, Description of the Distribution Law and Non-Linear Dynamics of Growth of Comments Number in News and Blogs Based on the Fokker-Planck Equation, Mathematics, Vol. 10 Issue 6, 2022. URL: https://www.mdpi.com/2227-7390/10/6/989/htm.

[6] J. B. Bassingthwaighte, G. M. Raymond, Evaluating Rescaled Range Analysis for Time Series, Annals of Biomedical Engineering, Vol. 22, 1994, pp. 432-444. URL: https://link.springer.com/article/10.1007/BF02368250#citeas. doi: https://doi.org/10.1007/BF02368250.

[7] Y. Chen, Random walk: statistical data analysis and application in finance, in Proc. SPIE 12259, 2nd International Conference on Applied Mathematics, Modelling, and Intelligent Computing (CAMMIC 2022), 122590E (2022). doi: https://doi.org/10.1117/12.2639464.

[8] Y. Hu, F. Xiao, A novel method for forecasting time series based on directed visibility graph and improved random walk, Physica A: Statistical Mechanics and its Applications, Vol. 594, 2022. doi: https://doi.org/10.1016/j.physa.2022.127029.

[9] S. U. Rehman, B. Said, Analysis of Investor Overreaction Effect and Random Walk: A Case Study of Pakistan Stock Exchange, Vol. 5, No. 1, 2019. doi: https://doi.org/10.31529/sjms.2018.5.1.1.

[10] M. A. Nia, R. E. Atani, B. Fabian, E. Babulak, On detecting unidentified network traffic using pattern-based random walk, Security and Communication Networks, Vol. 9, No. 16, 2016, pp. 3509-3526. doi: https://doi.org/10.1002/sec.1557.

[11] M. Rhanoui, S. Yousfi, M. Mikram, H. Merizak, Forecasting Financial Budget Time Series: ARIMA Random Walk vs LSTM Neural Network, IAES International Journal of Artificial Intelligence, Vol. 8, No. 4, 2019, pp. 317-327. doi: http://doi.org/10.11591/ijai.v8.i4.pp317-327.

[12] X. Qi, H. Sheng, Hurst Index Analysis of Social Electricity Consumption Change Trend Based on R/S Analysis, in IOP Conference Series: Materials Science and Engineering (2020). URL: http://dx.doi.org/10.1088/1757-899X/750/1/012150.

[13] A. Kuchansky, A. Biloshchytskyi, S. Bronin, S. Biloshchytska, Y. Andrashko, Use of the Fractal Analysis of Non-stationary Time Series in Mobile Foreign Exchange Trading for M-Learning, Advances in Intelligent Systems and Computing, Vol. 1192, Springer, Cham (2021). doi: https://doi.org/10.1007/978-3-030-49932-7_88.

[14] M. S. Raimundo, J. Jr. Okamoto, Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX Securities, International Journal of Modeling and Optimization, Vol. 8, No. 2, 2018, pp. 116-124.

[15] H. Danylchuk, O. Kovtun, L. Kibalnyk, O. Sysoiev, Monitoring and Modelling of Cryptocurrency Trend Resistance by Recurrent and R/S-Analysis, 2020. URL: https://ssrn.com/abstract=3618074.

[16] S. Wu, An User Intention Mining Model Based on Fractal Time Series Pattern, Fractals, Vol. 28, No. 08, P. 2040017, 2020. doi: https://doi.org/10.1142/S0218348X20400174.

[17] M. Fernández-Martínez, M. A. Sánchez-Granero, J. E. Trinidad Segovia, I. M. Román-Sánchez, An accurate algorithm to calculate the Hurst exponent of self-similar processes, Physics Letters A, Vol. 378, Issues 32-33, 2014, pp. 2355-2362. doi: https://doi.org/10.1016/j.physleta.2014.06.018.

[18] O. Zmeškal, M. Nežádal, B. Komendová, M. Julínek, T. Bžatek, Fractal analysis of printed structure images, Institute of Physical and Applied Chemistry, of the methods used to perform analysis listed above, Brno University of Technology, Brno, Czech Republic. URL: http://imagesci.fch.vut.cz/download/en07_brat03.pdf.

[19] W. Leland, M. Taqqu, W. Willinger, On the self-similar nature of IP-trafic, IEEE/ACM Transactions on Networking, No. 3, 1997, pp. 423-431.

[20] H. A. Kuchuk, Method of research of fractal network traffic, Information processing systems, Vol. 5 (45), Kharkiv University of Reinforced Forces named after Ivan Kozhedub, Kharkiv, 2005, pp. 74-84. (in Ukrainian)
[21] H. A. Kuchuk, O. O. Mozhaiev, O. V. Vorobiov, Analysis and models of self-similar traffic, Aerospace Engineering and Technology, Vol. 9 (35), 2006, pp. 173-180. (in Ukrainian)
[22] Y. Zhao, L. Wu, Comparison and Application of Estimation of Hurst Exponent, Computer Engineering and Applications, No. 16, 2014, pp. 154-158.
[23] V. A. Malinnikov, D. V. Uchaev, Analysis of methods for forming a multifractal measure based on wavelet processing of experimental data, Izvestiya vuzov, Series geodesy and aerial photography, No. 6, 2007, pp. 57-61. (in Russian)
[24] L. Ya. Nych, R. M. Kaminskyi, Determination of Hurst exponent using fractal dimension calculated by cellular method on the example of short times series, Information systems and networks, Vol. 814, Bulletin of the National University of Lviv Polytechnic, 2015, pp. 100-111. (in Ukrainian)
[25] V. Ya. Voropaieva, V. I. Bessarab, V. V. Turupalov, V. V. Chervynskyi, Theory of teletraffic, Donetsk National Technical University, Donetsk, 2011. (in Ukrainian)
[26] I. O. Romanenko, R. M. Zhyvotovskyi, S. M. Petruk, A. V. Shyshatskyi, O. O. Voloshyn, Mathematical model of load distribution in special purpose telecommunication networks, Information processing systems, No. 3(149), 2017, pp. 61-71. (in Ukrainian)
[27] H. Drieieva, O. Smirnov, O. Drieiev, Y. Polishchuk, R. Brzhanov, M. Aleksander, Method of Fractal Traffic Generation by a Model of Generator on the Graph, COAPSN, CEUR-WS, Vol. 2616, Lviv, Ukraine, 2020. URL: http://ceur-ws.org/Vol-2616/paper31.pdf.
[28] H. M. Drieieva, O. M. Drieiev, O. O. Denysenko, Determination of the fractal dimension of a numerical sequence by the probability distribution of the values of its elements, Machinery in agricultural production, industrial engineering, automation, Vol. 31, Kropyvnytskyi, 2018, pp. 119-128. URL: http://nbuv.gov.ua/UJRN/znpkntu_2018_31_17. (in Ukrainian)
[29] Ju. P. Ryzhikov, Algorithmic approach to queuing problems: monograph, Military Space Academy named after A. F. Mozhaisky, Saint Petersburg, 2013. (in Russian)
[30] V. N. Zadorozhnyj, Analytical and simulation studies of large network structures: monograph, Omsk State Technical University Publishing House, Omsk, 2011. (in Russian)
[31] J. Park, How can calculate Hurst exponent in python? (lags parameter issue), 2018. URL: https://quantopian-archive.netlify.app/forum/threads/how-can-calculate-hurst-exponent-in-python-lags-parameter-issue.html.
[32] Ye. Meleshko, O. Drieiev, M. Yakymenko, D. Lysytsia, Developing a model of the dynamics of states of a recommendation system under conditions of profile injection attacks, Eastern-European Journal of Enterprise Technologies, Vol. 4, No. 2(106), 2020, pp. 14-24. URL: https://www.scopus.com/record/display.uri?eid=2-s2.0-85096707995&origin=resultslist.
[33] Ye. Meleshko, L. Raskin, S. Semenov, O. Sira, Methodology of probabilistic analysis of state dynamics of multi-dimensional semi-Markov dynamic systems, Eastern-European Journal of Enterprise Technologies, Vol. 6, No. 4(102), 2019, pp. 6-13. URL: https://www.scopus.com/record/display.uri?eid=2-s2.0-85078054250&origin=resultslist.
[34] T. D. Dimitrakos, E. G. Kyriakidis, A semi-Markov decision algorithm for the maintenance of a production system with buffer capacity and continuous repair times, International Journal of Production Economics, Vol. 111(2), 2008, pp. 752-762. doi: https://doi.org/10.1016/j.ijpe.2007.03.010.
[35] Q.-L. Li, J. C. S. Lui, Block-structured supermarket models, Discrete Event Dynamic Systems, Vol. 26(2), 2014, pp. 147-182. doi: https:// doi.org/10.1007/s10626-014-0199-1.
[36] H. Okamura, S. Miyata, T. Dohi, A Markov Decision Process Approach to Dynamic Power Management in a Cluster System, IEEE Access, Vol. 3, 2015, pp. 3039-3047. doi: https://doi.org/10.1109/access.2015.2508601.
[37] Q.-L. Li, Nonlinear Markov processes in big networks. Special Matrices, Vol. 4(1), 2016. doi: https://doi.org/10.1515/spma-2016-0019.
[38] E. A. Feinberg, F. Yang, Optimal pricing for a GI/M/k/N queue with several customer types and holding costs, Queueing Systems, Vol. 82(1-2), 2015, pp. 103-120. doi: https://doi.org/10.1007/s11134-015-9457-7