UDC 681.3:519.2

Radivilova T.
Department of organization of studing process,
Kharkov national university of radioelectronics

R/S - ANALYSIS OF NETWORKING TRAFFIC FOR REVEALING LONG-TERM AND SHORT-TERM DEPENDENCE

Recent analyses of Internet traffic have shown that traffic volume fluctuations in wide-area and local-area networks are characterized by self-similarity, or long-range dependency. Self-similarity is a scale invariant property under timescale translation, i.e. it yields the existence of clustering and bursty characteristics in the flow over wide time scales. Thus, self-similar traffic causes larger queuing delays than the estimation by Poissonian traffic. Self-similarity is relatively a new concept in the computer networking community. It is very different from past views that considered computer network traffic as a Poisson process. In the case of self-similarity, burstness is present in all time-scales and does not tend to pure white Gaussian noise (see fig.1-2).

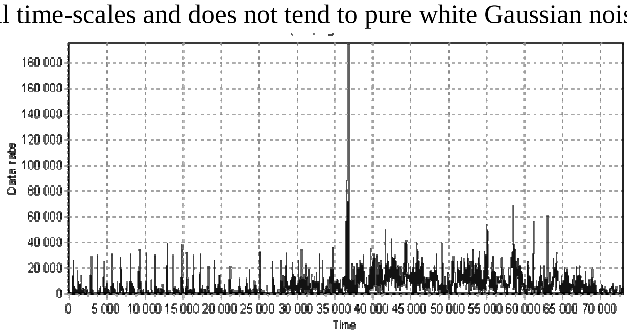

Fig.1 Aggregation by level of 30 sec.

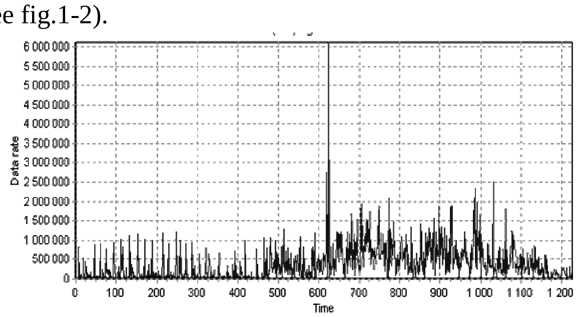

Fig.2 Aggregation by level of 3 min.

A real stochastic time series is said to be self-similar if the following holds: $X(t) =_d a^{-H} X(at)$, where $t \geq 0,\ a > 0,\ =_d$ means equality of the finite dimensional distributions and $0 < H < 1$ represents the Hurst exponent, a parameter that quantifies the degree of self-similarity. Self-similarity implies that if a process is scaled in time and normalized, it must maintain the same distributional properties as the original process.

Long-range dependence or long memory means that the autocorrelation function of a stochastic time series decays slowly or hyperbolically in time. This slowly decaying behavior gives rise to the presence of large values with non-negligible probability. In long-range dependence, the sum of the autocorrelation function is infinity. Let $X(t)$ be a stationary process, $X(t)$ is said to be long-range dependent if its autocorrelation function (fig. 3), ACF, $p(k)$ has the following asymptotic form: $p(k) \sim C_p k^{-\beta}$, where C_p is a constant and $\beta \in (0;1)$ is a real number, β is related to the Hurst exponent by the relation $H = 1 - \beta/2$.

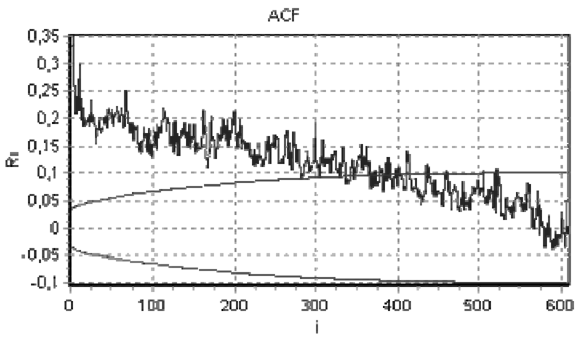

Fig.3 Autocorrelation function

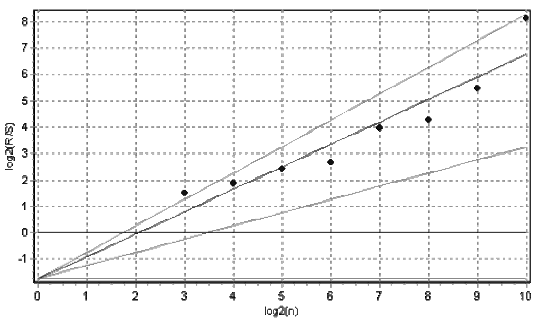

Fig.4 R/S -statistic

The R/S -statistic is one of the basic methods for inferring self-similarity in time series. First, a partition the original series into K non-ovelaping blocks is done and then for each lag n we compute the R/S -statistic, it means:

$$\frac{R}{S}(n)=\frac{1}{S(n)}\left[\max_{0\le t\le n}(X(t)-\frac{t}{n}X(n))-\min_{0\le t\le n}(X(t)-\frac{t}{n}X(n))\right],$$

where $X(t)$ is the partial sum series. The lag values should be equally spaced, begin at every starting point of each block and must not be greater than the original series length. As in the previous case a set of series for each lag value n is obtained. The next step is then a reduction of this set to only one series containing the mean of each series. Since $M\{\frac{R}{S}(n)\}=n^H$ then a log-log plot of $\frac{R}{S}(n)$ versus n should give a straight line with slope equal to $b=H$.

The analysis of change of a Hurst's parameter depending on length of time series is analyzed in work . Various stochastic processes which are used in modeling time series, can possess various memory: short-range, long- range and infinite. Besides, the same process can include different types of time dependence. R/S - analysis allows to reveal presence some kind of memory.

The typical dependence of $Log(R/S)$ on length of time series for processes with independent increments is presented on fig. 5. The dashed line shows theoretically expected values $Log(R/S)$ for a case $H=0.5$.

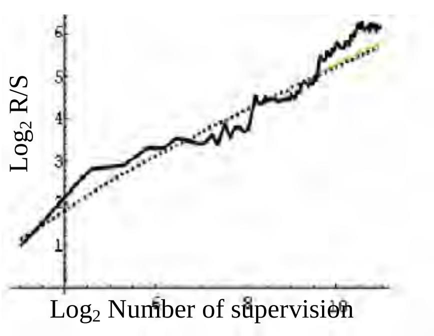

Fig. 5 Process with independent increments.

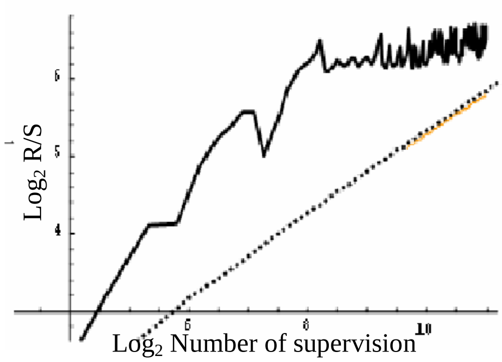

Fig. 6 Determinision chaotic process.

R/S -analysis allows to define average length of nonperiodic cycles, typical for chaotic systems. Values of a parameter H cease to change since the certain period as dynamics of system is limited by an attractor. This period is characterizes average length of a chaotic cycle. R/S -analysis for one time realization of chaotic Lorenc's attractor is shown on fig. 6.

R/S -analysis allows to reveal and eliminate short-term dependence, characteristic for autoregression processes. Linear dependence displaces values of a Hurst parameter and shows long- range memory. For removing short- range dependence is necessary a number of increments S_t to regress as a dependent variable against S_{t-1} and to find linear dependence between them. After that we analyze the residual $X_t=S_t-(a+b\cdot S_{t-1})$. If an initial time series had long-range dependence it is kept, whereas short- range dependence is removed (fig. 7).

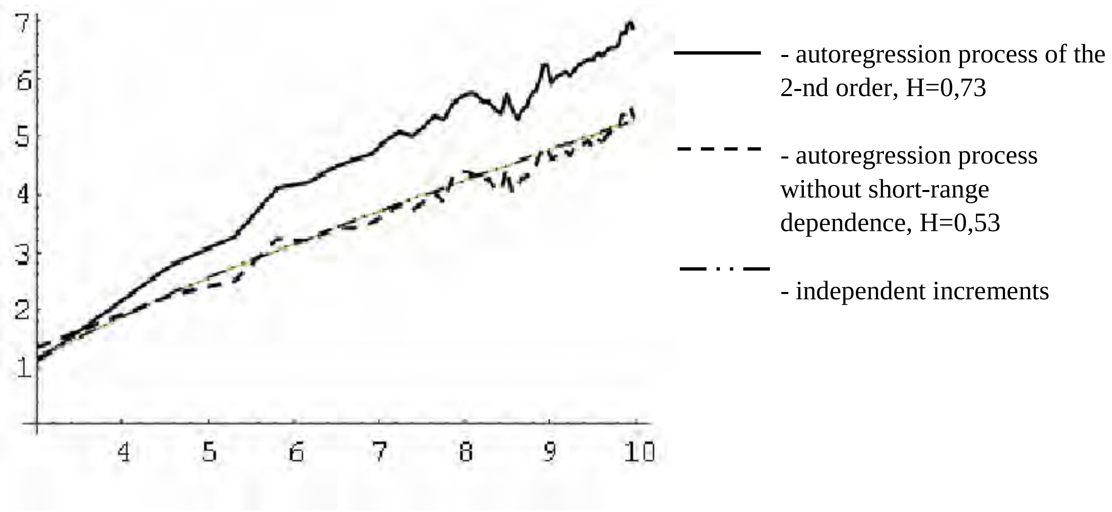

Fig.7 Deletion of short-range dependence.

Time realizations of the traffic working under report TCP have been investigated in work. Change of parameter H depending on length of realization is shown on fig. 8. Values H for various realizations are in a range $0{,}7 < H < 0{,}9$. R/S-analysis has not revealed presence of cycles typical for the determined chaotic processes. Research has shown presence of enough strong autoregression dependence $S_t = a + bS_{t-1}$, $0{,}5 < b < 0{,}6$.

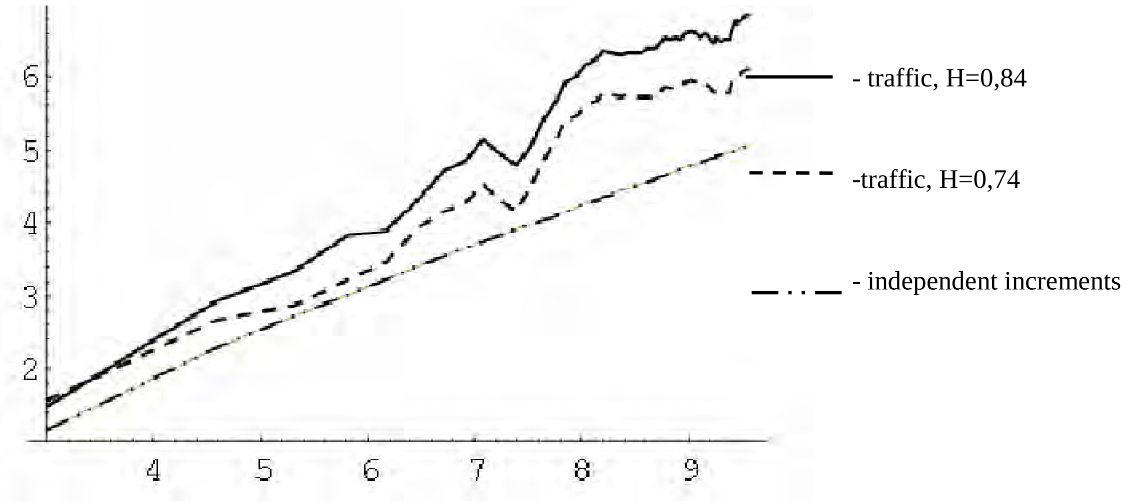

Fig.8 R/S -analysis of networking traffic.

After removal of short-term dependence, value H decreases. On the basis of the received data it is possible to draw a conclusion, that time realization of the traffic is possesses properties of long-term and short-term dependence. These properties are necessary for considering at modelling and forecasting network self-similarity the traffic.

References

[1] Peters E. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics, 2000.
[2] W.Stollings. High-speed networks and Internets. Performance and quality of service. 2002.