Self-Similar Network Traffic: From Chaos and Fractals to Forecasting and QoS

Vitaly Petroff

Abstract – It is well known nowadays, that packet network traffic has the self-similar nature [4], and conventional models, such as simple Markovian (memory-less) models, have faults. Moreover, the self-similar structure of the traffic leads to a number of undesirable effects like high buffer overflow rates, large delays and persistent periods of congestion. This paper presents the basic concept of the theory of self-similar teletraffic, its relationship with non-linear dynamics and fractals. The paper also studies the problem of providing QoS in the presence of self-similarity from the point of view the traffic prediction technique.

Keywords – Self-Similarity, Long-Range Dependence, Long memory, Forecasting, Prediction, Hurst, Heavy tails, Fractal dimension, Chaos.

I. Introduction

Numerous researches of traffic of the computer network testify that it has the property of self-similarity [1]. Meanwhile, the methods of calculation of the computer network (capacity of channels, capacity of buffers, etc.) based on the Markovian models and the Erlang formulas, which are successfully used in design of telephone networks, give unfairly optimistic solutions and result in underestimation of load. Besides, the self-similar traffic has a special structure, which holds true on many scales – there are a number of very big spikes at relatively low average levels of the traffic. The phenomenon considerably worsens characteristics (magnifies the losses, delays and jitter of packets) as the self-similar traffic passes through network nodes. In this respect, the efforts of development engineers and telecommunications standardization committees should be aimed at studying the effect of the self-similarity of the traffic and creating new algorithms for its optimal processing. This article is organized as follows: chapter II deals with an algorithm of processing the self-similar traffic. The algorithm is based on prediction and, according to the author, permits to lower losses and improve QoS characteristics. The third chapter treats the essence of the phenomenon of self-similarity, its link with the fractal structures, as well as the possible perspectives of prediction of the self-similar traffic from the point of view of the concepts of long memory, the Hurst exponent and heavy-tailed distribution. Chapter IV gives the results of the main features of the network traffic as interpreted in terms of non-linear dynamics. In particular, it is about two chaotic systems positioned as the models of the network traffic. This seems to be especially interesting, as it allows to closely approach the identification of the deterministic (possibly chaotic) component of the traffic. In chapter V, the author analyses the network traffic and compares it with the purely deterministic time series (the Lorenz chaotic system) and the purely random series (white noise) in terms of predictability. The final chapter draws major conclusions and suggests possible directions and perspectives of the further research to improve QoS characteristics under the conditions of the self-similar teletraffic.

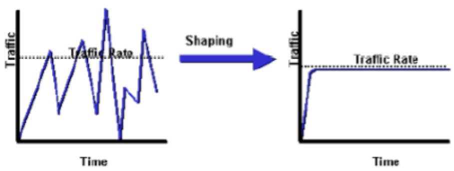
II. QoS through forecasting

Prediction of network traffics [14], [15], [17], [18] is one of the most challenging and insufficiently studied directions, and is of interest in terms of telecommunications most diversified systems and algorithms. The main ideas in the given area are connected with application of prediction to improve the mechanisms of congestion control, network safety, optimization of algorithms of dynamic routing, development of the adaptive network applications, the management efficiency and development of the network, as well as the rise in productivity in processing of the traffic in network nodes. It is the last problem that this research deals with. The number of real time applications (videoconferencing, an ip-telephony, etc.) with higher demands for the quality of network resources (packet delivery time, loss rate, jitter, free transmission range, etc.) has increased recently. Therefore, the provision of adequate quality of service attracts special attention. The author believes that the use of the prediction technique will allow solving a number of problems in this field. In fact, if we know the level of the traffic to pass through a certain network node at a certain moment of the future, we can try to process it most effectively (that is to improve quality of service characteristics). Let us consider two popular algorithms of providing QoS, representing the mechanisms of traffic regulation,

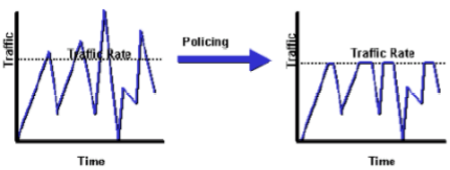
V. Petroff is with the Radio Receiving Equipment Department of the Moscow Power Engineering Institute (MPEI), Krasnokazarmennaya 13, Moscow, Russia.

E-mail: v_petrov@skc.ru, vpetroff@nm.ru

implemented by the leading telecommunications equipment company, Cisco Systems: Traffic Shaping (TS) and Traffic Policing (TP) [1]. As is clear from Fig. 1, the idea of these algorithms is as follows: TS smoothes the traffic and transfers it with a constant intensity, committed information rate (CIR), through queueing (buffering) packets with transmission rates exceeding the CIR; in its turn, the TP mechanism simply discards packets which intensity exceeds the CIR. On the one hand, as TS does not admit discarding packets, it becomes attractive for control over real time information communication (voice, real video). On the other hand, it brings buffering related delays, which has a negative effect on the characteristics of the transmitted traffic. To get rid of the above-mentioned disadvantages is enough to increase the bandwidth to the value of the maximum traffic spike, but a problem of low utilization (underexploitation of the bandwidth) will appear. Besides it is absolutely unacceptable in case of the self-similar traffic, as it always has a number of big enough spikes against the background of a relatively small average level, which is caused by heavy-tailed distribution (see below). However, the problem can have a solution - bandwidth on demand.



a)



b)

Fig. 1. Operating principles of traffic limitation mechanisms: Traffic Shaping (a) and Traffic Policing (b).

That implies early warning of the network of the requirements to the bandwidth at a certain interval in the future. The operating principle of this algorithm is shown in Fig. 2. It is worth mentioning that, in an ideal case alongside with the absence of discarding and buffering of packets, this mechanism provides a high degree of utilization of the channel. It is easy to imagine a situation where application of such algorithm should be a success: two virtual channels share a single physical link under the conditions of the limited bandwidth.

Let us allow that the first virtual channel is designed to transmit information sensitive to QoS parameters (real time traffic) and, therefore, has high priority, and the second channel is, for example, for access to ftp and http resources. Then, giving a bandwidth to the real video traffic on demand, which provides the required QoS, we will worsen the characteristics of the second channel.

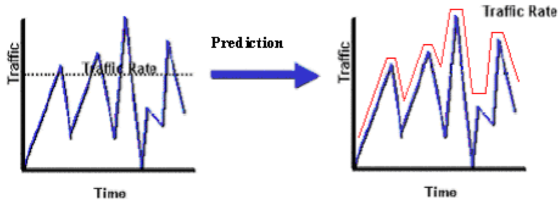


Fig. 2. Operating principle of traffic prediction.

However, that will be much less salient as http and ftp services are capable of working adequately under the conditions of big delays and losses of part of packets much due to the algorithm of guaranteed delivery of TCP protocol used on the network level, while real time network applications, as a rule, are based on UDP protocol, which has no such mechanism and, therefore, works faster (that justifies its use for these purposes). One of the realization principles of the method of bandwidth control with the help of traffic prediction [8] is shown in Fig. 3.

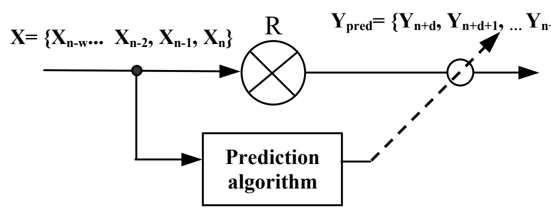


Fig. 3. Realization principle of bandwidth adaptive control.

The idea of this procedure is as follows: from the observational data $X = \{X_{n-w}...X_{n-2}, X_{n-1}, X_n\}$ of the traffic, coming into the input buffer of a certain network node R, we predict the bandwidth $Y_{pred} = \{Y_{n+d}, Y_{n+d+1}, ... Y_{n+d+m}\}$ for $d+m$ steps forward. Here, d – is the time necessary for generating the prediction, w – the length of the window. Before selecting a suitable algorithm of prediction of the network traffic we shall give the main results of research of its characteristic properties.

III. Fractal properties or self-similarity of network traffic

The notion of ***fractal*** was first introduced by Benua Mandelbrot. The important property that almost all fractals have is the property ***of self-similarity*** (scale invariance). It seems the fractal can be divided into small parts so that each part appears simply a reduction of the whole. In other words, if we look at the fractal in a microscope, we shall see the same picture, as without a microscope! The fern shown in Fig. 4 is an example of a natural fractal object. In fact, most things in existence are not circles, squares or lines. Instead, they are fractals, and the creation of these fractals is usually determined by the equations ***of chaos***. In this respect, application of the mechanism of the theory of non-linear dynamics (the theory of chaos) for research of the self-similar teletraffic also seems to be a most perspective direction and reasonable development of the ideas of fractal traffic. It is worth mentioning that the term of ***chaos*** means the word collocation ***deterministic chaos***, however, in informal conversation the word *deterministic* is frequently omitted. In this respect, the principle of determinancy can have a potential of playing a significant role not only for prediction of the network traffic and many similar processes apparent random at first sight.

Fig. 4. Fractal object: the fern

Unlike deterministic fractals, stochastic fractal objects (processes) are described as a rule by scale invariance (self-similarity) of statistical characteristics of the second order (the property of invariance of a correlation coefficient at scaling). These are such stochastic fractals that we will come across while studying the characteristics of the network traffic. In this connection in the literature the notions of fractal and self-similar teletraffic are used as synonyms often. Let us give a definition of exactly second-order self-similar process of discrete argument.

A. Self-similarity

Let $X=(X_1, X_2, \ldots)$ be a semi-infinite segment of a second-order-stationary stochastic process of discrete argument (time) $t \in N \overset{\Delta}{=} \{1,2,\ldots\}$. Let us designate through $\mu<\infty$ and $\sigma^2<\infty$ the average and the average of the process X accordingly, and through

$$r(k) \overset{\Delta}{=} \frac{\overline{(X_{t+k}-\mu)(X_t-\mu)}}{\sigma^2},$$

$$k \in Z_+ \overset{\Delta}{=} \{0,1,2\ldots\}$$

- the autocorrelation function of the process X. As the process X is second-order-stationary, the average $M[X]=\mu$, the dispersion $D[X]=\sigma^2$, the correlation coefficient $r(k)$ do not depend on time t and $r(k)=r(-k)$. Let us allow [2] that the process X has the autocorrelation function of the following kind:

$$r(k) \sim k^{-\beta} L_1(k), \quad k \to \infty \qquad (3.1)$$

with $0<\beta<1$ and L_1 being the function slowly varying at infinity. Let us mark through $X^{(m)}=(X_1^{(m)}, X_2^{(m)}, \ldots)$ - the averaged on blocks of length m process X, which components are determined by expression:

$$X_t^{(m)} \overset{\Delta}{=} \frac{1}{m}(X_{tm-m+1}+\ldots+X_{tm}), \quad m,n \in N \qquad (3.2)$$

Hereinafter we will call the process $X^{(m)}$ - ***aggregated*** process**.** Let us mark the correlation coefficient of the process $X^{(m)}$ through $r_m(k)$

Definition [2]. The process X is referred to as ***exactly second-order self-similar (es-s)*** with the parameter $H = 1-(\beta/2)$, $0 < \beta < 1$, if

$$r_m(k) = r(k), \; k \in Z_+, \quad m \in \{2,3\ldots\} \qquad (3.3)$$

That is, the es-s process does not change its correlation coefficient after it is averaged on blocks of length *m*. In other words, X – es-s, if the aggregated process $X^{(m)}$ is indistinguishable from the initial process X at least in terms of statistical characteristics of the second order.

B. Long-range dependence and forecasting

Let us consider one more notion having a key value in the theory of self-similar processes - long-range dependence (LRD). LRD describes the property of ***long memory*** [16] that is challenging in terms of prediction. At an intuitive level this property can be set forth as follows: the future of the process will be determined by its past, with a decreasing degree of influence as the past retreats from the present (that is the process with long memory a sort of forgets its relatively old past as time passes). Here come some definitions.

Definition [2]. They say, that the process X has ***long-range dependence (LRD)*** if (3.1) is satisfied. Thus, the processes with LRD are characterized by the autocorrelation function that decreases hyperbolically (under the power law), as the time delay (lag) increases.

Unlike the processes with LRD, ***processes with short-range dependence (SRD)*** have the following exponential decreasing autocorrelation function

$$r(k) \sim \rho^k, \quad k \to \infty, \quad 0 < \rho < 1 \qquad (3.7)$$

In frequency area LRD affects the characteristic power law of the behavior of the spectral density of the process. In fact, equivalently to (3.1) it is possible to state that the process $\boldsymbol{X}$ has LRD if

$$f(\lambda) \sim \lambda^{\beta-1} L_2(\lambda), \; \lambda \to 0, 0 < \beta < 1 \qquad (3.8)$$

L_2 is – a slowly varying in zero function, $f(\lambda) = \sum_k r(k) e^{ik\lambda}$ is a spectral density. Thus, from the standpoint of spectral analysis the LRD process has a spectral density with a singularity in zero (i.e. the spectral density of $f(\lambda)$ of such a process tends to infinity as frequency λ tends to zero). Such a process is frequently referred to as “*1/f* - noise” or “flicker-noise”.

C. Heavy tailed distribution and forecasting

Numerous measurements of the network traffic have shown that it best described by the so-called “heavy-tailed” distribution. To start with we will give some definitions and consider the most typical cases [4].

Definition. The random variable is considered to have ***heavy-tailed distribution*** if

$$P[Z > x] \sim c \cdot x^{-a}, x \to \infty \qquad (3.9)$$

where 0<a<2 and is referred to as a shape parameter, c - a positive constant. Unlike ***light-tailed distribution***, such as exponential or Gaussian, which has exponential decrease of the tail, heavy-tailed distribution has tails that decrease under the power law. With $0 < a < 2$, heavy-tailed distribution has an infinite dispersion, and with $0 < a \le 1$, it also has an infinite average. Speaking of the network, the case of $1 < a < 2$ is of particular interest.

In the class of heavy-tailed distributions, the Pareto distribution is most frequently used with the distribution function:

$$P[Z \le x] = 1 - \left(\frac{b}{x}\right)^a, \quad b \le x \qquad (3.10)$$

The main property of the random variable, distributed according to heavy-tailed distribution, is that it shows high variability. I.e. sample capture from heavy-tailed distribution represents mostly relatively low values, and yet it also contains quite a number of very high values. It can be shown that heavy-tailed distribution is tightly bound to the notion of long memory and LRD. Let us consider the predictability of some random variable having heavy-tailed distribution [4].

Assume that Z is a random variable having heavy-tailed distribution and interpreted as the life time (duration) of the session (TCP session, for example). Now suppose that the session is active for some time $\tau > 0$. Then the conditional probability $L(\tau)$ that the session remaining active for a time $1 \le t \le \tau$ will exist during the following $\delta > 1$ steps into the future is estimated in the case of light-tailed distribution, in particular for exponential-tailed distribution $P\{Z > x\} \sim c_1 e^{-c2x}$, as $L(\tau) \sim e^{-c2\delta}$ I.e. in the case of “exponentially light tails” the duration of activity session in the past does not impact the forecast. For heavy tails similar calculations result to

$$L(\tau) = (1 + \delta / \tau)^{-a} \qquad (3.11)$$

which means

$$L(\tau) \to 1, \quad \tau \to \infty. \qquad (3.12)$$

Thus, the more the period of observational activity of the session is, the higher is a probability that the session will continue to exist in the future. I.e. the process has persistence and, therefore, with great enough τ, the prediction error may be as small as we please.

D. *Hurst exponent and forecasting*

During centuries annual floods of the Nile were a basis of the agriculture of many known civilizations of Africa. Good irrigation means a good harvest while low water resulted in crop failure and shortage of food. Having reviewed the annals for 800 years for floods of the Nile, British official Harold Edwin Hurst detected that there was a tendency when a year of good flooding was followed by another fertile year, and, on the contrary, a year of low water was followed by another foodless year. In other words, it seemed that the foodless and fertile years were not random. To prove this fact Hurst introduced the coefficient $0 < H < 1$, which is now named after him as the Hurst coefficient. If the levels of annual floods were independent from each other, it would be logical to present the process of floods by the usual Brownian motion (BM) with independent increments, with the Hurst coefficient $H = 0.5$. However, as was found out by Hurst, for the Nile $H \approx 0.7$.

One of the ways to calculate the coefficient H is the analysis of the so-called ***R/S statistics*** (the rescaled adjusted range). Let us designate ξ an annual water level in the Nile, then the average level of water for τ years:

$$M[\xi] = \frac{1}{\tau}\sum_{i=1}^{\tau}\xi_i \qquad (3.13)$$

Let us get a new (cumulative) time series, representing the sum in time t of annual variation of water level of the Nile in reference to the average $M[\xi]$:

$$X(t,\tau) = \sum_{i=1}^{t}(\xi_i - M[\xi]), \quad 1 \le t \le \tau \qquad (3.14)$$

In this case the range between the maximum and minimum values of $X(t,\tau)$ in time τ is designated by $R(\tau)$:

$$R(\tau) = max(X(t,\tau)) - min(X(t,\tau)), \quad 1 \le t \le \tau \qquad (3.15)$$

Then the R/S statistics will be determined by the non-dimensional relation of range R(τ) to standard deviation ξ:

$$R/S = \frac{R(\tau)}{S(\tau)} = \frac{R(\tau)}{\sqrt{\frac{1}{\tau}\sum_{i=1}^{\tau}(\xi_i - M[\xi])^2}} \qquad (3.16)$$

Hurst showed that the dependence:

$$M\left[\frac{R(\tau)}{S(\tau)}\right] \sim c\tau^H, \quad \tau \to \infty \qquad (3.17)$$

holds true for many natural phenomena, with c being a positive constant not dependent from τ.

In particular, if the increments of temporal series (3.15) are independent, i.e. the time series represents BM with independent increments, then the Hurst coefficient in (3.17) is $H = 0.5$. However, for the Nile Hurst discovered that $H \approx 0.7$, which proved a certain dependence between consecutive samples ξ_i and ξ_{i+1}!

It should be pointed out that in case of $0.5 < H < 1$ one speaks of ***persistent*** behavior of the process or that the process has long memory. In other words, if during some time in the past positive increments of the process were observed, that is increase occurred, then in future on the average increase will be the case as well. In other words, the probability that the process at *i+1* step deviates from the average in the same direction as at *i* step is as great as *H* parameter is close to *1*. I.e. persistent stochastic processes show well-defined varying tendencies over relatively low noise.

In case of $0 < H < 0.5$ one speaks of ***antipersistence*** of the process. In this case the high values of the process follow the low ones, and vice versa. In other words, the probability that at *i+1* step the process deviates from the average in the opposite direction (in relation to the deviation at i step) is as great as *H* parameter is close to *0*.

With $H = 0.5$, deviations of the process from the average are really random and do not depend on the previous values, which corresponds to the case of BM.

Let us remark that it is the property of persistence that justifies application for simulation and prediction of self-similar time series of AR models as follows

$$X_n = \varphi_0 + \sum_{r=1}^{p} \varphi_r \cdot X_{n-r} + \varepsilon_n \qquad (3.18)$$

with φ_i being constants and ε_n being uncorrelated random variables (white noise) with the zero average. The formula (3.18) shows how to predict the future of the process knowing its past. In particular, autoregressive models, such as ARMA (autoregressive moving average model), ARIMA (autoregressive integrated moving average model) and FARIMA (autoregressive fractional integrated moving average model) have become widely common.

IV. Non-linear dynamic methods

Another challenging research area of the network traffic is the use of non-linear dynamic methods for its simulation and prediction [5], [19], [3], [21]. It is known that chaotic systems have the following main properties: non-linearity, determinancy and sensitivity to the initial conditions. Besides, chaotic time series looks like a stochastic process. Also, the attractor of a non-linear chaotic system is frequently fractal. If it is possible to detect the feature of deterministic chaos in the traffic, we will obtain a new model of the traffic and a new algorithm of its prediction due to the chaos deterministic nature. In 2000 Andras Veres and Miklos Boda from Ericsson Research [3] performed an interesting work in which with the help of ns-2 simulation modeling it is demonstrated that the traffic model of TCP protocol (TCP Tahoe version was used) can be both a simple periodic process and, under some conditions, have a complex behavior compatible with the concept of ***deterministic chaos***. In particular, the researchers obtained a trajectory of the system in phase space (Fig. 6) that they referred to the class of strange attractors. An ***attractor*** is a cluster set of trajectories in the phase space of the system to which all the trajectories from a neighborhood of this set tend.

The work demonstrated that the Hausdorff dimension of such an attractor exceeds *1*, but not so much as *2*, specifically is equal to *1,61*. Besides, for simultaneous operation of *30* TCP sessions they estimated the Lyapunov exponent $\lambda \approx 1.11$ meaning that after perturbation is introduced, the distinction between the sessions accrues with the average rate of $e^{\lambda} \approx 3.03$ per second, which is the evidence of sensitivity to the initial conditions.

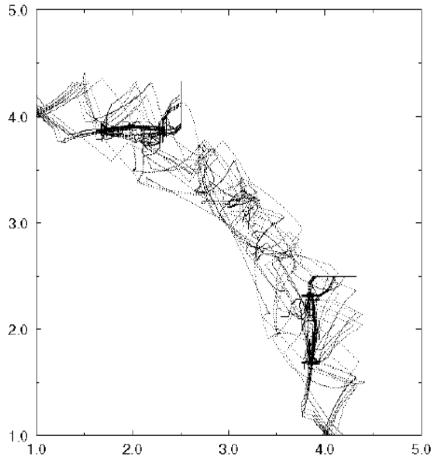


Fig. 6. "Strange" attractor for TCP model connection.

The combination of these two facts (the non-integer dimension and the positivity of the Lyapunov exponent) gives the researchers reason to speak about strangeness of the attractor and, as a consequence, about the presence in the system the features of deterministic chaos.

Let us remark, however, that from the chaos theoretical stand, the trajectory of a strange attractor must not be periodic. Therefore, as the attractor represented in fig. 6 is periodic (though its period is long enough - about 4 hours), let us refer to it as almost strange (i.e. "strange"). Still the work is a very important step forward, as it demonstrates that a system of TCP sessions simultaneously operating on one connection may enter, on some conditions, to the mode of deterministic chaos and make the traffic having invisible order but which looks like an absolutely random process and was simulated earlier with the use of the theory of stochastic processes. Here it is important that it is a complex, looking like random, but, at the same time, deterministic process. Setting absolutely precise initial conditions, we can repeat this process as many times as possible, with the trajectories being absolutely identical. However, if arbitrarily small deviation from the initial conditions occurs, the trajectories diverge, with the distance between them in time increasing exponentially. But again, the system future is always completely determined by its past.

The possibilities of dynamic systems in simulation of the network traffic are also studied in a series of research works by A. Erramilli and others [5]. They deal with the properties of chaotic maps as follows

$$x_{n+1} = f_1(x_n), y_n = 0, (0 < x_n < d)$$

$$x_{n+1} = f_2(x_n), y_n = 1, (d < x_n < 1)$$

with $f_1(.)$ and $f_2(.)$ being functions that satisfy the requirement of sensitivity to the initial conditions.

Besides, the construction of the model implies the source of the traffic to be actively or passively depending on whether x_n is more or less than limit d. The adequacy of the given model is confirmed by its characteristic properties of long-range dependence and heavy-tailed distribution. For one of the sets of parameters of chaotic map the correlation dimension of the attractor is estimated as *0.91*.

From the standpoint of traffic control, the presence the chaotic regimes in the traffic means a theoretical possibility of its prediction but if the precise dependence is established, certainly. It is necessary to note, however, that this approach to prediction has the following disadvantage: since it is physically impossible to ensure absolutely accurate initial conditions for the predictive function that correspond to this traffic, the discrepancy between the real and predicted traffic will quickly increase as the prediction interval (time covering the prediction of the process) increases. And, as follows from the theory of non-linear dynamics, this increase will occur under the exponential law.

V. Results

The analysis of the network traffic is actually reduced to the task of processing the time series. In turn, the theory of non-linear dynamics provides a potential to study, identification and prediction of the time series that have some specific properties.

The application of the Takens theorem on embedding the attractor into spaces of various dimensions is one of the key concepts of the theory of non-linear dynamics. This method allows to reconstruct the parameters of the dynamic system from one-dimensional time series by studying the system trajectory in m-dimensional phase space, with coordinates being the components of the following vector: $Z^{(m)} = \{X_i, X_{i+\tau}, \dots X_{i+(m+1)\tau}\}$, where τ is a time delay. This operation is referred to as embedding the attractor into m-dimensional space. A successful embedding results to the specific behavior of the system trajectory in the space of the given dimension. The absence of the specific behavior means either an incorrect selection of the space dimension or nonpossession of the system of the attractor. As the world around us has only three dimensions, we can imagine embedding the attractor into the space which dimension does not exceed three. However, it may be required to embed the attractor into spaces of greater dimension and select the most suitable dimension. For this purpose there are some special methods. One of them is the method [13] of ***False Nearest Neighbors (FNN)***. This method is designed for determination of the minimum acceptable dimension of the embedding space. Its principle is clear at an intuitive level. Let $z_i^{(m)}$ and $z_j^{(m)}$ be two near neighbors in the reconstruction of dimension m (that is $\left\| z_i^{(m)} - z_j^{(m)} \right\|$ is not enough), and $z_i^{(m+1)}$ and $z_j^{(m+1)}$ be in accord with them in reconstruction $m+1$. If we deal with really near neighbors, as a rule, they are close in both reconstructions. If the neighbors close in reconstruction m become distinct in reconstruction $m+1$ ($\left\| z_i^{(m+1)} - z_j^{(m+1)} \right\|$ - too much), they are referred to as false nearest neighbors. If now we increase m and estimate the quantity of FNN, then when we reach the necessary dimension at which the correct reconstruction is achieved, that quantity sharply decreases. It is obvious that from a diagram of dependence of the quantity of false nearest neighbors on the dimension of the embedding space (on its minimum or decrease in zero) we can make a conclusion about the minimum possible dimension of the phase space. Fig. 7 shows typical FNN dependences for self-similar network traffic (*BC-Oct89Ext.TL* [9]), as well as for white noise and deterministic chaos (in the form of the Lorenz system). Analyzing visually the curves of Fig. 7 it is possible to notice that the characteristics of the real traffic are between the situations of a complete disarray (white noise) and the complete order (of the Lorenz system), which is an illustrative example and points out to the possibilities to apply the methods of non-linear dynamics when studying the traffic.

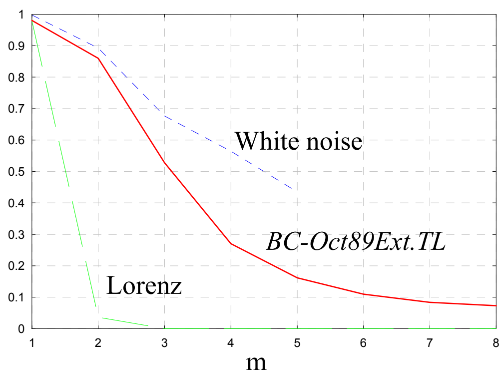


Fig. 7. FNN diagrams for the traffic, Lorenz chaotic system and white noise.

The concept of the correlation integral that allows estimating the dimension D_2 of attractor, embedded in the space of dimension m, is another basic concepts of non-linear dynamics. The estimation of the attractor dimension from a scalar time series is of interest in turn, as it allows estimating the minimum number of the essential dynamic variables necessary for the description of the process. The most popular algorithm for the calculation of correlation integral was offered by P. Grassberger and I. Prokaccia and is based on relation (5.1)

$$C(\varepsilon, m) = \frac{\text{quantity of pairs with } \left\| z_i - z_j \right\| < \varepsilon}{\text{total quantity of pairs } z_i, z_j} \quad (5.1)$$

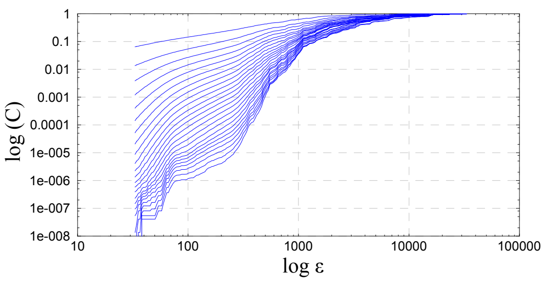
In this case z_i, z_j are vectors of the coordinates of points in the phase space of dimension m. This being the case the following relation holds true

$$C(m, \varepsilon) \sim \varepsilon^{D2}$$

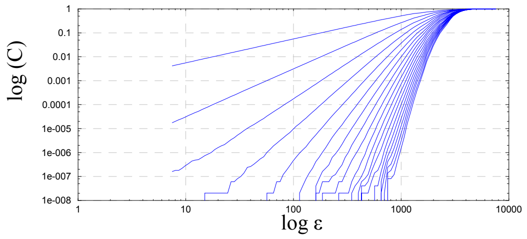
permitting to estimate dimension D_2 of the attractor by the slope of the most linear region of the diagram

$$\log C(\varepsilon, m) \cong -D_2 \log \varepsilon + const$$

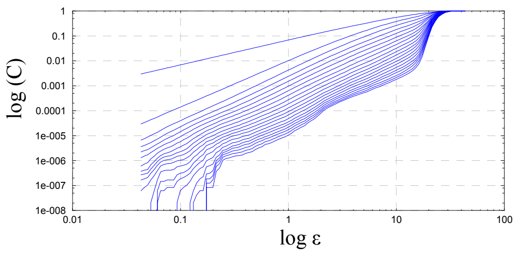
As m increases, such estimation of D_2 must tend to the true value of the correlation dimension of the attractor [20]. Fig. 8 shows a family of curves of correlation integral $C(\varepsilon)$ for the Lorenz system, white noise and *BC-Oct89Ext.TL* traffic. The parameter of the family is the dimension m of embedding.



a)



b)

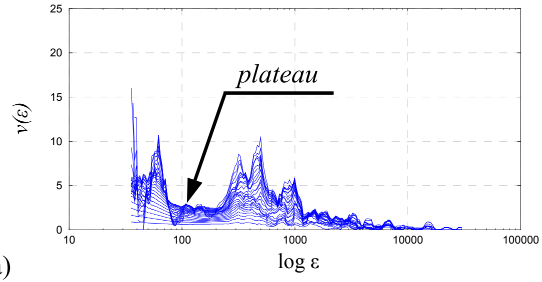


c)

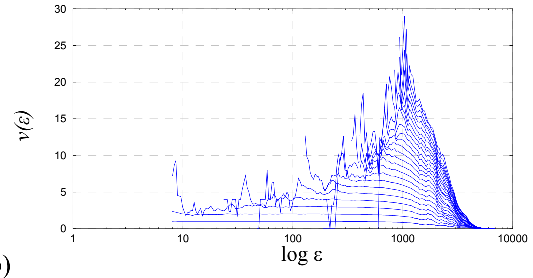
Fig. 8. Correlation integral diagrams for
a) *BC-Oct89Ext.TL* traffic
b) white noise
c) Lorenz system

Studying the dependences it is possible to make a conclusion that for the traffic as well as for the Lorenz system and in the contrary from the case of white noise there is characteristic curving. This curving is the cause of a plateau on the diagram of the local slopes for correlation integral (Fig. 9).

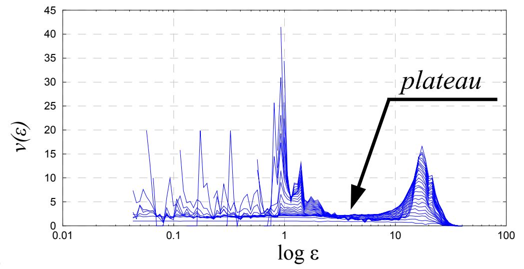
So, as clear form Fig. 9a, for the traffic case the most adequate dimension of attractor is $D_2 \sim 2.5$. But its value is varying from one traffic to another. Nevertheless, this result is holds true for most realization of traffic and has important significance, because allows to identify traffic as deterministic process with some level of noise.



a)



b)



c)

Fig. 8. Local slopes of the correlation integral for
a) *BC-Oct89Ext.TL* traffic
b) white noise
c) Lorenz system

As the determination of the statistical relation between the previous and following samples of time series is the primary problem for the prediction the traffic successfully, it is necessary to check up the statement about a statistical dependence between consecutive samples of the time series. The so-called BDS-test [10] based on the correlation integral properties has become rather famous. For this purpose the following statistics (5.2) is calculated

$$w_{m,n}(\varepsilon) = \sqrt{n - m + 1}\, \frac{C(\varepsilon, m, n) - C^m(\varepsilon, 1, n)}{\sigma_{m,n}(\varepsilon)} \quad (5.2)$$

Under the conditions of a null hypothesis about the independent and identical distributed (i.i.d) samples of the time series, statistics $w_{m,n}(\varepsilon)$ has normal distribution $N(0,1)$. By the deviation from this distribution, a statistical relation between the consecutive samples of time series is judged. Let us carry out the test for the traffic, Lorenz system and white noise time series by means of BDS program by D. Dechert. The results of calculations are given in Chart 1.

Studying the results it is possible to conclude that the null hypothesis of i.i.d for the Lorenz system and the network traffic time series is rejected at the *5 %* significance level as $|w_{m,n}(\varepsilon)| < 1.96$. While in the case of white noise the null hypothesis is not rejected for any value of m and ε.

a)

ε	m	w
0.394670	3	1.9851E+0001
0.394670	4	2.6883E+0001
0.394670	5	3.2429E+0001
0.394670	6	3.6749E+0001
0.394670	7	4.0058E+0001
0.394670	8	4.2538E+0001
0.343826	2	1.4466E+0001
0.343826	3	2.4942E+0001
0.343826	4	3.2417E+0001
0.343826	5	3.7635E+0001
0.343826	6	4.1159E+0001
0.343826	7	4.3417E+0001
0.343826	8	4.4737E+0001
0.299532	2	1.6972E+0001
0.299532	3	2.7988E+0001
0.299532	4	3.4943E+0001
0.299532	5	3.9130E+0001
0.299532	6	4.1441E+0001
0.299532	7	4.2501E+0001
0.299532	8	4.2732E+0001
0.260944	2	1.8564E+0001
0.260944	3	2.9283E+0001
0.260944	4	3.5168E+0001
0.260944	5	3.8082E+0001
0.260944	6	3.9204E+0001
0.260944	7	3.9270E+0001
0.260944	8	3.8723E+0001
0.227327	2	1.9361E+0001
0.227327	3	2.9221E+0001
0.227327	4	3.3809E+0001
0.227327	5	3.5498E+0001
0.227327	6	3.5647E+0001
0.227327	7	3.5013E+0001

b)

ε	m	w
0.127511	3	6.8946E+0000
0.127511	4	1.0213E+0001
0.127511	5	1.3261E+0001
0.127511	6	1.6072E+0001
0.127511	7	1.8668E+0001
0.127511	8	2.1102E+0001
0.111111	2	3.8698E+0000
0.111111	3	7.7632E+0000
0.111111	4	1.1325E+0001
0.111111	5	1.4573E+0001
0.111111	6	1.7537E+0001
0.111111	7	2.0255E+0001
0.111111	8	2.2774E+0001
0.096821	2	4.2689E+0000
0.096821	3	8.4566E+0000
0.096821	4	1.2206E+0001
0.096821	5	1.5610E+0001
0.096821	6	1.8697E+0001
0.096821	7	2.1473E+0001
0.096821	8	2.3998E+0001
0.084368	2	4.5665E+0000
0.084368	3	8.9713E+0000
0.084368	4	1.2909E+0001
0.084368	5	1.6430E+0001
0.084368	6	1.9594E+0001
0.084368	7	2.2420E+0001
0.084368	8	2.4941E+0001
0.073517	2	4.8576E+0000
0.073517	3	9.4901E+0000
0.073517	4	1.3595E+0001
0.073517	5	1.7198E+0001
0.073517	6	2.0404E+0001
0.073517	7	2.3181E+0001

c)

ε	m	w
0.280897	3	8.6298E-0002
0.280897	4	1.7640E-0001
0.280897	5	2.4202E-0001
0.280897	6	2.6935E-0001
0.280897	7	2.6535E-0001
0.280897	8	2.8305E-0001
0.244361	2	-1.8969E-0003
0.244361	3	8.6625E-0002
0.244361	4	1.7790E-0001
0.244361	5	2.3287E-0001
0.244361	6	2.4508E-0001
0.244361	7	2.2344E-0001
0.244361	8	2.2093E-0001
0.212577	2	-1.0176E-0002
0.212577	3	8.4664E-0002
0.212577	4	1.6756E-0001
0.212577	5	2.0341E-0001
0.212577	6	2.0031E-0001
0.212577	7	1.6980E-0001
0.212577	8	1.5283E-0001
0.184928	2	-1.6581E-0002
0.184928	3	7.5984E-0002
0.184928	4	1.4147E-0001
0.184928	5	1.6165E-0001
0.184928	6	1.4856E-0001
0.184928	7	1.1643E-0001
0.184928	8	9.4440E-0002
0.160874	2	-2.5229E-0002
0.160874	3	5.6593E-0002
0.160874	4	1.0341E-0001
0.160874	5	1.0948E-0001
0.160874	6	9.1606E-0002
0.160874	7	6.4733E-0002

Chart 1. Calculation results of BDS- statistics
a) For the Lorenz system
b) For *BC-Oct89Ext.TL* network traffic
c) For white noise

When analyzing of the traffic of the real network, the author noticed the presence in the spectrums of the traffic time series of some harmonic components with levels that are not compatible with the concept of flicker-noise (3.8). A typical power spectrum W(f) of time series for the traffic LBL-PKT-5.TCP [9] (averaged on blocks of 0.1 sec.) is given in Fig.10.

Analyzing the given spectrum it is possible to mark the presence of a strong harmonic component with the frequency of *~5 Hz*. As the aggregation level decreases, it is also possible to mark the presence of the harmonic of *~10 Hz*. Similar observations can be made on studying LBL-TCP-3 [9] traffic.

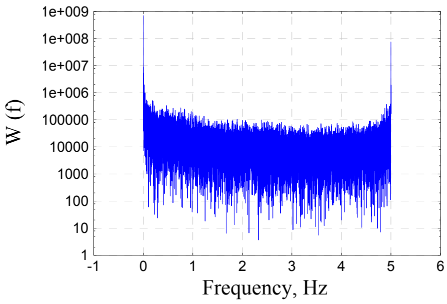


Fig. 10. LBL-PKT-5.TCP time series on a logarithmic scale on a y-axis.

Let us mark that in [6] and [7], when studying the traffic, similar results are obtained and the presence of harmonics with frequencies of *~ 11.9 Hz* and *1.25 Hz* in the wan-traffic is revealed.

The discovered phenomenon has an important significance as it reveals the presence of a regular deterministic component in the network traffic, which may be of interest in solving the problems of teletraffic forecasting.

VI. Conclusions

This work looks into the possibilities to provide the necessary level of QoS under the conditions of the self-similar traffic, characteristic of telecommunication networks with packet transmission. In particular, an algorithm based on traffic prediction and capable of enhancing the effectiveness of traffic processing in network nodes from the point of view of QoS provision is at issue. The work covers some results from the theory of the self-similar teletraffic pointing out to the possibilities of traffic forecasting. At the same time the network traffic is considered from the standpoints of the theory non-linear dynamics: while the majority of existing methods of traffic simulation are based on the models of stochastic processes, this work indicates the possibilities of traffic simulation by means of purely deterministic (chaotic) objects. In turn, the identification of the deterministic (possibly chaotic) component in traffic will result into more accurate forecasts. Also, the application of the FNN test and the calculation of the correlation integral demonstrate that the traffic of the real network (as well as chaotic models) has deterministic structure, possibly. And, it cannot be related to the class of purely stochastic processes. In other words, analyzing the obtained characteristics it is possible to classify the traffic as a strongly noisy process having some deterministic (chaotic perhaps) features. By means of calculation of BDS-statistics the work demonstrates that the hypothesis on a statistical independence of the consecutive samples of time series is rejected for all recommended values of ε and m at the *5%* significance level. This fact enables to draw a conclusion on basic possibility to predict the traffic. Besides, the work examines a power spectrum of a time series of the aggregated network traffic. The spectrum possesses a strong harmonic component at the frequency of *~5 Hz*. In other words, a principle deterministic component in the traffic shows up. This fact may also has significance in terms of prediction.

REFERENCES

[1] Policing and Shaping Overview, QC: Cisco IOS Release 12.0 Quality of Service Solutions Configuration Guide http://www.cisco.com

[2] W.E.Leland, M.S.Taqqu, W.Willinger, and D.V.Wilson. On the self-similar nature of ethernet traffic (extended version).IEEE/ACM Transactions of Networking, 2(1):1-15,1994.

[3] A. Veres, M. Boda. The Chaotic Nature of TCP Congestion Control // Proceedings of IEEE INFOCOM'2000, March 2000.

[4] K. Park, W. Willinger. Self-Similar Network Traffic and Performance Evaluation, John Wiley & Sons, 2000.

[5] A. Erramilli, R.P. Singh, P. Pruthi. Chaotic Map As Model of Packet Traffic. 1994

[6] G. Olowoyeye, B. Kim, K. Chandra. Modelling Spectral Features in TCP Traffic

[7] K. Chandra, C. You, G. Olowoyeye and C. Thompson, Non-linear Time-Series Models of Ethernet Traffic. Submitted to INFOCOM’99, July 1998.

[8] G. Chong, S. Li and J. Grosh. Predictive Dynamic Bandwidth Allocation for Efficient Transport of Real-Time VBR Video over ATM. IEEE Journal on Selected Areas in Communications. Vol .13, No 1, January 1995, pp.12-23.

[9] http://ita.ee.lbl.gov/ -Internet traffic archive

[10] W. A. Brock, W. D. Dechert, J.A. Scheinkman and B. LeBaron. A test for independence based on correlation dimension, Econometric Reviews 15: 197-235, 1996

[11] Tsybakov B., Georganas N. Self-similar processes in communications networks // IEEE Trans. Inform. Theory, vol. 44, pp. 1713--1725, Sep. 1998.

[12] The TISEAN software is publicly available at http://lists.mpipks-dresden.mpg.de/~tisean/TISEAN_2.1

[13] M. B. Kennel, R. Brown and H. D. I. Abarbanel, Determining embedding dimension for phase-spase reconstruction using a geometrical construction, Phys. Rev. A 45, 3403 (1992)

[14] Östring S., Sirisena H., The Influence the Long-Range Dependence on Traffic Prediction, Proceedings of ICC'01, Helsinki, June 2001.

[15] Miloucheva I., Müller E., Anzaloni A., A practical appoach to forecast Quality of Service parameters considering outliers. 2003.

[16] Beran J. Statistical Methods for Data with Long-Range Dependence. //Statistical Science, Volume 7, Issue 4 (Nov., 1992), 404-416.

[17] H. Zhao, N. Ansari and Y.Q. Shi. A Fast Non-linear Algorithm for Video TrafficPrediction. ITCC’ 2002.

[18] B. Chen, S. Peng, K. Wang. Traffic Modeling, Prediction, and Congestion Control for High-Speed Networks: A Fuzzy AR Approach. IEEE *Trans. On Fuzzy Systems* Vol. 8, №5, Oct. 2000.

[19] A. Erramilli, R.P. Singh, P. Pruthi. An Application of Deterministic Chaotic Maps to Model Packet Traffic. Queueing Systems, 20(3):171-206, 1995.

[20] J. Theiler. Some comments on the correlation dimension of $1/f^{\alpha}$ noise // Phys. Lett. A. **155** (1991). P480-493.

[21] Kugiumtzis D. Boudourides M. Chaotic Analysis of Internet Ping Data: Just a Random Generator? // SOEIS meeting at Bielefeld, March 27-28, 1998.