



Researching Measured and Modeled Traffic with Self-similar Properties for Ateb-Modeling Method Improvement

Olga Fedevych¹(✉), Ivanna Dronyuk¹, and Danylo Lizanets²

¹ ACS Department, Lviv Polytechnic National University,
12 Bandera Street, Lviv, Ukraine

olha.fedevych@gmail.com, ivanna.droniuk@gmail.com

² Division of Microengineering and Photovoltaics,
Wrocław University of Science and Technology,
27 Wybrzeże Wyspiańskiego, Wrocław, Poland
danylo.lizanets@pwr.edu.pl

<http://www.lp.edu.ua/asu>

<http://www-old.wemif.pwr.wroc.pl/zmf/>

Abstract. In this paper, improvements in the traffic behavior modeling method based on the consideration of traffic self-similarity properties was proposed. Two methods for calculating the Hurst parameter were used: R/S plot method and V/S analysis method. The research was carried out on real traffic network data, collected from LPNU ACS Department. The selected methods were implemented in the corresponding software. The work of the methods was illustrated in experimental calculations. Also, the results was shown in the form of both tables and plots.

Keywords: Computer network traffic · Ateb-function
Traffic modeling · Self-similar properties · Hurst parameter

1 Introduction

One of the biggest challenges nowadays is the gradual shift from the Internet of Devices concept to the new Internet of Things concept, given the aspirations of those who developed these concepts to boost their user-friendliness and the growing demand for dynamic scalability of networks that differ by scale, type or functionality. Also, the revolutionary influence on traffic parameters changing has an alteration of its internal structure. Based on today's Internet research, provided by Cisco, mobile traffic on 75% consists of the video traffic [1].

According to the forecasts of the same company, the share of video content in the future will only be increasing. Another factor that significantly influences the structure of traffic is the widespread introduction of cloud technologies. The cloud computing paradigm has become the foundation of the modern economy by offering subscription-based services anytime, anywhere with services paid by

users. The transition of many technologies to distributed computing, software-managed networks and data processing at the network level in nodes of network equipment are also among the modern trends that open up new opportunities for traffic management [2, 3].

At the same time, computer networks have certain disadvantages, such as: the complexity of network management, the high cost of network equipment, the lack of efficient use of the channel through the transmission of a large amount of information to manage the network instead of useful traffic. Multimedia traffic samples were selected for the research.

The purpose of this development is to expand the functionality of the previously [4] proposed method of modeling the behavior of traffic based on the consideration of self-similar properties of traffic. In this article, under the term of traffic modeling (forecasting), the conception of the process of collecting data for real traffic and based on these data and the corresponding formulas of the mathematical model, the calculation of the following traffic values was considered.

2 Simulation of Traffic Behavior for Model Construction

One of the most pressing problems of traffic research is adequate consideration of its features. Studies show that the traffic of modern telecommunication networks has a special structure that does not allow to use classical methods to design network equipment based on Markov models and Erlang's formulas [4]. This is the effect of self-similarity of traffic, which leads to the ripples of its receipt. This phenomenon significantly degrades the characteristics (increases losses, delays, jitter) while passing self-similar traffic through the network equipment, so the study of this indicator will predict and reduce the impact of such undesirable factors.

Research of various types of network traffic over the past fifteen years has shown that network traffic is self-similar or fractal in its nature. It follows from this that the classical methods that are used for modeling of network systems based on the use of Poisson flows [5], do not give a complete and accurate picture of what is happening online. In [4] a mathematical model for modeling traffic behaviour in computer networks based on differential equations describing the oscillatory motion was proposed.

In addition, self-similar traffic has a special structure, stored at multiple scaling. When passing traffic on the network, as a rule, there is a certain number of pulsations with a relatively small average traffic level. In practice, this is implemented in such a way that packets, at high speed of their network movement, arrive at the site not separately, but as a whole package, which can lead to their losses due to limited buffer, calculated according to classical techniques.

Many contemporary researchers note [6] that combining of traffic from multiple variable ON/OFF sources increases self-similar properties of traffic. Traffic becomes highly autocorrelated with long-term dependence. The unification of a large number of data sources is characterized by a syndrome of infinite dispersion and, as a result, gives a self-similar unified network traffic that seeks for a

fractal Brownian motion. In addition, the research of various sources of traffic indicates that the very variable behavior of traffic is a function inherent in the client/server architecture [7].

There is no single causative factor that causes self-similarity. Different correlations that exist in self-similar network traffic and that take effect on different time scales may occur for various reasons, changing the characteristics at specific time scales.

3 Methods of Traffic Self-similar Properties Determination

To analyze the self-similar properties of traffic, the Hurst coefficient was selected. At the same time, the Hurst method, being robust, can reveal such properties as clustering, the tendency to be in the direction of the trend, strong aftereffect, separate memory, rapid change of sequential values, fractality, the presence of periodic and nonperiodic cycles, the ability to distinguish “stochastic” and “chaotic” nature of noise in the statistical data.

The value of the Hurst coefficient can be found in several methods [8–10]:

1. Variance method. Logarithmic selective dispersion in comparison with the level of aggregation should be a straight line with a slope of more than -1 .
2. R/S plot method. The logarithmic sample of R/S statistics compared with the number of points in the aggregated series is a straight line with a slope.
3. Absolute moments method. The aggregated series $X(m)$ determines the use of different sizes of blocks m .
4. Variance of residuals method. Logarithmic aggregation in comparison with the average dispersion level of the residues should be a straight line with a slope of $H/2$.
5. Abrý-Veitch estimation method. To estimate the H parameter, the energy of the series on different scales is used.
6. Whittle estimation method is based on the minimization of the likelihood function, which applies to the time series period, gives the assessment of H and dependence on the confidence interval.
7. V/S analysis, described in detail in [11].

All seven mentioned methods were tested for realization of scientific researches. During the computational experiments, R/S plot method and V/S analysis were selected as the best methods for determining the proximity of the Hurst coefficient for modeled and real traffic. These methods were chosen among others due to the ease of their program implementation.

In addition to the key contribution of Hurst to the development of the theory of R/S-method and its application in practice, a significant role was played by Mandelbrot [12]. According to this method, not the data that compose a dynamic time series are analyzed, but the scale of the amount of deviations of these data from the mean arithmetic, normalized by dividing on standard deviation. The sum of these deviations is calculated for different periods of time

(or for different number of successive observation points), which act as a scale of measurement.

The main difference between R/S estimation method and other existing statistical methods for analyzing of time series is that this method includes the time direction in its analysis, while other known methods for this time are invariant.

The Hurst H coefficient is described by the empirical relation

$$\frac{R}{S} = N^H \quad (1)$$

where R – the range of deviations values of the series x , S – standard deviation x , N – number of observations.

Express the Hurst H factor as:

$$H = \frac{\log \frac{R}{S}}{\log N} \quad (2)$$

Let $\bar{x}(N)$ – average value of random variable

$$\bar{x}(N) = \frac{1}{N} \sum_{i=1}^N x(t_i) \quad (3)$$

The standard deviation is determined from the formula

$$S(N) = \sqrt{\frac{1}{N} \sum_{i=1}^N [x(t_i) - \bar{x}(N)]^2} \quad (4)$$

Denote by the following formula the accumulated deviation of the values of the random variable $x(t)$ from its mean value $\bar{x}(N)$ for time t :

$$X(t, N) = \sum_{u=1}^t [x(u) - \bar{x}(N)] \quad (5)$$

Difference between maximum and minimum values $X(t, N)$ is called a scope

$$R(N) = \max X(t, N) - \min X(t, N) \quad (6)$$

where $1 \leq t \leq N$.

Calculated deviations (4), (5) are substituted in the formula (1) and there is a Hurst coefficient.

Two methods for calculating the Hurst parameter are implemented in the developed software module. Conducted calculations of the Hurst parameter are implemented for the real traffic values and the modeled traffic values based on the R/S method and the V/S method described in detail in [11, 12].

4 Development of Software Solution

This work envisaged the creation of two separate software solutions that collectively carried out modeling and traffic analysis in order to model its behavior. First, a solution for analysis of traffic samples and the implementation of the short-term behaviour modeling method was developed, followed by a separate module for analyzing and calculating the Hurst parameter in the collected traffic data.

To carry out these studies, an additional software module has been implemented to calculate the Hurst parameter for different time intervals, samples of traffic, as well as to calculate the Hurst parameter for pre-recorded modeled values of the short-term behaviour modeling method of traffic values.

Figure 1 shows the main window of the software module for calculating the Hurst parameter.

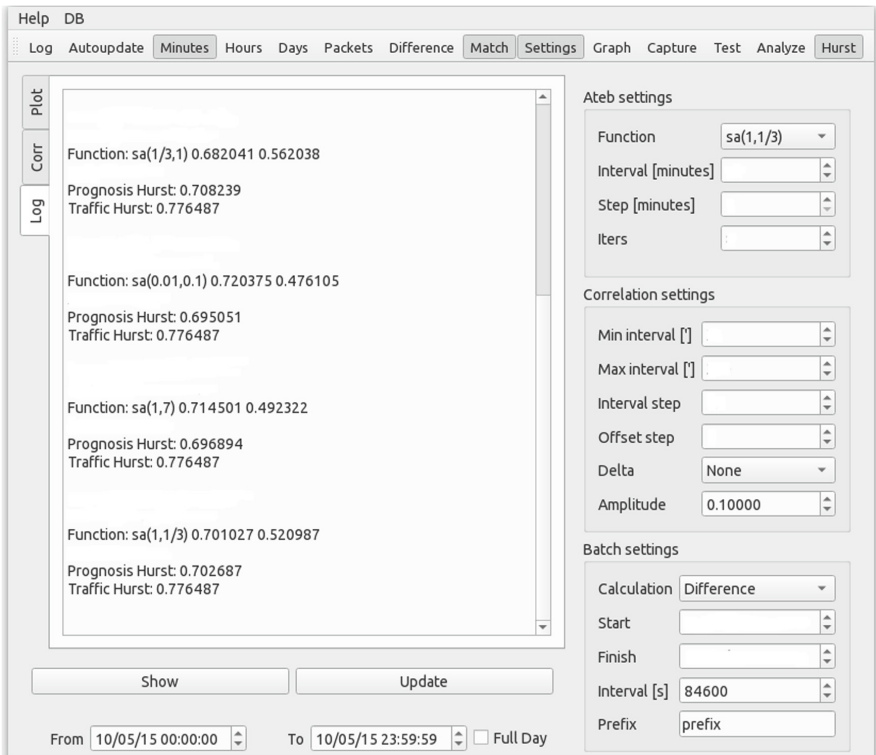


Fig. 1. Program module interface for Hurst parameter calculation

5 Analysis of Experimental Calculations

There is an ability to open captured traffic in the analyzer program, analyze and calculate the properties of the traffic. Opening of pcap files does not require much resources since the reading of the file occurs in the thread, though opening of the same file in other programs may require a lot of resources.

To test the traffic analysis software, data was captured in two different programs - Wireshark and CommView. Data capture was carried out in different places and different devices: a wireless adapter, a network card, a virtual network adapter.

The results were analyzed by a software solution for traffic analysis. It was established that since the captured data is exported to the standardized format of the pcap file, there were no errors in the reading. Traffic properties were determined correctly.

Developed software was also tested on different machines with different operating systems. After the testing, the software continued to work in as stable and predictable manner as before.

5.1 Description of Conducted Experiments

Traffic was captured at the point where it had already been merged and sent (received) from the Internet. The network of the ACS department contains about 20 working computers, loaded on average from 8:30 to 17:30; about 4 computers are loaded until 21:00, and 3 computer classes with 32 workstations are loaded on average from 8:30 to 16:00. The measurements were carried out during the month on each working day from 9.00 to 13.00 o'clock, at an average data transfer rate of 150 Mbit/sec. However, for visualization of the conducted research, three days in a row were randomly chosen.

The collected data was analyzed by the developed software solution in terms of traffic; the results of the analysis are shown in Table 1.

Duration of data collection differs in samples, which may slightly affect the Hurst coefficient. The simulation time is chosen identically to the time of data capture on a network.

In the previously proposed method of modeling, the property of the alternating period of the Ateb-functions was used to select modeling parameters.

In [13] is proved that the Ateb-cosine and the Ateb-sine are periodic with period $2\Pi(m, n)$, where $\Pi(m, n)$ is shown by the formula

$$\Pi(m, n) = \frac{\Gamma\left(\frac{1}{n+1}\right) \Gamma\left(\frac{1}{m+1}\right)}{\Gamma\left(\frac{1}{n+1} + \frac{1}{m+1}\right)} \quad (7)$$

where $\Gamma\left(\frac{1}{n+1}\right), \Gamma\left(\frac{1}{m+1}\right)$ – gamma function.

For the modeling, the parameters n, m were selected in such a way that the modeling period corresponded to the period of real traffic. In order to apply the self-similarity property to improve the traffic behavior modeling, the Hurst parameter for Ateb-functions in the half-period was calculated by the formula (7). These calculations are presented in Table 1. Parameter t - modeling time, s.

Table 1. Dependency of Hurst parameter from Ateb-sine parameter

Function type	Hurst parameter value $H(m, n)$
sa(1, 1, t)	0.861452
sa(1/7, 3, t)	0.884723
sa(1/5, 1, t)	0.873587
sa(1/3, 1, t)	0.875528
sa(0.01, 0.1, t)	0.823017
sa(1, 7, t)	0.879125
sa(1, 1/3, t)	0.855861

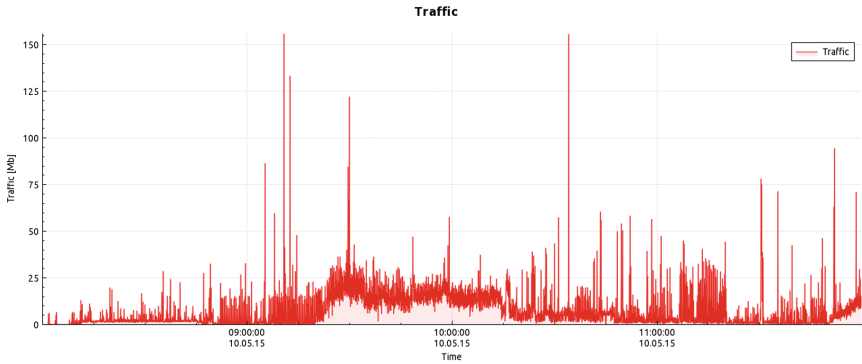


Fig. 2. The example of investigated traffic sample

Not only the proximity of the period of real and modeled traffic, but also the proximity of the Hurst parameter as the second factor are taken into account when choosing the values of the parameters n, m , when implementing modeling in the improved method (Fig. 2).

$$\min_{n, m} |H(m, n) - H_{tr}| \rightarrow 0 \quad (8)$$

where $H(m, n)$ - Hurst parameter for Ateb-function, H_{tr} - Hurst parameter for real traffic.

The traffic modeling is implemented using the method with the addition of the delta function after the parameters n, m are chosen with the consideration of the period and self-similarity using the formula (8).

5.2 Evaluation and Analysis of Computational Results

The obtained data was analyzed and the Hurst coefficient was obtained. The Hurst coefficient is consistently greater than 0.5 for all collected and simulated traffic data, which indicates a self-similarity effect appearance.

The Hurst coefficient (H) is a measure of the stability of a statistical phenomenon or measure of the duration of the long-term dependence of the process. The closer the value of H is to 1, the higher the degree of stability of long-term dependence and self-similarity is.

Thus, in the follow-up of the long-term observations, there is the possibility to learn the tendency of traffic and to be able to model network traffic load. If $0 \leq H \leq 0.5$, then the collected data is trend-resistant, if $0.5 \leq H \leq 1$, the series is the self-similar and trend-stable and thus the tendency of its changes can be predicted.

The calculation were carried out for the R/S method (Tables 1, 2 and 3, Fig. 3), as well as for the V/S method with different values of the amount of division of the time interval k , where $k = 100$ (Tables 4 and 5, Fig. 4) and $k = 300$ (Tables 6 and 7, Fig. 5).

Table 2. Hurst parameter calculation results for real traffic using R/S analysis

Hurst parameter value (traffic for 10.05.2017)	Hurst parameter value (traffic for 11.05.2017)	Traffic for Hurst parameter value (12.05.2017)
0,776487	0,755443	0,578647

Table 3. Hurst parameter calculation results for modeled traffic using R/S analysis

Function type	Hurst parameter value (10.05.2017)	Hurst parameter value (11.05.2017)	Hurst parameter value (12.05.2017)
sa(1, 1, t)	0,702865	0,757899	0,704181
sa(1/7, 3, t)	0,706573	0,762734	0,699697
sa(1/5, 1, t)	0,70686	0,759841	0,692659
sa(1/3, 1, t)	0,708239	0,749252	0,701112
sa(0.01, 0.1, t)	0,695051	0,745324	0,697232
sa(1, 7, t)	0,696894	0,755396	0,699143
sa(1, 1/3, t)	0,702687	0,762055	0,708672

In Figs. 3 and 4, the first column of the chart corresponds to the selected data of real traffic sample, the Hurst coefficient values of which are shown in Tables 3 and 4 in the first line. Comparing the results of calculations for real and modeled traffic shows a larger spread of computing results for real traffic samples compared with the model traffic samples. The comparison of the calculations by

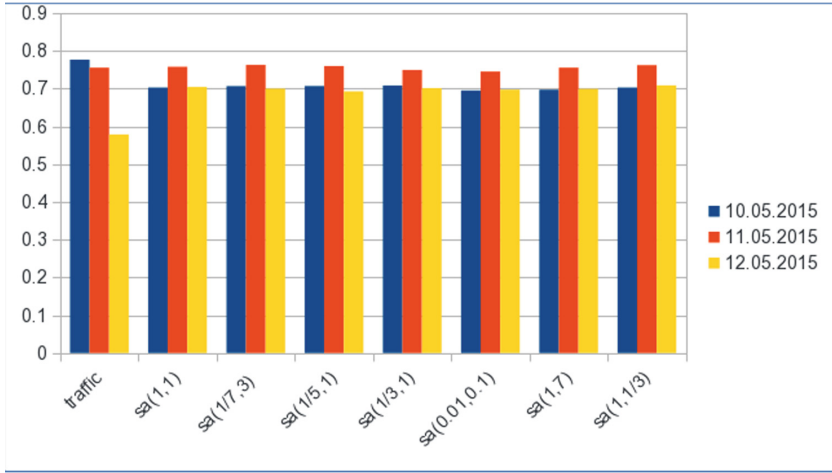


Fig. 3. Hurst parameter values for real and modeled traffic samples using R/S analysis.

Table 4. Hurst parameter calculation results for real traffic using V/S analysis for $k = 100$

Hurst parameter value (traffic for 10.05.2017)	Hurst parameter value (traffic for 11.05.2017)	Hurst parameter value (traffic for 12.05.2017)
0,472811	0,380671	0,313827

Table 5. Hurst parameter calculation results for modeled traffic using V/S analysis for $k = 100$

Function type	Hurst parameter value (10.05.2017)	Hurst parameter value (11.05.2017)	Hurst parameter value (12.05.2017)
sa(1, 1, t)	0,608668	0,605909	0,605868
sa(1/7, 3, t)	0,600347	0,595662	0,600275
sa(1/5, 1, t)	0,600398	0,596081	0,60033
sa(1/3, 1, t)	0,599482	0,596307	0,599413
sa(0.01, 0.1, t)	0,609914	0,607638	0,610486
sa(1, 7, t)	0,60995	0,60788	0,610553
sa(1, 1/3, t)	0,607961	0,607886	0,60801

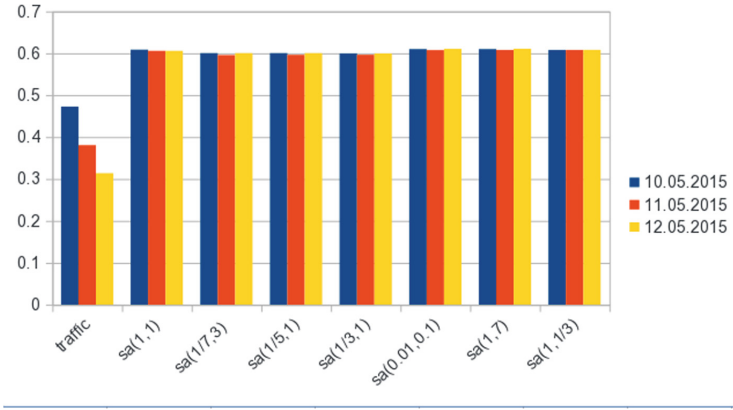


Fig. 4. Hurst parameter values for real and modeled traffic samples using V/S analysis for $k = 100$

Table 6. Hurst parameter calculation results for real traffic using V/S analysis for $k = 300$

Name	Hurst parameter value (10.05.2017)	Hurst parameter value (11.05.2017)	Hurst parameter value (12.05.2017)
Traffic	0,510714	0,442446	0,346464

Table 7. Hurst parameter calculation results for modeled traffic using V/S analysis for $k = 300$

Function type	Hurst parameter value (10.05.2017)	Hurst parameter value (11.05.2017)	Hurst parameter value (12.05.2017)
Traffic	0,510714	0,442446	0,346464
sa(1, 1, t)	0,642291	0,642654	0,64264
sa(1/7, 3, t)	0,636954	0,63623	0,636924
sa(1/5, 1, t)	0,637107	0,636658	0,637082
sa(1/3, 1, t)	0,636594	0,636614	0,636561
sa(0.01, 0.1, t)	0,642945	0,642956	0,642529
sa(1, 7, t)	0,643276	0,643293	0,642884
sa(1, 1/3, t)	0,642028	0,642069	0,642105

R/S method (Tables 1, 2 and 3, Fig. 3) and the V/S method (Tables 4, 5, 6 and 7, Figs. 4 and 5) shows the higher values of the calculated Hurst coefficient by about 0.7 according to the R/S method than by the V/S method is about 0.6.

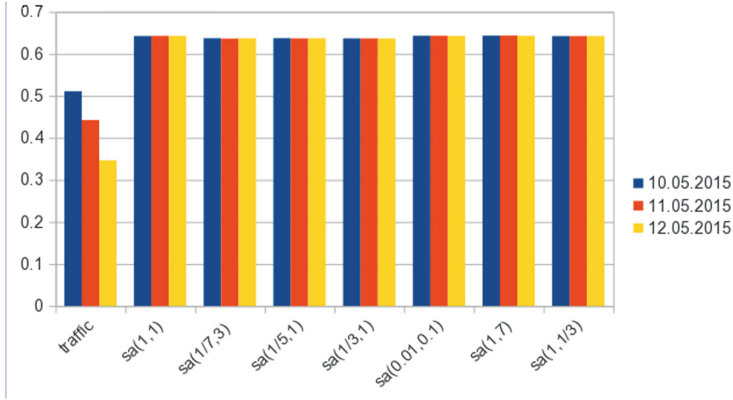


Fig. 5. Hurst parameter values for real and modeled traffic samples using V/S analysis for $k = 300$

Consequently, it can be concluded that for the studied traffic samples, the R/S method gives higher values of the Hurst coefficient.

The comparison of calculations of the Hurst coefficient for real traffic samples using the V/S method for different values of the coefficient k (Table 4, $k = 100$) and (Table 6, $k = 300$) shows that an increase of k value corresponds to increasing of the Hurst coefficient value. On the contrary, for modeled traffic samples, the increase of k value corresponds to reducing of the Hurst coefficient value (Table 5, $k = 100$) and (Table 7, $k = 300$).

6 Conclusions

New trends in the development of computer networks create the need for new approaches and strategies for their research, as well as a reappraisal of models designed to address such issues as scalability, elasticity, reliability, security, sustainability and application of traffic behavior patterns.

This paper shows improvements in the traffic behavior modeling method based on the consideration of traffic self-similarity parameters. Two methods for calculating the Hurst parameter were used: R/S plot method and V/S analysis method.

Committed studies have shown that for the traffic that is available in the network of the ACS NULP department, the method that gives the higher value of the Hurst parameter is the R/S method. This was observed in all cases that were considered. In addition, this method has a faster computation time compared to the V/S method in 7 times in average. So it can be concluded that the R/S method can be used for the implementation in the network nodes for determination of the Hurst parameter to improve the Ateb-modeling method, developed before.

The selected methods are implemented in the corresponding software. The work of the methods is illustrated in experimental calculations. The work has a

practical application for studying the self-similar properties of traffic. The growth of traffic volumes, transmitted through the network, as well as a significant increase in the proportion of video files in traffic, causes the use of additional traffic management tools directly at the nodes of the network. In order to manage traffic at a network node and redistribute it to reduce delays, it is efficient to use traffic modeling methods. Thereby, the investigation of self-similarity parameters of traffic helps to increase the effective traffic management in the nodes of the network.

References

1. The Zettabyte Era: Trends and Analysis. <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html>
2. Gelembé, E.: Steps towards self-aware networks. *Commun. ACM* **52**(7), 66–75 (2009). <https://doi.org/10.1109/CIT.2010.508>
3. Toral-Cruz, H., Pathan, A.K., Ramirez Pacheco, J.C.: Accurate modeling of VoIP traffic QoS parameters in current and future networks with multifractal and Markov models. *Math. Comput. Model.* **57**(11–12), 2832–2845 (2013). <https://doi.org/10.1016/j.mcm.2011.12.007>
4. Dronyuk, I., Fedevych, O.: Traffic flows ateb-prediction method with fluctuation modeling using Dirac functions. *Commun. Comput. Inf. Sci.* **718**, 3–13 (2017). https://doi.org/10.1007/978-3-319-59767-6_1
5. Mishura, Y., Ral'chenko, K., Seleznev, O., Shevchenko, G.: Asymptotic properties of drift parameter estimator based on discrete observations of stochastic differential equation driven by fractional brownian motion. In: Korolyuk, V., Limnios, N., Mishura, Y., Sakhno, L., Shevchenko, G. (eds.) *Modern Stochastics and Applications*. SOIA, vol. 90, pp. 303–318. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-03512-3_17
6. Park, K., Willinger, W.: Self-similar network traffic: an overview. In: *Self-Similar Network Traffic and Performance Evaluation*. Wiley (2000). <https://doi.org/10.1002/047120644X.ch1>
7. Demydov, I., Klymash, M., Kharkhalis, Z., Strykhaliuk, B., Komada, P., Shredrejeva, I., Targeusizova, A., Iskakova, A.: The research of the availability at cloud service systems In: *Proceedings Volume 10445, Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017*, 104451V (2017). <https://doi.org/10.1117/12.2280885>
8. Gnatushenko, V.V., Ali, D.: Studies of self-similar processes traffic based on the ON/OFF model. In: *Bulletin of the National University “Lviv Polytechnic”, Series of “Computer Science and Information Technologies” - Lviv*, No. 751, pp. 87–94 (2013). (in Ukrainian)
9. Ramirez-Pacheco, J.C., Torres-Roman, D., Toral-Cruz, H., Estrada-Vargas, L.: A high performance tool for the test of long-memory and self-similarity. In: *Simulation Technologies in Networking and Communications: Selecting the Best Tool for the Test*, pp. 93–114. CRC Press, Taylor & Francis Group, USA (2014). <https://doi.org/10.1201/b17650-6>
10. Estrada Vargas, L., Torres Romn, D., Toral-Cruz, H.: A study of wavelet analysis and data extraction from second-order self-similar time series. *Math. Probl. Eng.* **2013**, 1–14 (2013). <https://doi.org/10.1155/2013/102834>

11. Cajueiro, D., Tabak, B.: The rescaled variance statistic and the determination of the Hurst exponent. *Math. Comput. Simul.* **70**, 172–179 (2005). <https://doi.org/10.1016/j.matcom.2005.06.005>
12. Mandelbrot, B.B.: *The Fractal Geometry of Nature*. W. H. Freeman, New York (1982). <https://doi.org/10.1119/1.13295>. 550 p. 8
13. Senyk, P.: Reversal of incomplete Beta function. *Ukrainian Math. J.* **3**, 325–333 (1969). (in Ukrainian)