Managing Multifractal Properties of The Binary Sequence Generated With The Markov Chains

Hanna Drieieva[1[0000-0002-8557-3443]], Oleksii Smirnov[1[0000-0001-9543-874X]] and Oleksandr Drieiev[1[0000-0001-6951-2002]] Volodymyr Simakhin[2[0000-0003-4497-0925]], Serhiy Bondar[2[0000-0003-4140-7985]] and Roman Odarchenko [3,4[0000-0003-4140-7985]]

[1] Central Ukrainian National Technical University, Ukraine, Kropyvnytskyi
[2] International Research and Training Center for Information Technologies and Systems, Ukraine, Kyiv
[3] National Aviation University, Kyiv, Ukraine, 03058
[4]Yessenov University, Aktau, Kazahstan
dr.SmirnovOA@gmail.com

Abstract. With the development of modern telecommunications systems and the exponential growth in the demand for information transmission, there is a constant lack of bandwidth for the available telecommunication channels under the management of routers and communicators, because it is necessary to distribute the load of the network segments taking into account their self-similar nature of the traffic. Now, it is not possible to analytically build criteria and algorithms for optimal traffic management to ensure QoS measurements. The correspondence of the practical results with the theoretical ones was experimentally confirmed, although non-standard metric of the coverage measure expression was used in the derivation of analytical dependencies. That's why it was decided to develop theoretical method and provide experimental confirmation using conventional methods of estimating the time series' fractal dimension. The actual possibilities of adjusting the Hearst index on a given time scaling were established as a result of the numerical experiment. The obtained time series with the help of a cascade binary sequence generator have multifractal properties. That is, the cascade generator has more possibilities for receiving traffic, which will correspond to the examples of binary traffic in real telecommunication networks. However, the cascade generator requires further theoretical study to analytically express the coefficients to obtain the desired fractal characteristics and flow intensity.

Keywords: multifractal properties, binary sequence, Markov Chains, traffic modeling, Hurst exponent.

1 Introduction

With the development of modern computer telecommunications systems and the exponential growth in the demand for information transmission, there is a constant lack of bandwidth for the available telecommunication channels under the management of

routers and communicators, because it is necessary to distribute the load of the network segments taking into account their self-similar nature. Unfortunately, it is not possible to analytically build criteria and algorithms for optimal traffic management to ensure QoS measurements. Therefore, to select suitable modes of work for real telecommunication equipment the simulation systems (in particular: ModelNet, Ohio Network Emulator, ENDE, EMPOWERb OPNET, Emulab, NISTNET, DummyNet, NS, GTNeS, NSE, NETWARS) are commonly used [1,2]. As a result, such systems should have telecommunication traffic simulation sources with self-similar properties that would meet the actual characteristics determined experimentally at the simulation preparation stage.

The analysis of imitation methods for telecommunication traffic sources distinguish the following: Poisson processes; fractal Brownian motion-based generators; fractal Gaussian noise with discrete wavelet transform generation (where wavelet coefficients are independent random variables with normal distribution); Levi's fractal movement (a generalized Brownian movement, which is self-similar and forms distributions with "heavy tails"); autoregressive models (which assume that the current value of the process is the sum of a constant weighted sum of the previous values and the error of the model); neural network models (that are trained with experimental traffic to predict a new element); Markov chains (that allow you to create relatively simple models of discrete traffic generation with a wide range of properties) [3-5].

In most cases, the choice of traffic generation method should be based on the traffic properties obtained experimentally from the telecommunication system, for which traffic management algorithms will be implemented.

In [6], the authors analyzed Markov chain-based generation methods that are characterized by low computational complexity and a wide range of applications. An analytical expression of the generator parameters was solved to ensure the intensity of the simulated flow and its fractal dimension. However, the generated traffic, as shown in further studies in this article, contains multifractal properties that did not meet the real indicators, which led to the goal of improving the traffic generation simulator for telecommunication networks to replicate and manage multifractal properties.

2 Literature Analysis and Problem Statement

As a result of the previous study [6], a traffic generator G (Fig. 1) was used, which contains states "1", "0" and the probabilities to keep the state are $p("0" \rightarrow "0") = p_0$ and $p("1" \rightarrow "1") = p_1$. The generator outputs the value of the current state discretely upon an event of keeping or changing the state.

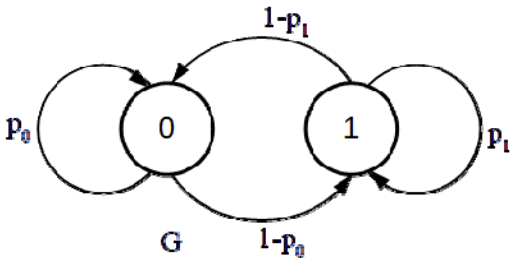

Fig. 1. Fractal Binary Sequence Generator

For such a generator flow intensity is determined as [6]:

$$\lambda = \frac{1 - p_0}{2 - (p_0 + p_1)} \tag{1}$$

Note that with $p_0 = p_1$ flow is generated with an intensity of 0.5; but as the probabilities to keep the next generated value is changed, flow's fractal dimension changes as well. A number of simulation experiments must be performed to confirm the theoretical conclusions.

The theoretical estimation of fractal dimension of the generated binary traffic was carried out [6], which depends not only on the probabilities of the generation parameters, but also on the size of the sub-band, i.e. scaling over time. To abstract from the size of the sub-band, a limit was found when the length of the sub-band tends to one. However, the obtained results are theoretical and experimental confirmation using conventional methods of estimating the time series' fractal dimension is required.

3 Evaluation of the Hurst exponent

The experimental simulation involves performing R/S analysis on the generated binary sequence. One of the R/S analysis implementations is described in [7]. The measurement is performed 200 times on different implementations of the sequence due to low accuracy of the method. The result of the experiment is Hurst exponent values, which are shown in Fig. 2 and Fig. 3, where the probabilities of keeping the state unchanged are 0.1 and 0.9, respectively:

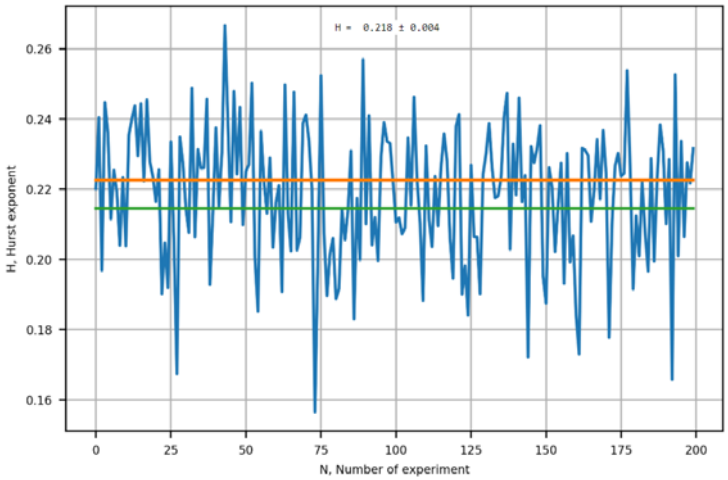

Fig. 2. Hurst exponent measurement series at 5, 10, 15, 20 samples, $p_0 = p_1 = 0.1$

Fig. 2 and Fig. 3 show that the probability of keeping the next state value equal to the previous does not change the intensity of the flow of "1", but the probability of obtaining long chains with values "0" and "1" is greatly changed. With a high probability of keeping the next value equal to the previous $(p_0 = p_1 = 0.9)$, a persistent series (Fig. 3) with Hurst exponent value $H = 0.947 \pm 0.002$ is obtained. In the case of high probability to keep the next value opposite to the previous one $(p_0 = p_1 = 0.1)$ – an antipersistent series with Hurst exponent $H = 0.218 \pm 0.004$ is obtained.

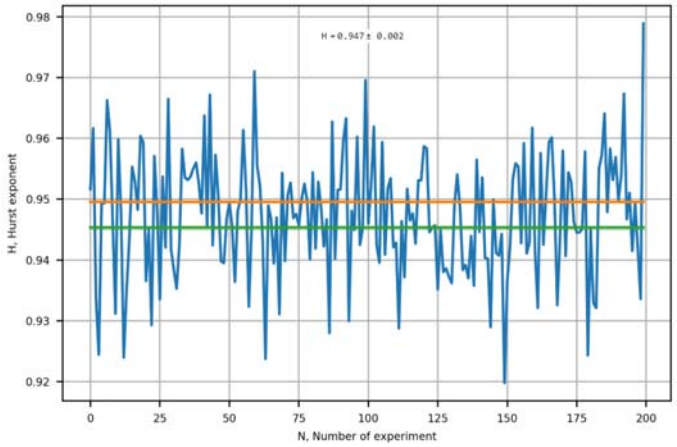

Fig. 3. Hurst exponent measurement series at 5, 10, 15, 20 samples, $p_0 = p_1 = 0.9$

In [6] the authors have obtained an estimate of the fractal dimension of a series produced by the traffic generator G (Fig. 1) on another metric, when the coverage was considered zero if the series contains only "1" or "0" elements. The fractal dimension of the series relating to the Hurst exponent as (2):

$$H(p) = \frac{p \log(p)}{1 - p}, \tag{2}$$

where $p = p_0 = p_1$. The dependence of the Hurst exponent $H(p)$ on the probability of the next value change for p is shown in Fig. 4 where theoretical dependence is upper red curve and practically measured dependence is blue curve.

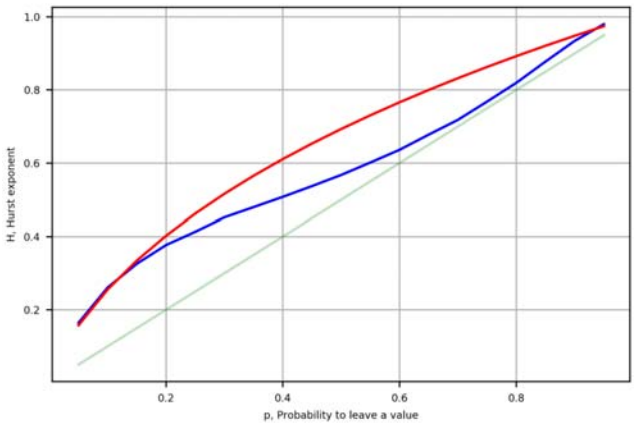

Fig. 4. The dependence of the Hurst exponent on the probability of re-generation by the traffic generator G

There are procedures of conformity assessment, formation and implementation of control actions, efficiency estimation during regular mode. In case of intelligence based mode the additional procedures are executed when efficiency decreases.

The authors assumed that the problem of efficiency decreasing can be solved based on artificial intelligence principles utilization.

Let's consider the algorithms for data processing. The list of algorithms can be presented as:

- A_1 is an algorithm of data collecting;
- A_2 is an algorithm of conformity assessment;
- A_3 is an algorithm of parameters stability assessment in case of changepoint;
- A_4 is an algorithm of preventive and corrective actions formation;
- A_5 is an algorithm of preventive and corrective actions implementation;
- A_6 is an algorithm of decision making about efficiency providing after control actions implementation;

– A_7 is an algorithm of decision making about additional data processing procedures;

– A_8 is an algorithm of decision making about usage of intelligence based procedures;

– A_9 is an algorithm of statistical processing for diagnostic variables;

– A_{10} is an algorithm of statistical processing for reliability parameters.

Algorithms A_7, A_8, A_9, A_{10} have complex structure and contain the set of procedures.

All algorithms are generalized. For detailed description of algorithms it is necessary to solve synthesis and analysis problems, to choice best option for criterion of maximum efficiency, etc. Initial information for synthesis and analysis problems is measured data trends model.

Considered algorithms contain detection, estimation, filtration, extrapolation, interpolation, and other procedures. There are algorithms with known sample size, and sequential algorithms. The sequential procedures have advantages in duration of decision making [3].

Usage of adaptability principles is based on the following approaches:

– logic based solution finding;

– fuzzy logic;

– Bayesian network;

– adaptable learning after observation;

– semantic network;

– neural network, etc [4].

More over, during diagnostic variables measuring, expert evaluation and subjective probability based estimates can be used [1].

There are different types of adaptation:

1) adaptation to the models and models parameters;

2) adaptation to the external conditions;

3) adaptation to internal changes in operation system (OS);

4) adaptation to the new requirements of regulatory and normative documents;

5) adaptation to OS aims, etc.

4 Estimation of the Hurst exponent at different scales over time

For the binary series, the fractal dimension depends on the length of the sequence for which rescaled range (R/S) analysis is performed [8, 9]. Estimation value of the Hurst exponent can be calculated by estimating the dependence of the R/S on the time span n (where n is set of time span intervals, i.e. $n = N, N/2, N/4, \ldots$). For example, under the same conditions as in Fig. 2 and 3, an increase in scaling (when the range and dispersion are calculated not at intervals of 5-20 samples, but in 450-600 samples) results the Hurst exponent values $H(0.1) \approx H(0.9) \approx 0.5$. Fig. 5 and 6 shows the results of the conducted experiment, which displays the "degeneration" of the Hurst exponent up to 0.5.

In Fig. 5 the generator G was used with the probabilities of keeping the previous state unchanged as $p_0 = p_1 = 0.1$, which should result in a visible antipersistence. However, it is obvious from that graph that aggregating cumulative sums of 450/500/550/600 samples keeps the Hurst exponent close to 0.5. Scilicet, on the specified time scale the generated sequence does not differ from a random time series in which the next element is independent of the previous one. This shows that the central limit theorem is valid, and therefore, as the quantity of samples increases for the number series' cumulative sums, these sums are getting closer to the normal distribution.

A similar impact of the cumulative sums values distribution approximation to the normal distribution is observed for the persistent series with high probability to keep the previous state $p_0 = p_1 = 0.9$, which can be seen from Fig. 6 as a decrease in the Hurst exponent values from 0.95 to 0.55.

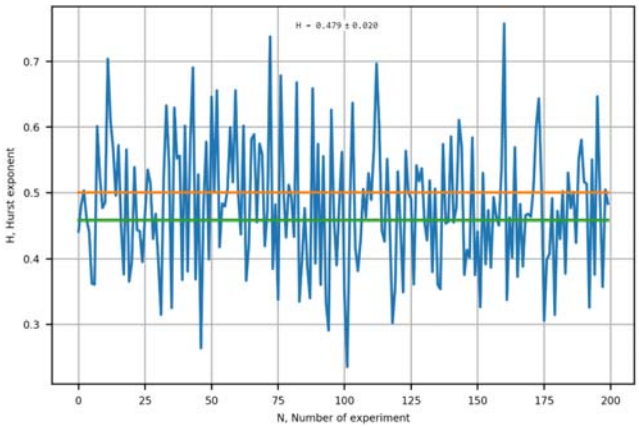

Fig. 5. Hurst exponent values in the series at 450, 500, 550, 600 samples, $p_0 = p_1 = 0.1$

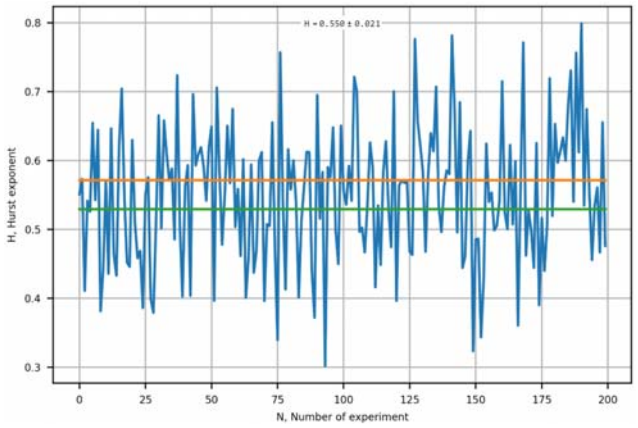

Fig. 6. Hurst exponent values in the series at 450, 500, 550, 600 samples, $p0 = p1 = 0.9$

As shown in [10] (Fig. 7), the change in fractal dimension with time scaling of real traffic data is relevant for quite a long time, as evidenced by publications on the development of an effective multifractal traffic generator [11, 12]. In this case, the fractal dimension can both increase and decrease with changing time scale on the real data.

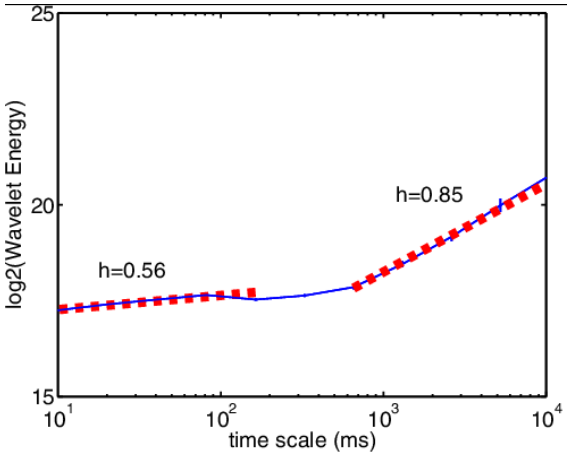

Fig. 7. Change of the Hurst exponent from time scale [10, Fig. 1.a]

In order to reveal the dependence of the Hurst exponent on the length of partial (cumulative) sums in R/S analysis, the dependence graph of the Hurst exponent on the length of these sums (Fig. 8) is shown.

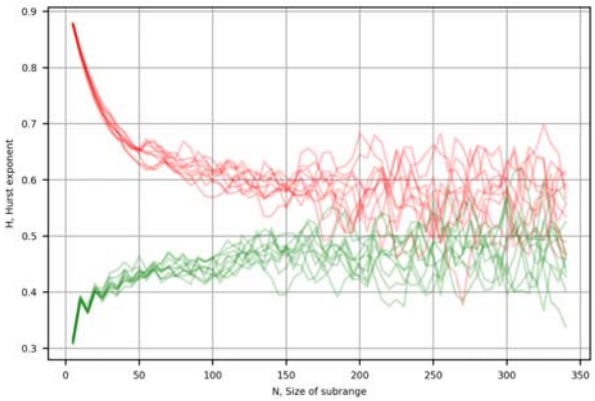

Fig. 8. Dependence of the Hurst exponent on the N interval's length in case of R/S analysis

Fig. 8 contains two bundles of curves, where each curve corresponds to one experimental determination of the Hurst exponent with the specified scaling. The upper bundle (red curves) corresponds to the experiment of measuring the Hurst exponent

with the generator G parameters set to $p_0 = p_1 = 0.9$, and the lower bundle (green curves) parameters are $p_0 = p_1 = 0.1$. It is obvious that both bundles approach $H = 0.5$.

Practical implementations of the communication networks' multifractal traffic may vary at different scales, so the purpose of this work is to add to the binary traffic generator based on stochastic automaton G a mechanism that would be able to control the fractal dimension of the binary series at different scales. This will greatly increase the relevance of the generated data to actual network traffic and extend the applicability of the generator.

Adjustment of the fractal dimension of the generated traffic based on the generator G can be performed with the proposed model (Fig. 9), which is based on a cascade binary traffic generator [13, 14]:

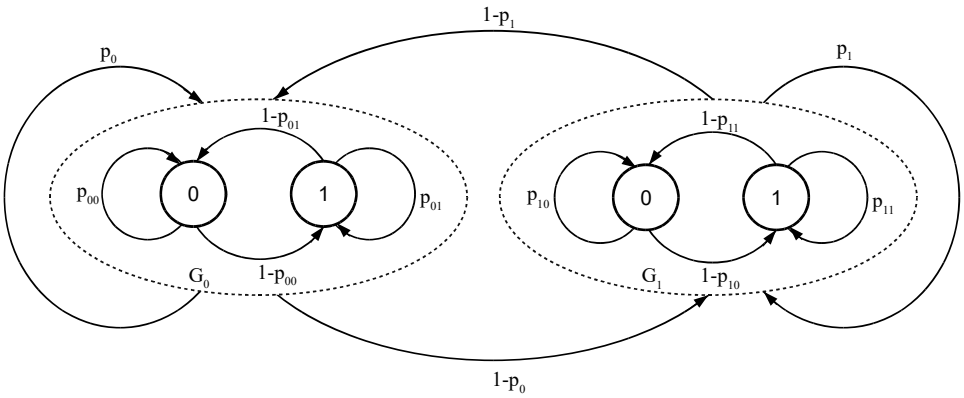

Fig. 9. Multifractal binary sequence generator

The proposed generator is cascaded and contains an external generator G, which aggregates the generators G_0 and G_1. Such generator G is characterized by the following properties: p_{00}, p_{01} – the probability of keeping the next value unchanged for the generator G_0; p_{10}, p_{11} – the probability of keeping the next value unchanged for generator G_1; p_0, p_1 – the probability of keeping the next set of d values for the same G_1 generator as in the previous series.

According to the formation of cumulative sums, the generators G_0 and G_1 must differ in a sufficiently long period, which is not possible at the same intensity of the generated flow according to the limit theorem. Therefore, the generators are configured to generate sequences with the same fractal dimension but with different flow intensities. Therefore, on a large scale, the generator G is responsible for balancing the total flow intensity, which in this example is symmetrical and thus averages the flows from the generators G_0 and G_1.

In order to visually evaluate the formed sequence, a binary series that is aggregated by 25 counts of the generator G cycles is shown in Fig. 10. The result is a char-

acteristic of high-intensity pulsating flow $\tau = 0.5$. This series is obtained with the following generation parameters (2):

$$d = 10; p_0 = 0.9; p_1 = 0.9; p_{00} = 0.95; p_{01} = 0.15; \\ p_{10} = 0.15; p_{11} = 0.95 \quad (2)$$

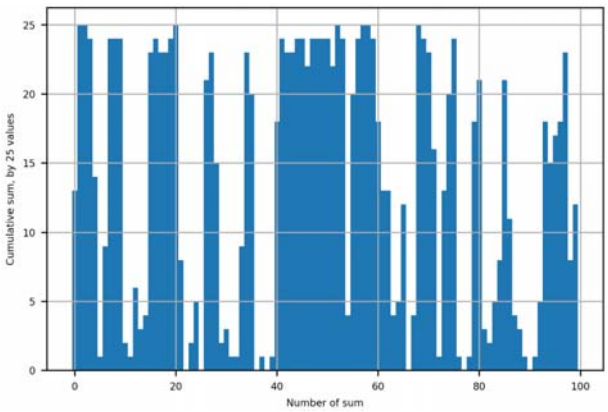

Fig. 10. An example of a cascaded traffic generator that has persistence on a wider time scale

The Fig. 11 is similarly constructed, but with the following parameters (3):

$$d = 10; p_0 = 0.05; p_1 = 0.05; p_{00} = 0.95; p_{01} = 0.15; \\ p_{10} = 0.15; p_{11} = 0.95 \quad (3)$$

It should be emphasized that the changes occurred only for the traffic generator G.

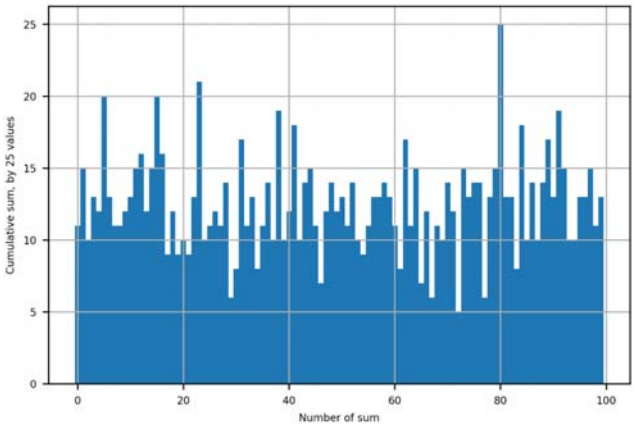

Fig. 11. An example of a cascaded traffic generator that is persistent and antipersistent on a different time scale

For the cascade generator G, a graph for the dependence of the Hurst exponent on time scale is plotted (Fig. 12). The graph contains three bundles of curves that correspond to the parameter of generator G described as (2) as (3), which have already been shown in Fig. 8, and the parameter set described below (4):

$$d = 10; p_0 = 0.0; p_1 = 1.0; p_{00} = 0.9; p_{01} = 0.9; p_{10} = 0.9; p_{11} = 0.9 \quad (4)$$

Set of parameters (4) means that the modulator switches to the operation only of the generator G_1, with parameters $p_{10} = 0.90; p_{11} = 0.90$, and it corresponds to the operation of the generator G, without correction of the fractal dimension in large time scales. In Fig. 12 parameters set (4) is shown by the green bundle of curves, which is equivalent to the upper bundle of curves in the graph. 8. Red and blue bundles are equivalent to sets (2) and (3) respectively.

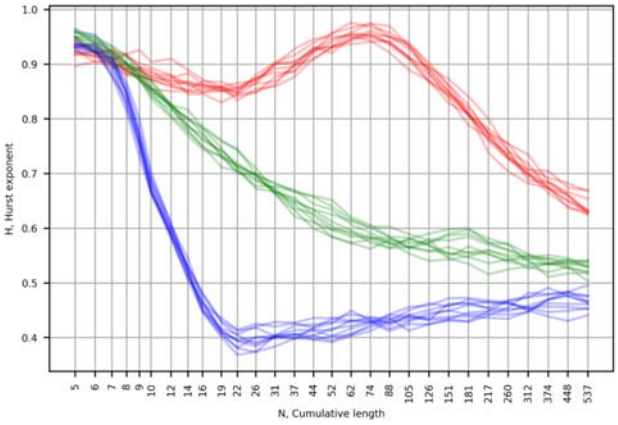

Fig. 12. Comparison of the dependence of the Hurst exponent on time scale for the cascade and standard generators G

The most important property for using a cascade generator is its ease of setup, which actually means being able to get the coefficients $d, p_0, p_1, p_{00}, p_{01}, p_{10}, p_{11}$ from the real traffic sample. In [6] the authors have already obtained the following ratio for the generator G:

$$G(0) = \frac{1 - p_1}{2 - (p_0 + p_1)}; G(1) = \frac{1 - p_0}{2 - (p_0 + p_1)} \quad (5)$$

where $G(1)$ is the probability for the modulator to use the first generator and $G(1)$ – to use a zero generator.

The corresponding flow intensities for the dependent generators can be expressed from (1) as (6):

$$\lambda_0 = \frac{1 - p_{00}}{2 - (p_{00} + p_{01})}; \lambda_1 = \frac{1 - p_{10}}{2 - (p_{10} + p_{11})} \tag{6}$$

The total flow intensity for the cascade generator can be expressed as (7) due to the known "portions" of the work of the both generators G_0 and G_1:

$$\lambda = \frac{(1 - p_1)}{2 - (p_0 + p_1)} \cdot \frac{(1 - p_{00})}{2 - (p_{00} + p_{01})} + \frac{(1 - p_0)}{2 - (p_0 + p_1)} \cdot \frac{(1 - p_{10})}{2 - (p_{10} \_ p_{11})} \tag{7}$$

According to (7), the expression of flow intensity in traffic depends on six parameters, which indicates the possibility to change the properties of the generated traffic within wide limits. However, in practice, this dependency is not applicable, because the fractal dimension of traffic on two temporal scales needs to be linked here, but this problem has not yet been solved.

5 Conclusion

The correspondence of the practical results with the theoretical ones was experimentally confirmed, although non-standard metric of the coverage measure expression was used in the derivation of analytical dependencies.

The actual possibilities of adjusting the Hearst index on a given time scaling were established as a result of the numerical experiment. The obtained time series with the help of a cascade binary sequence generator have multifractal properties. That is, the cascade generator has more possibilities for receiving traffic, which will correspond to the examples of binary traffic in real telecommunication networks.

However, the cascade generator requires further theoretical study to analytically express the coefficients $d, p_0, p_1, p_{00}, p_{01}, p_{10}, p_{11}$ to obtain the desired fractal characteristics and flow intensity.

The generalized diagram of processing procedure is considered in the paper. This diagram suggests two modes of OS: regular mode and adaptable mode. The adaptability principles utilization expands the possibilities of flexible control of operations. Three strategies of OS components inspection were analyzed. The numerical example showed advantages of catchall inspection.

References

1. Srivastava, S., Anmulwar, S., Sapkal, A.M., Batra, T., Gupta, A.K. and Kumar, V.: Comparative study of various traffic generator tools. In: Recent Advances in Engineering and Computational Sciences (RAECS), University Institute of Engineering and Technology, Panjab University, Chandigarh, pp. 1-6 (2014)
2. Avallone, S., Pescape, A. and Ventre, G.: Analysis and experimentation of internet traffic generator. In: Next Generation Teletraffic and Wired/Wireless Advanced Networking, February 02-06, 2004, St. Petersburg, Russia, pp 70-75 (2004)
3. Robert, S. and Le Boudec, J.Y.: New models for pseudo self-similar traffic. In: Performance Evaluation, vol. 30 (1-2), pp 57-68 (1997)
4. Horn, G., Kvalbein, A., Blomskøld, J. and Nilsen, E.: An empirical comparison of generators for self similar simulated traffic. In: Performance Evaluation, vol. 64 (2), pp 162-190 (2007)
5. Sobh, T., Elleithy, K. and Mahmood, A. eds.: Novel algorithms and techniques in telecommunications and networking. In: Springer, pp 41-46 (2010)
6. Drieieva H.M., Smirnov O.A., Drieiev O.M., Smirnova T.V.; A fractal analysis of a Markov chain based self-similar traffic generator. In: Central Ukrainian Scientific Bulletin. Engineering sciences. vol. 2 (33), pp 161-172 (2019)
7. GitHub - Mottl/hurst: Hurst exponent evaluation and R/S-analysis in Python. https://github.com/Mottl/hurst
8. Nazarychev, S.A., Zagretdinov, A.R., Ziganshin, S.G. and Vankov, Y.V.: Classification of time series using the Hurst exponent. In: Journal of Physics: Conference Series, vol. 1328, pp 66-73. (2019)
9. Fontugne, R., Abry, P., Fukuda, K., Veitch, D., Cho, K., Borgnat, P. and Wendt, H.: Scaling in internet traffic: a 14 year and 3 day longitudinal study, with multiscale analyses and random projections. In: IEEE/ACM Transactions on Networking, vol. 25 (4), pp 2152-2165 (2017)
10. Vinay J. Ribeiro, Zhi-Li Zhang, Sue Moon, and Christophe Diot: Small-time scaling behavior of Internet backbone traffic.In: Computer Networks, vol. 48, pp.315–334 (2005)
11. G. Millána, G. Lefranc.: A Fast Multifractal Model for Self-Similar Traffic Flows in High-Speed Computer Networks. In: Information Technology and Quantitative Management. ITQM 2013 Procedia Computer Science, vol. 3, pp 420 – 425 (2013)
12. Areström, E. and Carlsson, N.: Early Online Classification of Encrypted Traffic Streams using Multi-fractal Features. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops, pp 84-89 (2019)
13. Kant, K. On aggregate traffic generation with multifractal properties. In: Seamless Interconnection for Universal Services, Global Telecommunications Conference. GLOBECOM'99. vol. 2, pp 1179-1183 (1999)
14. Anderson, D., Cleveland, W.S. and Xi, B.: Multifractal and Gaussian fractional sum–difference models for Internet traffic. In: Performance Evaluation, vol. 107, pp 1-33 (2017)