MATHEMATICAL SIMULATION OF SELF-SIMILAR NETWORK TRAFFIC WITH AIMED PARAMETERS

L. Kirichenko

Kharkiv National University of Radioelectronics, Department of Applied Mathematics

T. Radivilova

Kharkiv National University of Radioelectronics, Department of Telecommunication Systems

Abed Saif Alghawli

Salman Bin Abdulaziz University, Department of Computer Science

ABSTRACT: The objective of the given work is to study the queues resulting in the buffer while self-similar traffic passing through a network nod and to build a mathematical model of an actual traffic.
The latest researches of different types of network traffics bring out clearly that network traffic is defined by self-similarity and long-term dependence. The self-similar traffic has specific structure being reserved at various measures – its realization is characterized by some vast emissions when respecting low medium-scale traffic. This fact degrades the performance significantly (increases, losses and delays) while running through the network nods. Hence, it follows that commonly used methods of simulation and network system calculations rested on traditional assumptions do not reflect the real situation taking place in the network. [L+94, CTB98, PW00, JOB12, L+10].
KEYWORDS: self-similar, network traffic, stochastic processes, queue.

1.BASIC THEORETICAL INFORMATION ABOUT SELF-SIMILAR STOCHASTIC PROCESSES [L+84, CTB98, L+10, SSA07]

Stochastic process with continuous time $X(t)$ is called self-similar in restricted sense with a parameter $H,\ 0.5<H<1$, if for any real value a>0 the expression $a^{-H}X(at) \sim X(t)$ is true. That is the $a^{-H}X(at)$ and $X(t)$ have the same finite-dimensional distributions.

Stochastic process $X(t)$ is called self-similar in wide sense if the process $a^{-H}X(at)$ possesses the same statistical characteristics of the second order, as $X(t)$:

mathematical expectation $M[X(t)] = \dfrac{M[X(at)]}{a^H}$;

autocorrelation function: $r_x(t,s) = \dfrac{r_x(at,as)}{a^{2H}}$.

H parameter, named Hurst exponent, presents the measure of self-similarity or the measure of persistence of long-term dependence of the stochastic process. The value H=0,5 denotes the absence of long-term dependence. The closer is the value H to 1, the higher is the degree of positive stability of long-term dependence.

Let us consider the term of self-similarity for the processes with discrete time. Assume that $X=(X_1,X_2,...)$ - time realization of wide-sense stationary stochastic process of discrete time $t \in N=\{1,2,...\}$.

Let us denote through $X^{(m)}=\{X_1^{(m)},X_2^{(m)},...\}$ - the averaged along length blocks m process X, whose components are defined by the equation

$$X_t^{(m)} = \tfrac{1}{m}(X_{tm-m+1}+...+X_{tm}),\quad m,t\in N .$$

The series $X^{(m)}$ is called aggregative.

The process X is called self-similar with a parameter in restricted sense $H,\ 0.5<H<1$, if the expression $m^{1-H}X^{(m)} \sim X,\ \ m\in N$ is true, which is regarded as the equality of distributions.

The process X is called self-similar with a parameter H, in wide sense if the correlation function of an aggregative function is

$$r_m(k)=r(k),\ \ k\in Z_+,\ \ m\in\{2,3,...\},$$

that is the process leaves its correlation index unchanged after averaging over length blocks m.

One of the essential properties of the network traffic as a stochastic process is the presence of heavy tails of its one-dimensional distribution functions.

A random variable ξ имеет распределение с has heavy-tailed distribution, if $P[\xi > x] \sim c \cdot x^{-\alpha}, x \to \infty$, that is a distribution tail fades out along the power law. The parameter α denotes the level of the tail heaviness. When $0 < \alpha < 2$ the value ξ possesses nonterminating dispersion, and when $0 < \alpha < 1$ its mathematical expectation is nonterminating as well.

One of the characteristics of a heavy-tailed random variable is its high unsteadiness. That is to say, the sample collection of such random variable presents relatively low values in the main, yet it also contains a fair number of very big values.

2.INVESTIGATION OF THE PROPERTIES OF NETWORK TRAFFIC

The data coming into a global network from local Internet provider's network were chosen as an actual network traffic. Osi.dat, Osi.dst and Osi.04.src implementations represent the time dependence of the size of the network protocol frames of the second level of OSI model (Ethernet-frame, in this case) coming through a local area network to the global Internet.

Tcp.dat implementation represents time dependence of the size of the network protocol of the fourth frame level of OSI model (TCP in this case) through a local area network to the global network Internet. The schedule of implementation of the Tcp.dat fragment, aggregated at the level of 10 seconds is shown in Fig. 1: The abscissa represents time samples, and the vertical axis - lengths of Ethernet-frames.

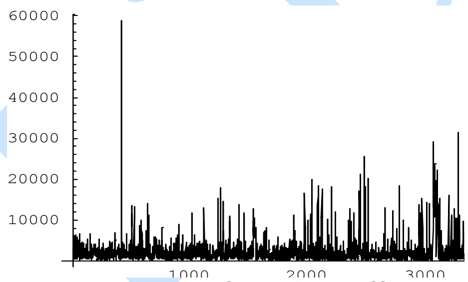

Fig. 1. Fragment of Tcp.dat traffic

The investigations conducted showed that the following implementations of data network traffic are self-similar with Hurst exponent $0.6 < H < 0.9$.

The analyzed implementations exhibit high variability: their averages are relatively small values, while at the same time the implementation provides a sufficient number of high emissions. In Figure 3 (a) a histogram for Tcp.dat traffic is shown, suggesting the presence of a heavy tail. The studies have testified that the distribution of the considered implementations of network traffic possess heavy tails with a parameter of tail heaviness in the range $1.2 < \alpha < 1.8$.

3.THE STUDY OF QUEUES IN THE BUFFER

Networks built on TCP / IP often give rise to strong emission peaks. These time localized releases cause significant packet losses, even when the total demand of all flows is far from the maximum allowable values. Numerous studies show that if the input traffic is self-similar, then for any multiplexing of self-similar flows the duration of delays will be high and require buffers of larger sizes. Moreover, for large values of H the demand in the buffer is beginning to grow rapidly even with a modes load factor. If it is necessary to achieve a high degree of use for self-similar traffic, the buffers required are much larger than it is predicted by classical analysis of queues. [SSA07, YWL07, NOS07, C+10].

The paper presents the simulation of a channel loading and the emergence of queuing in the buffer for the implementations of network traffic. Taking into account the digital nature of modern high-speed communication networks, the communication system can be viewed as a queuing system (QS) of the form, $G/D/C/\infty/d$, where G means that the input traffic has an arbitrary distribution; D – a deterministic service time equal to one; C – the number of servers equal to the channel bandwidth; ∞ means an infinite buffer size and d - discipline operating the system. [TG97, Sto02].

The input buffer receives traffic $Y = (Y_1, Y_2, Y_3, \dots)$, where Y_t denotes the number of packets that arrive at a time moment t. It is assumed that in QS in every moment t the discipline decides which of the following alternatives should be applied to the package in the system: 1) to initiate the transmission (service) of the package at the time moment t; 2) to store the package in a buffer until the moment $t+1$; 3) to reset (to lose) the package at the time moment t.

In each window t (a window is a time interval $[t, t+1]$), the channel can transmit no more than C packages, which are taken either from the buffer, or from Y_t new packages. The package of the buffer which is passed to the window t, leaves the channel and the system itself at a time moment $t+1$.

In theory of teletraffic, one of the most commonly used characteristics of data flow is Fano parameter F, which is defined as the ratio of the variance of the number of events at a given time interval T to the mathematical expectation of this quantity: $F(T) = \frac{D[N(T)]}{M[N(T)]}$, where $N(T)$ – a random variable

that determines the number of events of the studied flow on that interval T.
It is shown that the length of the queue in the buffer is determined by three main parameters of the traffic: traffic intensity, Fano parameter and Hurst exponent. Large values of the Fano parameter correspond to the large scatter of values in the input stream, which, even at low intensity, creates queues. Hurst exponent $H > 0.5$ implies that the high values of the process to be followed by the same as high ones, which does not allow the buffer to be cleaned up fast. Figure 3 shows the queue formed in the buffer at every time moment when passing through the above mentioned queuing traffic area shown in Figure 1. The load capacity of the system in this case was 90%.

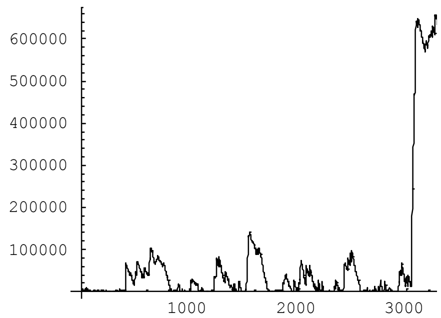

Fig. 2. The buffer queue for Tcp.dat traffic

Figure 3 (b) shows the histogram corresponding to the values of the queue, which allows to make an assumption about the presence of a heavy tail.

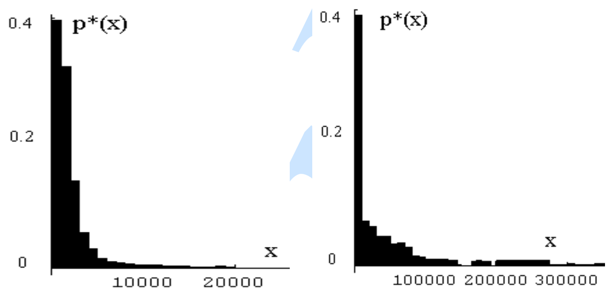

Fig. 3 a) histogram of Tcp.dat traffic; b) histogram of the corresponding queue in the buffer

Numerical analysis of queues *B(t)*, resulting in a buffer of infinite size, has showed that *B(t)* is self-similar stochastic process. Hurst index for such a process coincides with the Hurst index of input traffic. . However, the process *B(t)* is far more heavy-tailed: the heaviness parameter of a tail resulting in the queue buffer lies in the range $0.2 < \alpha < 0.7$
Table 1 shows the values of average intensity of the traffic $\overline{X}$ implementations by the length $N = 4000$ of counts, evaluation of Fano F and Hurst H exponents, as well as the values of the average queue length $\overline{B}$, formed in the buffer when the system is loaded by 90% and .parameters for the tail heaviness resulting in a queue buffer.

Table 1. The parameters of the actual traffic

Implementation	$\overline{X}$	F	H	$\overline{B}$	α_B
Osi.dat (1)	4500	5900	0.8	540159	0.45
Tcp.dat(2)	4200	7200	0.81	470564	0.35
Osi.src (3)	3700	7100	0.82	410256	0.271
Osi.dst (4)	2700	5200	0.78	251368	0.31

During the simulation the average system load varied from 70 to 95 percent. It was studied that there is a dependence of the average queue size in the buffer on the channel loading size. The studies conducted showed that when the system loads at 80-90%, the average queue in the buffer exceeds the intensity of the traffic to hundreds of times.
Figure 5 shows the average size of the buffer memory when the system loads from 70% to 95% for the investigated implementations of traffics.

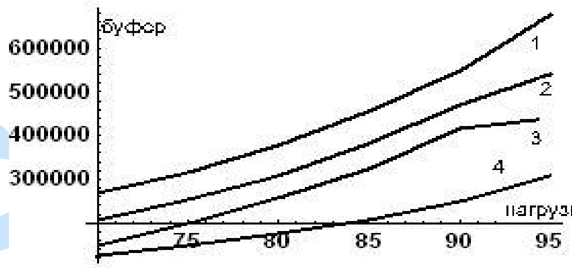

Fig. 4. Dependence of the average queue length on the load for the implementations of actual traffic

The task of self-similar traffic simulating is not the generation of a sample implementation being completely identical to the actual traffic (which is an impossible task), but only getting the samples with the same statistical characteristics (magnitude and frequency of bursts, the intensity of traffic), as in real traffic. The main criterion for the traffic model adequacy can be considered as identical behavior in the communication channel, i.e. the analogy of the dependency of the queue length on the system load.

4.MATHEMATICAL SIMULATION OF SELF-SIMILAR TRAFFIC

Fractional Brownian motion (FBM) is often considered as a stochastic process possessing fractal properties and it is widely used in physics, chemistry, biology, economics and the theory of network traffic. [Sto02, Fed88, Cro95]
Gaussian process X (t) is called a fractal Brownian motion with a parameter $H, 0 < H < 1$, if the increments of the stochastic process $\Delta X(\tau) = X(t+\tau) - X(t)$ possess Gaussian distribution $\Delta X \sim N(0, s_0 \times t^H)$, i.e.

$$P(\Delta X < x) = \frac{1}{\sqrt{2\pi}\sigma_0 t^H} \times \int_{-\infty}^{x} \mathrm{Exp}\left[-\frac{z^2}{2\sigma_0^2 t^{2H}}\right] dz,$$

where σ_0 - the diffusion coefficient.

Fractional Brownian motion with a parameter $H = 0.5$ coincides with ordinary Brownian motion. The increment process of fBM is known as fractional Gaussian noise (fGN). The fGN dispersion is subject to relation $D[X(t+t) - X(t)] = \sigma_0^2 t^{2H}$.

The increments of fractional Brownian motion ΔX possess the ability to self-similarity in the narrow sense, that is, $X(t+\tau) - X(t) \cong \frac{X(t+a\tau) - X(t)}{a^H}$, for any $a > 0$.

There are several methods for fBM constructing for discrete time. The most practical one is the method of successive random addition of Voss.

The method includes the following algorithm [Fed88, Vos88]. The initial sequence of values of the coordinates at time ti = 0, 1/2, 1; the initial values of the coordinates are zero. Then to the values of the coordinates $X(t_1)$, $X(t_2)$, $X(t_3)$ random numbers are added selected from a normal distribution with mean zero and initial variance σ_1^2. The mean values for each time interval are then treated as additional nodes on the time axis and the values of the coordinates are measured in the interpolation. Again, random numbers with zero mean and reduced variance are added to coordinate values $\sigma_2^2 = \frac{\sigma_1^2}{2^{2H}}$.

After $n-$ repeated application of this algorithm we obtain the values of the coordinate of the generalized Brownian particle at $1 + 2^n$ time moments. The variance of the summands n-th of the generation is equal $\sigma_n^2 = \frac{\sigma_{n-1}^2}{2^{2H}} = \frac{\sigma_0^2}{2^{2Hn}}$. The process proposed by Voss, leads to the generalized Brownian motion at any solution.

Theoretically fGN can be considered as a model of self-similar traffic with a specified Hurst index and the corresponding long-term dependence. However, this model has serious shortcomings: zero means and absence of heavy tails, i.e. high-power spikes typical for network traffic.

In this paper an approach is suggested that is based on the functional transformation of fGN. The proposed transformation preserves the long-term dependence of the stochastic process and turns it into self-similar process with heavy tails.

In practice, the most commonly used are the Pareto, Cauchy, Levy heavy-tailed distributions and a log-normal distribution.

The value η possesses a log-normal distribution, if $\eta = \mathrm{Exp}[\xi]$, where the random variable ξ is normally distributed $N(a;\sigma)$. The density of the random variable η is

$$p_\eta(y) = \frac{1}{y\sqrt{2a\sigma}} e^{-\frac{(\ln y - a)^2}{2\sigma^2}}, \quad y > 0; \quad p_\eta(y) = 0, \quad y \le 0.$$

For a random variable with a log-normal distribution the expectation and variance are respectively

$$M[\eta] = e^a \cdot e^{\frac{\sigma^2}{2}}, \quad D[\eta] = e^{a^2} \cdot e^{\sigma^2} \cdot (e^{\sigma^2} - 1) \qquad (1)$$

Fig. 5 shows the density diagrams of log-normal distribution with different parameters a and σ. The tail heaviness $P[\xi > x]$ depends on the ratio $\frac{D[\eta]}{M[\eta]}$.

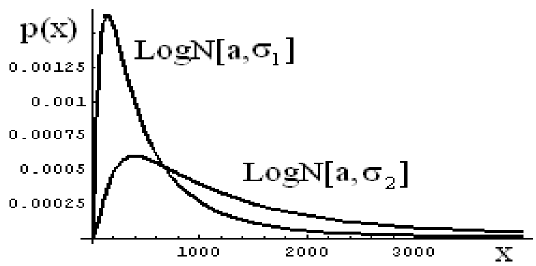

Fig. 5. The diagram of densities of log-normal distributions, $\sigma_2 > \sigma_1$

In this paper it is proposed to use the following functional transformation of fGN as a self-similar stochastic process that generates the simulative traffic implementations

$$Y(\tau) = b \cdot \mathrm{Exp}[k \cdot X(\tau)] \qquad (2)$$

where $X(\tau)$ - the series of fBM increments with the specified parameter H at a time interval *τ, b, k* - the parameters governing the frequency and magnitude of the bursts.

Increments of fBN $X(\tau)$ have normal distribution $X \cong N\left(0, \sigma^2(\tau)\right)$, with variance $\sigma^2(\tau) = \sigma_0^2 \tau^{2H}$, where σ_0 - the diffusion coefficient. In this case $Y(\tau)$ has a log-normal distribution $Y \cong \mathrm{LogN}(0, \sigma^2)$.

Assume that $X(\tau)$ - the series of fBM the self-similar parameter H at the time interval τ The stochastic process $Y(t) = \mathrm{Exp}[X(t)]$ is a self-similar stochastic process with the same Hurst index H, as the initial fGN.

Numerical characteristics of the stochastic process $Y(\tau) = b \cdot \mathrm{Exp}[k \cdot X(\tau)]$, because of (1), are as follows:

$$M[Y(\tau)] = b\cdot \mathrm{Exp}[\frac{1}{2}k^2\sigma^2(\tau)]$$
$$D[Y(\tau)] = b^2\cdot \mathrm{Exp}[k^2\sigma^2(\tau)]\left(\mathrm{Exp}[k^2\sigma^2(\tau)]-1\right) \quad (3)$$

Assume that $\tilde{Y}$ and $\tilde{F}$ both the intensity and the Fano parameter for the implementation of the simulative actual traffic τ in length are calculated by the formulas

$$\tilde{Y} = \frac{1}{\tau}\sum_{t=1}^{\tau} Y_t, \quad \tilde{F} = \frac{\sum_{t=1}^{\tau}\left(Y_t-\tilde{Y}\right)^2}{\sum_{t=1}^{\tau} Y_t} \quad (4)$$

When equating the estimates to their theoretical expressions for $M[Y(\tau)]$ and $F[Y(\tau)]$ in accordance with formulas (3), will obtain the system

$$\begin{cases} b\cdot \mathrm{Exp}[\frac{1}{2}k^2\sigma^2(\tau)] = \tilde{Y} \\ b^2\cdot \mathrm{Exp}[k^2\sigma^2(\tau)]\left(\mathrm{Exp}[k^2\sigma^2(\tau)]-1\right) = \tilde{F} \end{cases},$$

from which we find the values of parameters *b* and *k*

$$b = \frac{\tilde{Y}}{\sqrt{\frac{\tilde{F}+\tilde{Y}}{\tilde{Y}}}}, \quad k = \frac{\sqrt{\ln\left(\frac{\tilde{F}+\tilde{Y}}{\tilde{Y}}\right)}}{\sigma(\tau)} \quad (5)$$

Thus, the constructing algorithm of the simulative traffic implementation involves the following steps:

- when having the implementation of an actual network traffic, evaluate its performance $\tilde{Y}$, $\tilde{F}$, according to formulas (4), and $\tilde{H}$ one way of finding the Hurst exponent [Cle05];
- build fBN with a parameter $H = \tilde{H}$ on the interval τ by the Voss method;
- obtain the series of increments $X(\tau)$ with variance $\sigma^2(\tau)$;
- find the parameters b and k by the formulas (5);
- obtain a model of the simulative traffic implementation from the series of increments $X(\tau)$ by the formula (2).

5.PRACTICAL SIMULATING AND VALIDATION OF THE MODEL

The paper provides the simulation of the network traffic for implementations for the parameters presented in Table 1. Simulative implementations in length $\tau = 4096$ were obtained according to the algorithm above-described that takes into account the parameters of intensity, Fano and Hurst parameters. Figure 6 shows one of the network traffic implementations and its model counterpart.

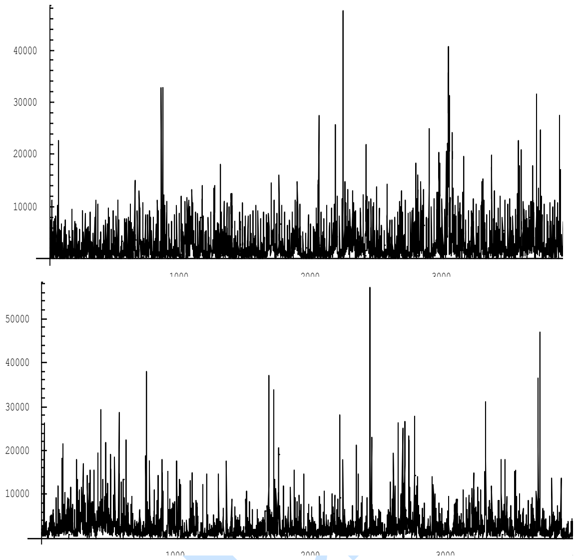

Fig. 6. Actual traffic (at the top) and simulative implementation (below)

For testing the adequacy of the model, the queues have been studied $B(t)$ that were formed in the buffer of infinite size while passing the simulative implementations through the QS. The numerical analysis of the queues $B(t)$ has showed that the queue is a statistically self-similar realization with the same Hurst index as the input realization possesses. The parameter of the tail heaviness α_B for the resulting buffer queues is close (5-10% tolerance) to the queues parameter for actual input traffics. The hypothesis of the equality of mean values of the queues in the buffer for the actual and simulative traffics has also been tested.

Table 2 shows the results of numerical simulation. Each of the simulative implementations had the same average value, the Fano parameter and Hurst exponent as the traffic implementations did. Obviously, the simulative implementations generate queues of similar size to those arising while the real traffic passing through the channel. The hypothesis of equality of mean values of the queues in the buffer for the actual and simulative traffics was accepted with a significance level $\alpha = 0.05$ in each case.

Table 2. Parameters of the simulative traffic

implementation	$\overline{X}$	F	H	$\overline{B}$	α_B
Osi.dat (1)	4500	5900	0.8	570000	0.45
Tcp.dat(2)	4200	7200	0.81	440000	0.35
Osi.src (3)	3700	7100	0.82	350000	0.271
Osi.dst (4)	2700	5200	0.78	210000	0.31

Figure 7 shows the dependency of the average size of the buffer memory when the system loads from 70% to 95% for the simulative traffic implementations under study. Thus assuming, for practical researches the behavior of the real and simulative traffics in a communicating channel can be considered as to be identical.

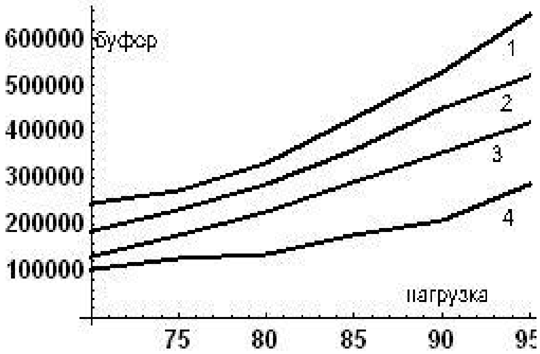

Fig. 7. Dependence of the average queue length on the load for the simulative traffic implementations

6.CONCLUSIONS

The paper presents mathematical model of the traffic, which takes into account the parameters determining the occurrence of queues when the traffic passing through the communication system: the average intensity, the Fano parameter and the Hurst exponent. The conducted simulation showed that the proposed model of the network traffic permits to adjust characteristics of the local network at the design stage or during its operation.

REFERENCES

[CTB98] **M. Crovella, M. Taqqu, A. Bestavros** - *Heavy-Tailed Probability Distribution in World Wide Web*, in A practical Guide to Heavy Tails: Statistical techniques and Application, 1998.

[Cle05] **R.G. Clegg.** *A practical guide to measuring the Hurst parameter*, Computing science technical report (№ CS–TR–916), 2005.

[Cro95] **R. M. Crownover**. *Introduction to fractals and chaos*, Jones & Bartlett Publishers, 1995. - 306 p.

[C+10] **L. Chang, Q. Xiaowei, Z. Sihai, Z. Wuyang** - *Analysis of Queuing Behaviors with Self-similar Traffic in Wireless Channels*, Journal of networks, vol. 5, no. 1, january 2010 pp.120-127

[Fed88] **J. Feder** – *Fractals*, Plenum, New York, 1988.

[JBC12] **J. Liebeherr, A. Burchard, F. Ciucu** - *Delay Bounds in Communication Networks with Heavy-Tailed and Self-Similar Traffic Information Theory*, IEEE Transactions, Vol.58 , Issue: 2, pp. 1010 – 1024, 2012

[L+10] **P. Loiseau, P. Goncalves, G. Dewaele, P. Borgnat, P. Abry, P. Vicat-Blanc Primet** – *Investigating self-similarity and heavy-tailed distributions on a large scale experimental facility Networking*, in IEEE/ACM Transactions, on Vol 18 , Issue: 4 , pp: 1261 – 1274, 2010

[L+94] **W. Leland, M. Taqqu, W. Willinger, D. Wilson** - *On the self-similarnature of ethernet traffic*, in IEEE/ACM Transactions of Networking, 2(1), 1994. P. 1-15

[NOS07] **T. Nakashima, A. Ono, T. Sueyoshi** - *Queueing analysis for self-similar traffic by simulator*, Proceedings of the Third International Conference on International Information Hiding and Multimedia Signal Processing (IIH-MSP 2007) – Vol. 02, pp. 51-54.

[PW00] **K. Park, W. Willinger** - *Self-similar network traffic and performance evaluation*, 2000. – № 14(1). – 398 p.

[Sto02] **W. Stollings** - *High-speed networks and Internets. Performance and quality of service*, New Jersey, 2002.

[SSA07] **O. Sheluhin, S. Smolskiy, O. Andrey** - *Self-similar processes in telecommuni-cations*, JohnWiley & Sons Ltd, Chichester, 2007.

[TG97] **B. Tsybakov, N. Georganas** - *Overflow probability in an ATM queue with self-similar input traffic*, in IEEE ICC'97, Montreal. - Conference Record. -Vol. 2.- 1997.-P. 822-826.

[Vos88] **R. F Voss.** *Fractals in nature: From characterization to simulation in The Science of Fractal Images*, Springer-Verlag, New York, 1988.

[YWL07] **C. Yu, Z. Weihua, W. Lei** - *Calculation of Loss Probability in a Finite Size Partitioned Buffer for Quantitative Assured Service*, in IEEE Transactions on Communications, vol. 55, pp. 1757-1771, 2007