SELF-SIMILARITY OF TRAFFIC IN EDUCATIONAL NETWORK

Liudvikas Kaklauskas
Šiauliai State College, Šiauliai University
Lithuania

Annotation
The present article deals with statistical university network traffic, by applying the methods of self-similarity and chaos analysis. The object of measurement is Šiauliai University LitNet network node maintaining institutions of education of the northern Lithuania region.
Key words: *correlation, Hurst coefficient, fractality, fractal measure, cluster analysis, standard deviation, persistent process, chaos, burstiness and jitter.*

Introduction

Empirical research of computer network packet traffic shows that it is attributed with self-similarity (Domańska and etc. 2015, Gallos, and etc. 2007, Петров, 2003, Park, Willinger, 2000, Erramilli and etc. 1996, Leland and etc. 1994). After estimating the latter feature, it is possible to adequately prognosticate the change of traffic and to apply the prognosis results in increase of network throughput and improvement its QoS quality of service, while regulating packet latency, fluctuation restriction and packet loss transportation on data and physical OSI layers (Cenedese, and etc. 2015, Miesowicz, and etc. 2015, Gebali, 2015, Kaklauskas, Sakalauskas, 2011, Sakalauskas, Kaklauskas, 2010, Bárány, 2009, Kaklauskas, Sakalauskas, 2008, He, and etc. 2004,).

In contemporary university studies, computer networks are widely applied; they often undergo non-prognosticated overload. For effective network control, it is necessary to perform monitoring of network nodes in order to prognosticate network node load and overload. Research suggests that classical Markov's models which are widely used while estimating indexes of classical telephone networks are no longer suitable for modelling parameters of modern computer network parameters because parameter measures are imprecise and lead to unreasoned prognoses (Kaj, 2002). On the base of A. Erramilli, O. Narayan and W. Willinger, in 1989, by empirical research of Ethernet local area network of 10 Mbps which was carried out at Bellcore laboratory (120 persons using network services, New Jersey State) it was estimated that Ethernet traffic characterisations bear fractal characteristics and are attributed with self-similarity with long-range dependence (Erramilli, and etc. 1996). Ingemar Kaj (2002) in the monoFig.s suggests the methods of statistical analysis of characteristics of modern communication traffic, by applying possibilities of contemporary mathematical modelling. V. V. Petrov analysis network traffic as a fractal process attributed with a statistical self-similarity which is characterised by a fractal measure (Петров, 2003). Methods of non-linear (chaos theory) are applied for modelling and description of network processes, while estimating the so-called heavy-tails which characterise large burstiness of network traffics.

The aim of this research is to analyse measurement results of Šiauliai University LitNet network node traffic and to estimate its self-similarity. It should be noted that analogous analysis of network data, in order to find out about the self-similarity, has not been carried out yet in Northern Lithuania. Programme and device tools for monitoring the network were used in analysis of network traffic; these tools registered data packets at the indicated interval. *ulogd* software for *Linux* operational system distributed under GPL licence was used for measurement (Welte, Ayuso, 2015). Data was registered while applying different levels of time discretisation, at different levels of network load present, while forming aggregated time queues. Measurement results were processed by calculating correlation measure of the queue, analysing attractors, estimating the fractal measure, calculating Hurst coefficient and statistically estimating reliability of analysis results. Queue analysis shows that measured traffics are attributed with self-similarity and high burstiness which strongly influence the patterns of loss and delay latency and loss of data packets in the network (Leland, Wilson, 1991).

1. The model for measurement of network traffics

For measurement of network traffics, Šiauliai University LitNet network node with the highest intensity of traffic load was chosen. In this node, received inter-city channel traffic of 1 Gbps is distributed to the university and educational institutions of Šiauliai region (see Figure 1). Only data packets arriving at the node M, while disregarding sent packets, were analysed. Obtains information was collected in external data base Porstgree SQL (DB). Initial measurement was carried out with exactness of one microsecond. Record on the data base was formed right after receiving TCP or other protocol's data frame. Service information was not withdrawn while saving framework data: title, feature of framework beginning, addresses of a sender and receiver, etc. The biggest length of fixed transport frameworks was up to 1518 bytes (Liu, and etc. 2015).

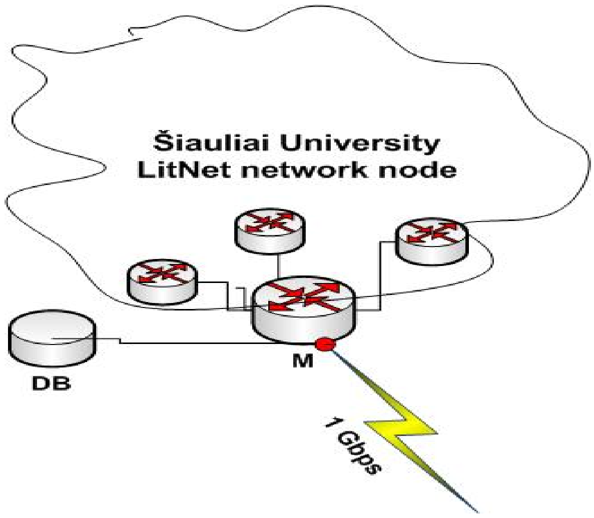


Fig. 1. The structural scheme of the network node analysed

ulogd software for Linux operational system distributed under GPL licence was used for measurement (Welte, 2000). *ulogd* is a userspace logging daemon for netfilter / iptables related logging. This includes per-packet logging of security violations, per-packet logging for accounting purpose as well as per-flow logging. Data was being fixed in incoming data traffic drives of the router M. Callback function called for every packet traversing certain point within network stack. Every pre-routed packet is registered by *ulogd* daemon in PostgreeSQL database (see Figure 2).

Data from January 4, 2008 13:30:35 to April 16, 2008, 12:00:00 were chosen from the data base for analysis. Within this period, more than three billion records were accumulated in the data base; they corresponded to 8936965 seconds or 103 days 10 hours 29 minutes and 25 seconds. Data for analysis was selected according to days of the week and part of the day, while estimating intensity of data traffic, i.e. those time queues were selected when data traffic was the least, medium or highest.

Analysis of intensity of data traffic suggests that the highest data traffic in the node is observed on week days, the least – on Sundays, and Saturday occupies the medium position.

While investigating the change of load during the day, hours when data traffic is the highest, medium and least were indicated for every day of the week. The method of k-averaged of cluster analysis was applied in analysis of intensity of data traffic in selected time intervals, while using Statistical Package for Social Sciences (SPSS). The method of non-hierarchic cluster analysis was applied, when the amount of cluster figures (k=24) is known, and distances between clusters and objects are calculated by using Euclidean square range metrics.

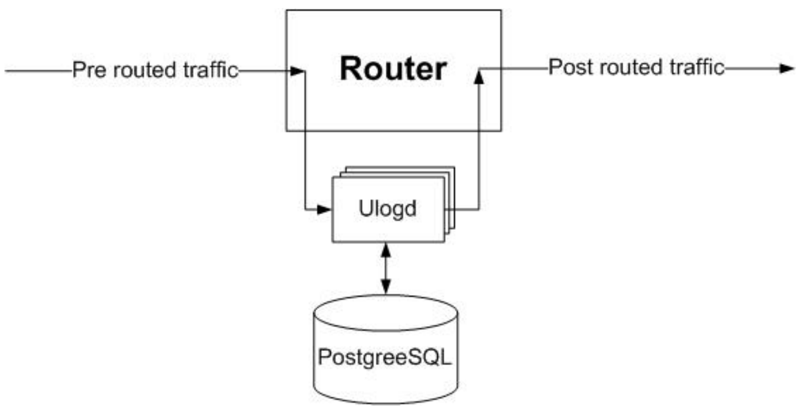


Fig. 2. The scheme of network traffic registration

A queue of one hour measurement consists of up to one-and-a-half million records. Due to technical problems inside the network, measurements of some of the days were omitted. 309 hours of when data traffic is the highest, medium and least were selected for further analysis.

2. Formation of time queues

The time queue obtained corresponds to measurements performed in time moments $\tau_1, \tau_2, ..., \tau_n$, here n is an amount of saved packets, and intervals between τ_i are unequal, i.e. $\tau_i - \tau_{i-1} \neq \tau_{i+1} - \tau_i$, where $i \in [1, n]$. Data packets $x_1, x_2, ..., x_n$ in certain time periods characterise business of the channel in various time moments. In order to analyse such time queue, it must be aggregated, i.e. to calculate data traffics in equal time intervals $t_i - t_{i-1} = t_{i+1} - t_i$. For aggregation of data, two methods were chosen.

The first one, the method of smoothing of moving surfaces was applied when an average traffic for a data queue is calculated in a chosen time interval Δt:

$$x_k^{\Delta} = \frac{\sum_{\tau_i \in [t_k, t_{k-1}]} x_i}{\Delta t},$$

here $t_k = k \cdot \Delta t + \tau_1$. Obtained time queues characterise average changes of data traffic in time moments Δt.

The second one applied transferred data traffic amount within the time interval Δt:

$$x_k^{\Sigma} = \sum_{\tau_i \in [t_k, t_{k-1}]} x_i,$$

here $t_k = k \cdot \Delta t + \tau_1$.

While forming the queues for the research, the time intervals $\Delta t \in$[100ms, 500ms, 1s] were chosen. After evaluating results of cluster analysis, the time queues when network load is the lowest – $x_1^{\min}, x_2^{\min}, ..., x_n^{\min}$, medium – $x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}$ and highest – $x_1^{\max}, x_2^{\max}, ..., x_n^{\max}$ were singled out. Out of 309 selected measurement sequences, 6 queue groups were formed, totally 1854 queues; in every, network load measures of one hour were used:

- Δt = 100ms, $x_t^{\Sigma} \in \{x_1^{\min}, x_2^{\min}, ..., x_n^{\min}; x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}; x_1^{\max}, x_2^{\max}, ..., x_n^{\max}\}$;
- Δt = 100ms, $x_t^{\Delta} \in \{x_1^{\min}, x_2^{\min}, ..., x_n^{\min}; x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}; x_1^{\max}, x_2^{\max}, ..., x_n^{\max}\}$;
- Δt = 500ms, $x_t^{\Sigma} \in \{x_1^{\min}, x_2^{\min}, ..., x_n^{\min}; x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}; x_1^{\max}, x_2^{\max}, ..., x_n^{\max}\}$;
- Δt = 500ms, $x_t^{\Delta} \in \{x_1^{\min}, x_2^{\min}, ..., x_n^{\min}; x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}; x_1^{\max}, x_2^{\max}, ..., x_n^{\max}\}$;
- Δt = 1s, $x_t^{\Sigma} \in \{x_1^{\min}, x_2^{\min}, ..., x_n^{\min}; x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}; x_1^{\max}, x_2^{\max}, ..., x_n^{\max}\}$;
- Δt = 1s, $x_t^{\Delta} \in \{x_1^{\min}, x_2^{\min}, ..., x_n^{\min}; x_1^{\Delta}, x_2^{\Delta}, ..., x_n^{\Delta}; x_1^{\max}, x_2^{\max}, ..., x_n^{\max}\}$.

Aggregated time queues, while estimating network load, are marked as follows: $x_t^{\Sigma_{\min}}$, $x_t^{\Delta_{\min}}$ – the lowest load, $x_t^{\Sigma_{\Delta}}$, $x_t^{\Delta_{\Delta}}$ – the average load and $x_t^{\Sigma_{\max}}$, $x_t^{\Delta_{\max}}$ – the highest load, when $\Delta t \in$[100ms, 500ms, 1s].

For estimation of time queues, the programme Fraktan 4.4 worked out by V. Sychov in 2003 was applied (Бельков, and etc. 2011, Богутский, Незамова, 2010, Петров, Богатырев, 2003). In the aggregated queues, the programme calculates the following: correlation and phasic space measure, auto-correlation function by singling out optimal time lag, correlation entropy, Hurst coefficient and fractal measure, presents Fig.ic image of numerical values and draws obtained attractors.

3. Research of attractors of network traffic time queues

The formed computer network data traffic queues remind of a determined chaos which depends on initial network parameters selected by an administrator: channel throughput, efficiency of network nodes, flexibility of used data transfer protocol, bug correction, defined rules on node protection, etc. The change of lengths of data packets transferred through the computer network node M and fixed by the programme ulogd is dynamic, hardly predictable and reminds of noise (see Figure 3). It is impossible to characterise it by traditional mathematical methods. For analysis of such queues, the chaos theory is the most suitable; it was scientifically reasoned and started to be developed by E. Lorenz (Massachusetts Institute) in his works of the late second decade and a French-American mathematician B. B. Mandelbrot. The chaos theory is based on two statements: it is impossible to precisely estimate the future due to measurement errors and no knowledge on all initial conditions; reliability of worked out prognoses rapidly decreases as time goes by. The main instruments of this theory are attractors and fractals which tend to describe dynamic systems, disregarding classical theories. An attractor is a geometric figure characterising behaviour of a system in phasic space when time draws closer to infinity. Phasic space is an abstract space the coordinates of which define degrees of the analysed system's unrestraint. While computer-aided analysing the system defined by three levels of unrestraint made up of three simple differential equations with three constants and three initial conditions, E. Lorenz described the first chaotic or strange attractor. Later, a Belgian-French mathematician-physician D. P. Ruelle and a German mathematician F. Takens described them in their works (Ruelle, Takens, 1994). The strange attractor bears certain maximum limits characterised as attractor's dimension, and its geometric expression is a fractal. Then, for evaluation of the strange attractor, one can apply the following features of a fractal: symmetry which defines self-similarity and fractal measure which is expressed through a fraction is a mathematical measure of fractal's inaccuracy (Riddle, 2014, Feder, 2013, Olsen, 2000).

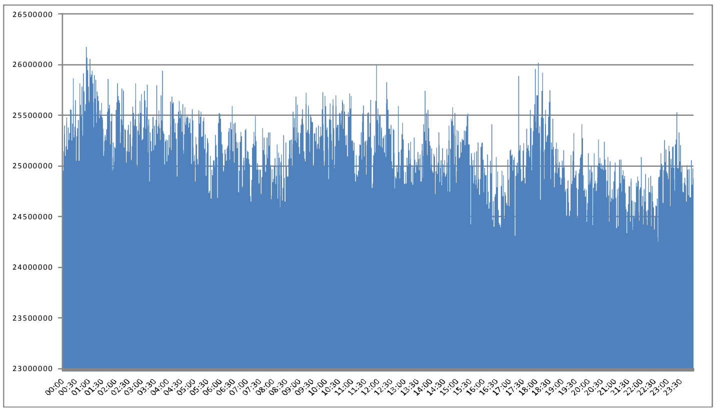


Fig. 3. Characteristic Fig.ic expression of one day's network traffic

While estimating self-similarity of measured data traffic of the network node *M*, the present section calculates attractors' dimensions and draws their charts. For estimation of attractors' dimension, we calculate Hausford measure which is obtained by analysing the strange Lorenz's attractor. For estimation of the system, we calculate Hausford measure *D* (Feder, 2013, Bárány, 2009, Beran, 1994):

$$D = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)},$$

here *N* is minimum amount of n-time blocks with ε facet length which cover points of a set, when facet length draws close to zero. This measure can be called a fractal measure. We analyse the system, when $1<D<2$, then the formula is as follows:

$$D = \frac{\ln N}{\ln(1/(2*r))},$$

here *N* – amount of elements used fro measuring fractal's undulation, $N \to 2$ when a fractal is in plane, *r* – radius of a circle used in 2-dimensional space. In computer networks, fractal measure characterises dynamics of formed data traffic time queue's changes, when $D \in [1,2]$ and one variable is used. 1854 charts were drawn and analysed.

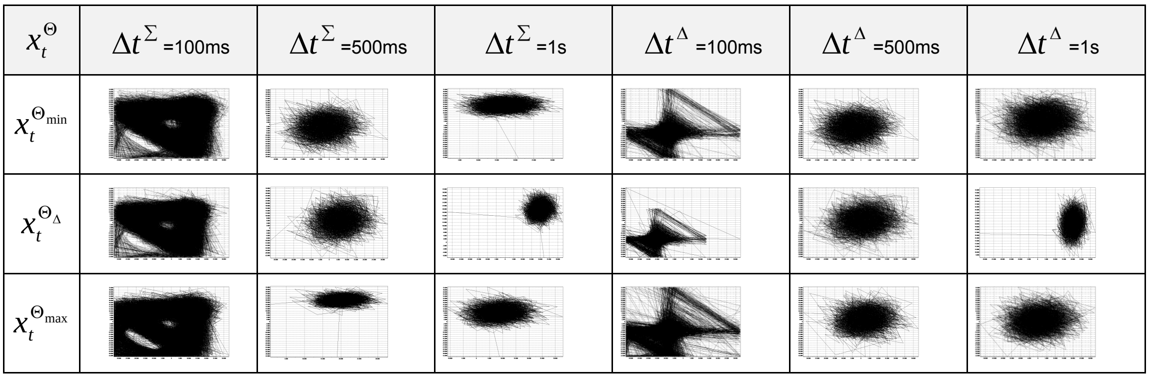


Fig. 4. Characteristic attractors obtained while analysing aggregated queues

After carrying out analysis, it was estimated that the attractors obtained belong to the group of the strange attractors. Figure 4 demonstrates characteristic attractors, when data traffic in computer network is minimal, medium and the highest. Fig.ic spheres of all the attractors are close and attributed with symmetry. This allows us drawing a conclusion that the time queues analysed are attributed with self-similarity.

4. Estimation of phasic space and correlative measures

In order to test whether LitNet node's *M* computer network data traffic queues really characterise a determined chaos for every queue, the fractal and correlative measures were analytically assessed. Grassberger and Procaccia (1983) suggested assessing the fractal measure according to the correlative measure, by using the correlative integral which defines probability that two points of an attractor are in a distance *R* one from each other.

Let us state x_t^{Θ} is an aggregated queue *X* of the analysed network traffic, which is obtained by smoothing according to averages – x_t^{Δ} and summing – x_t^{Σ} in equal time intervals Δt_n, where $i \in [1, n]$. It is stated that process *X* is attributed with long-range dependence, if the equation is satisfied:

$r(k) \approx k^{-\beta} \cdot L_i(k)$, when $k \to \infty, 0 < \beta < 1$, o $\lim\limits_{t\to\infty} \dfrac{L_i(tx)}{L_i(t)} = 1, x > 0$, where $L_i(t)$ is a constant or a function long-range changing in infinity, e.g. a logarithmic function. *r(k)* is an auto-correlative function calculated according to the classical formula:

$$r(k) = \frac{1}{N-\tau} \cdot \frac{\sum_i (X_i - \overline{X})(X_{i+k} - \overline{X})}{\sigma^2(X)},$$

$\overline{X}$ – average of a queue *X*, $\sigma^2(X)$ – dispersion, when $k \in Z_+ = \{1,2,...\}$. Such process is described by an auto-correlative hyperbolically decreasing function.

It is stated that the process *X* is attributed with short-range dependence if the aggregated process x_t^{Θ} and non-aggregated *X* are described by a short-range exponent formula:

$r(k) \approx \rho^k, k \to \infty, 0 < \rho < 1$ (Gebali, 2015, Feder, 2013, Kaklauskas, Sakalauskas, 2008, Петров, 2003, Olsen, 2000).

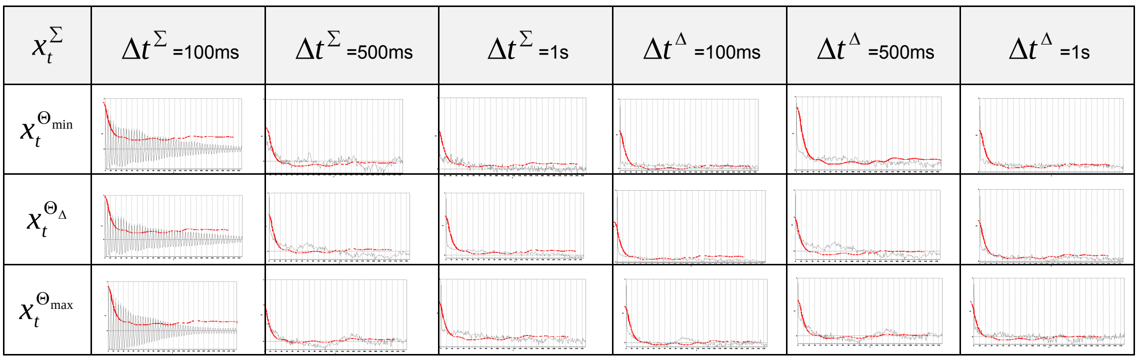


Fig. 5. Characteristic charts of the auto-correlation function

Figure 5 demonstrates charts where Lawrence's classical correlative curves of determined chaos are drawn in red using dotted line. From the charts, one can see that when time interval increases, the function of auto-correlation hyperbolically decreases; i.e. time queues describing transfer of computer network packets are attributed with long-range dependence. When time interval Δt decreases, dependence in aggregated network traffic queues is more expressed. This feature is attributed to approximately 85% out of 1854 analysed charts.

For the time queues analysed, phasic space P_{Θ} and correlation measure C_{Θ} were calculated according to network load: the lowest – $x_t^{\Sigma_{\min}}$, $x_t^{\Delta_{\min}}$, average – $x_t^{\Sigma_{\Delta}}$, $x_t^{\Delta_{\Delta}}$ and the highest – $x_t^{\Sigma_{\max}}$, $x_t^{\Delta_{\max}}$. Research proves that if time queue's phasic space measure is defined by finite numbers, it shows that the analysed process defines determined chaos and is analytically described by while using at least P_{Θ} amount of dynamic variables.

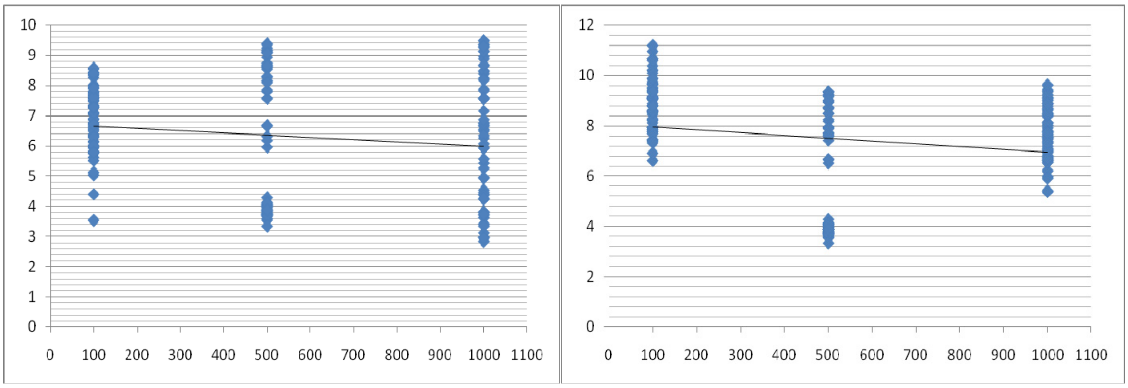


Fig. 6. Scattering of P_{Θ} measures: x_t^{Σ} – on the left, x_t^{Δ} – on the right

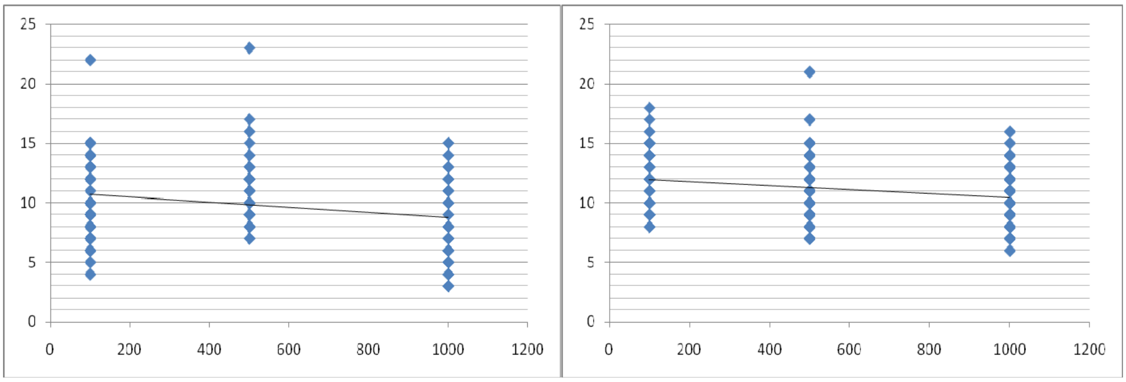


Fig. 7. Scattering of C_{Θ} measures: x_t^{Σ} – on the left, x_t^{Δ} – on the right

As it is seen in Figure 6, phasic space measures x_t^{Σ} for the queues are approximately 6.712, and x_t^{Δ} – 8.055, i.e. finite; and Figure 7 demonstrates that correlation measures x_t^{Σ} – 10.91, x_t^{Δ} – 8.055. The calculated numerical expressions are estimated analytically as well, i.e. for every queue group, according to the time interval, the average, median, standard deviation and dependent interval with reliability of 95% were calculated (see Table 1).

Table 1

Reliability estimations of phasic space and correlation measure

	$\overline{P_{\Theta}}$	$\overline{C_{\Theta}}$	$Md_{P_{\Theta}}$	$Md_{C_{\Theta}}$	$\sigma_{P_{\Theta}}$	$\sigma_{C_{\Theta}}$	$[\overline{P_{\Theta}}-1.96\sigma_{P_{\Theta}}, \overline{P_{\Theta}}+1.96\sigma_{P_{\Theta}}]$	$[\overline{C_{\Theta}}-1.96\sigma_{C_{\Theta}}, \overline{C_{\Theta}}+1.96\sigma_{C_{\Theta}}]$
$x_{t=100}^{\Sigma_{\min}}$	7.19	11.35	7.6	10	1.23	3.92	[7.58; 7.61]	[9.94; 10.06]
$x_{t=500}^{\Sigma_{\min}}$	5.32	10.59	3.97	10	2.17	3.55	[3.93; 4.00]	[9.96; 10.04]

	$\overline{P_\Theta}$	$\overline{C_\Theta}$	Md_{P_Θ}	Md_{C_Θ}	σ_{P_Θ}	σ_{C_Θ}	$[\overline{P_\Theta}-1.96\sigma_{P_\Theta},\overline{P_\Theta}+1.96\sigma_{P_\Theta}]$	$[\overline{C_\Theta}-1.96\sigma_{C_\Theta},\overline{C_\Theta}+1.96\sigma_{C_\Theta}]$
$x_{t=1}^{\Sigma_{\min}}$	5.13	6.65	4.52	6	2.04	3.04	[4.49; 4.56]	[5.95; 6.05]
$x_{t=100}^{\Delta_{\min}}$	8.62	12.47	8.61	14	1.02	2.58	[8.60; 8.63]	[13.96; 14.04]
$x_{t=500}^{\Delta_{\min}}$	5.23	10.35	3.95	10	1.99	2.29	[3.91; 3.98]	[9.97; 10.03]
$x_{t=1}^{\Delta_{\min}}$	7.28	10.41	7.31	11	0.98	2.35	[7.29; 7.32]	[10.96; 11.04]
$x_{t=100}^{\Sigma_{\Delta}}$	6.82	9.30	6.89	9	0.86	2.54	[6.88; 6.90]	[8.97; 9.03]
$x_{t=500}^{\Sigma_{\Delta}}$	6.28	11.04	6.65	10	2.30	2.98	[6.62; 6.67]	[9.96; 10.04]
$x_{t=1}^{\Sigma_{\Delta}}$	6.62	9.00	6.86	8	2.29	3.51	[6.84; 6.89]	[7.96; 8.04]
$x_{t=100}^{\Delta_{\Delta}}$	8.96	11.78	8.89	12	1.04	2.36	[8.88; 8.90]	[11.97; 12.03]
$x_{t=500}^{\Delta_{\Delta}}$	6.36	11.07	7.43	10	2.35	2.96	[7.40; 7.45]	[9.96; 10.04]
$x_{t=1}^{\Delta_{\Delta}}$	7.84	10.52	7.77	10	1.00	1.97	[7.76; 7.78]	[9.98; 10.02]
$x_{t=100}^{\Sigma_{\max}}$	6.80	10.07	6.88	10	1.24	2.84	[6.86; 6.90]	[9.95; 10.05]
$x_{t=500}^{\Sigma_{\max}}$	5.65	11.07	4.01	11	2.51	2.87	[3.97; 4.05]	[10.95; 11.05]
$x_{t=1}^{\Sigma_{\max}}$	6.69	8.67	6.64	8	1.81	2.99	[6.61; 6.67]	[7.95; 8.05]
$x_{t=100}^{\Delta_{\max}}$	8.90	12.53	7.80	12	1.59	2.61	[9.03; 9.06]	[11.96; 12.04]
$x_{t=500}^{\Delta_{\max}}$	5.72	11.07	3.96	11	2.51	2.71	[3.92; 4.00]	[10.96; 11.04]
$x_{t=1}^{\Delta_{\max}}$	7.78	11.00	8.02	11	1.09	2.80	[8.01; 8.04]	[10.95; 11.05]
$x_{t}^{\Sigma_{\min}}$	5.88	9.53	6.31	10	2.05	3.78	[6.30; 6.33]	[9.97; 10.03]
$x_{t}^{\Delta_{\min}}$	7.04	11.08	7.53	11	1.98	2.56	[7.51; 7.55]	[10.98; 11.02]
$x_{t}^{\Sigma_{\Delta}}$	6.57	9.78	6.86	10	1.93	3.13	[6.85; 6.88]	[9.98; 10.02]
$x_{t}^{\Delta_{\Delta}}$	7.72	11.12	8.12	11	1.90	2.49	[8.11; 8.13]	[10.98; 11.02]
$x_{t}^{\Sigma_{\max}}$	6.38	9.93	6.53	10	1.95	3.00	[6.51; 6.55]	[9.97; 10.03]
$x_{t}^{\Delta_{\max}}$	7.47	11.53	8.02	12	2.14	2.74	[8.00; 8.04]	[11.97; 12.03]

According to Figures 6, 7 and Table 1, every time queue has an obtained finite number which characterises the phasic space measure and correlation measure. Dependent intervals show that the phasic space measure is finite and changes from 3 to 10, depending on time interval and aggregation method. The correlation measure changes from 3 to 18, i.e. correlation integral is finite. According to results of this research, we can state that aggregated time queues describe determined chaos with finite amount of dynamic variables.

5. Estimation of Hurst statistics of computer network traffic time queues

As time queues formed of lengths of data frameworks transferred via the computer network do not satisfy the normal distribution, this section investigates their Hurst statistics. Hurst coefficient characterises whether the queue analysed is random, whether it has a short-range or long-range, also called Markov, dependence. If Hurst coefficient *H=0.5*, it means that sequence members are random and every subsequent its member does not depend on previous queue members; in an opposite case, we can state that previous events recorded in time queues have constant influence on further processes and this influence is the stronger the closer the event is to the past. Such queues are invariant from the viewpoint of time. Influence of the current process on future events is calculated by estimating its correlation (Gebali, 2015, Gallos, and etc. 2007, Beran, 1994):

$$C = 2^{2H-1} - 1,$$

where C – correlation measure, o H – Hurst coefficient. While evaluating self-similarity of a time queue, the value of Hurst coefficient, i.e. interval where it occurs, is very important.

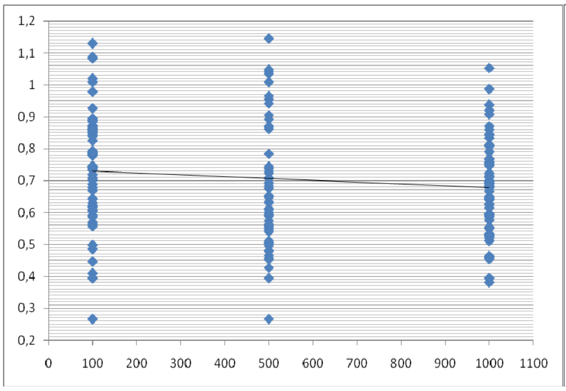
If $0 \le H < 0.5$, then the process characterised by the time queue is anti-dispersive, i.e. we can state that if increase is observed in one period, in other period decrease will definitely follow, and the probability is higher the closer H is to 0. In this case, correlation is negative and draws closer to 0.5. Such queues usually bear a feature of high changeability and are formed of frequent increases and decreases.

If $0.5 < H < 1.0$, then it is a persistent process with long-term memory, i.e. in the past, the process had a feature for increase, and it will retain it in future with the probability the higher the closer H is to 1, and correlation will draw closer to 1. Usually, such queues are called trend-resistant, while H draws closer to *0.5*, the amount of trends (noises) increases in the queue.

For formed and aggregated time queues x_t^{Θ} in the network node M, Hurst coefficient is calculated according to the formula $H = \log(R/S)/\log(n/2)$, where H – Hurst coefficient, *R/S* – *r/s* statistics acquired according to the formula:

$$R/S = \frac{R(n)}{S(n)} = \frac{Max(\sum_{i=1}^{\tau}(x_i^{\Theta} - \overline{x_t^{\Theta}})) - Min(\sum_{i=1}^{\tau}(x_i^{\Theta} - \overline{x_t^{\Theta}}))}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^{\Theta} - \overline{x_t^{\Theta}})^2}},$$

here $1 \le \tau \le n$, where n – amount of sequence members, $\overline{x_t^{\Theta}}$ – average value of the queue x_t^{Θ}, and $\sum_{i=1}^{\tau}(x_i^{\Theta} - \overline{x_t^{\Theta}})$ – the formed cumulative queue describing sum of changes throughout time τ. According to Hurst (1951), we can state that the expression is suitable for majority of natural phenomena: $M\left(\frac{R(n)}{S(n)}\right) \sim cn^H, n \to \infty$, where c – constant independent value – *a* constant (Park, Willinger, 2000). It was estimated that when amount of queue members (amount of monitoring) increases, Hurst coefficient draws closer to the value of 0.5, i.e. the memory effect decreases. Hurst coefficient is closely related with the fractal measure D which characterises local features of computer network data traffic, and Hurst coefficient describes characteristics of the whole process – memory of the process. In self-similar processes, local features are reflected in global ones and vice versa; because the time queue measure *N=1*, therefore the connection can be estimated by using the formula: *D=2-H*, where D – fractal measure, H – Hurst coefficient.



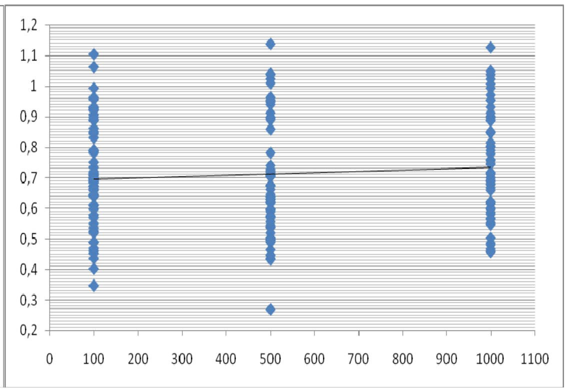


Fig. 8. Scattering of H coefficients: x_t^{Σ} – on the left, x_t^{Δ} – on the right

Scattering of calculated Hurst coefficient values is Fig.ically presented in Figure 8, dispersion of its values and analytical measures are presented in Table 2. We see that more than 80% of time queues *0.5<H<1.0*, therefore, we can state that the processes analysed are persistent with long-term memory. In order to find out whether analysed network processes retain their features when aggregated queue time intervals and intensity of data traffic change, the obtained Hurst coefficient values were compared with different queue parameters. After carrying out calculations with different time intervals $\Delta t \in$ *[100ms, 500ms, 1000ms]* and different network loads $x_t^{\Theta_{\min}}$, $x_t^{\Theta_{\Delta}}$, $x_t^{\Theta_{\max}}$, the formed queues retain their features, i.e. if a queue was attributed with a long-term memory, the same tendency remains with modified parameters (see Table 2).

Table 2

Dispersal of Hurst coefficient values

	$0.5<H<1.0$									$0 \le H < 0.5$									$1.0 \le H$								
x_t^{Σ}	81.92%									11.30%									6.78%								
	100ms			**500ms**			**1000ms**			**100ms**			**500ms**			**1000ms**			**100ms**			**500ms**			**1000ms**		
	81.36%			76.27%			88.14%			10.17%			15.25%			8.47%			8.47%			8.47%			3.39%		
	Lowest	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**
	88.24%	70.37%	93.33%	70.59%	74.07%	86.67%	76.47%	85.19%	93.33%	11.76%	14.81%	0.00%	11.76%	18.52%	13.33%	23.53%	11.11%	0.00%	0.00%	14.81%	6.67%	17.65%	7.41%	0.00%	0.00%	3.70%	6.67%
x_t^{Δ}	80.23%									12.43%									7.34%								
	100ms			**500ms**			**1000ms**			**100ms**			**500ms**			**1000ms**			**100ms**			**500ms**			**1000ms**		
	83.05%			76.27%			81.36%			13.56%			15.25%			8.47%			3.39%			8.47%			10.17%		
	Lowest	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**	**Lowest**	**Average**	**Highest**
	88.24%	81.48%	80.00%	70.59%	74.07%	86.67%	88.24%	77.78%	93.33%	11.76%	14.81%	13.33%	11.76%	18.52%	13.33%	5.88%	11.11%	0.00%	0.00%	3.70%	6.67%	17.65%	7.41%	0.00%	5.88%	14.81%	6.67%

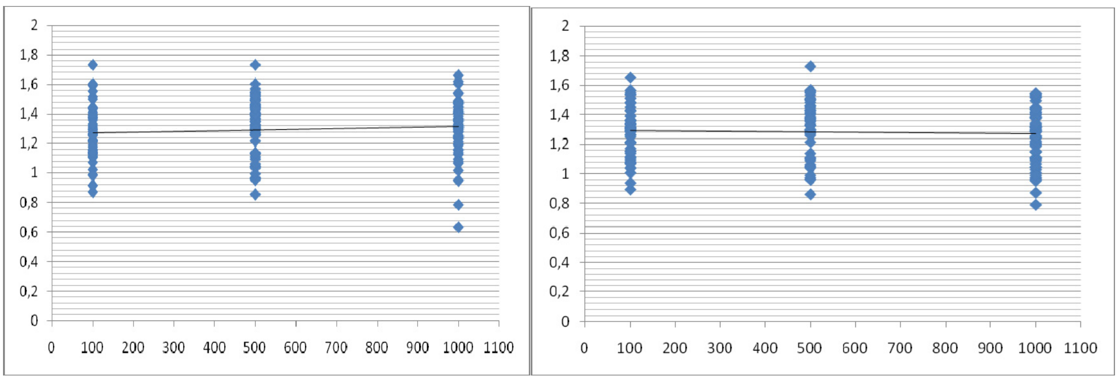


Fig. 9. D scattering: x_t^{Σ} – on the left, x_t^{Δ} – on the right

As it is seen in Figure 9, $D \approx 1.27$, it shows that the queues analysed are attributed with fractality; that is why they bear the features characteristic to fractals. Table 3 presents estimation of reliability of obtained calculation results.

Table 3

Estimations of Hurst coefficient and Fractal measure reliability

	$\overline{H_\Theta}$	$\overline{D_\Theta}$	Md_{H_Θ}	Md_{D_Θ}	σ_{H_Θ}	σ_{D_Θ}	$[\overline{H_\Theta}-1.96\sigma_{H_\Theta}, \overline{H_\Theta}+1.96\sigma_{H_\Theta}]$	$[\overline{D_\Theta}-1.96\sigma_{D_\Theta}, \overline{D_\Theta}+1.96\sigma_{D_\Theta}]$
$x_{t=100}^{\Sigma_{\min}}$	0.71	1.29	0.74	1.26	0.18	0,18	[0.74; 0.74]	[1.25; 1.26]
$x_{t=500}^{\Sigma_{\min}}$	0.69	1.31	0.63	1.37	0.24	0.24	[0.63; 0.64]	[1.36; 1.67]
$x_{t=1}^{\Sigma_{\min}}$	0.61	1.42	0.64	1.41	0.14	0.14	[0.64; 0.644]	[1.41; 1.42]
$x_{t=100}^{\Delta_{\min}}$	0,68	1,82	0,70	1,29	0,16	2,21	[0,69; 0,70]	[1.26; 1.33]
$x_{t=500}^{\Delta_{\min}}$	0.69	1.31	0.64	1.36	0.23	0.23	[0.63; 0.64]	[1.360; 1.367]
$x_{t=1}^{\Delta_{\min}}$	0.74	1.26	0.70	1.30	0.18	0.18	[0.70; 0.704]	[1.295; 1.301]
$x_{t=100}^{\Sigma_{\Delta}}$	0.76	1.70	0.79	1.21	0.19	2.43	[0.786; 0.79]	[1.18; 1.24]
$x_{t=500}^{\Sigma_{\Delta}}$	0.67	1.33	0.68	1.33	0.18	0.17	[0.68; 0.69]	[1.32; 1.33]
$x_{t=1}^{\Sigma_{\Delta}}$	0.73	1.25	0.69	1.30	0.18	0.19	[0.688; 0.692]	[1.302; 1.306]
$x_{t=100}^{\Delta_{\Delta}}$	0.71	1.28	0.68	1.31	0.18	0.18	[0.68; 0.69]	[1.31; 11.32]
$x_{t=500}^{\Delta_{\Delta}}$	0.67	2.27	0.67	1.39	0.17	3.42	[0.67; 0.68]	[1.35; 1.43]
$x_{t=1}^{\Delta_{\Delta}}$	0.73	1.27	0.69	0.20	0.20	0.20	[0.69; 0.693]	[1.306; 1.311]
$x_{t=100}^{\Sigma_{\max}}$	0.77	1.23	0.78	1.22	0.14	0.14	[0.776; 0.781]	[1.218; 1.223]
$x_{t=500}^{\Sigma_{\max}}$	0.69	1.31	0.68	1.32	0.16	0.16	[0.67; 0.68]	[1.321; 1.326]
$x_{t=1}^{\Sigma_{\max}}$	0.72	1.28	0.68	1.31	0.16	0.16	[0.681; 0.686]	[1.313;1.318]
$x_{t=100}^{\Delta_{\max}}$	0.74	1.26	0.75	1.25	0.17	0.17	[0.746; 0.752]	[1.247; 1.253]
$x_{t=500}^{\Delta_{\max}}$	0.69	1.31	0.68	1.32	0.16	0.16	[0.67; 0.68]	[1.321; 1.326]
$x_{t=1}^{\Delta_{\max}}$	0.79	1.21	0.76	1.24	0.15	0.15	[0.755; 0.760]	[1.239; 1.244]
$x_{t}^{\Sigma_{\min}}$	0.67	1.34	0.65	1.35	0.19	0.19	[0.649; 0.653]	[1.347; 1.350]
$x_{t}^{\Delta_{\min}}$	0.70	1.47	0.66	1.31	0.19	1.29	[0.660; 0.664]	[1.302; 1.324]
$x_{t}^{\Sigma_{\Delta}}$	0.72	1.74	0.70	1.30	0.18	2.43	[0.694; 0.696]	[1.287; 1.321]
$x_{t}^{\Delta_{\Delta}}$	0.71	1.29	0.68	1.32	0.18	0.18	[0.676; 0.679]	[1.314; 1.317]
$x_{t}^{\Sigma_{\max}}$	0.72	1.28	0.71	1.29	0.16	0.16	[0.706; 0.709]	[1.290; 1.293]
$x_{t}^{\Delta_{\max}}$	0.74	1.26	0.71	1.29	0.16	0.16	[0.709; 0.712]	[1.287; 1.290]
x_{t}^{Σ}	0.71	1.51	0.69	1.31	0.18	1.66	[0.693; 0.693]	[1.301; 1.317]
x_{t}^{Δ}	0.71	1.33	0.69	1.30	0.18	0.71	[0.690; 0.691]	[1.300; 1.307]

Hurst coefficient changes from 0.61 to 0.79, thus, the process of transferred via computer network data that is described by aggregated queues is a persistent process with long-term memory. Fractal measure *D* changes from 1.23 to 1.82 and dependent intervals are very narrow, and this proves that calculated values are reliable, and fractional expression of the fractal measure suggests that the queues bear features of fractals. This proves the conclusions drawn after analysis of queue attractors. After generalising research data of this section, we can state that the investigated time queues characterise persistent processes with long-term memory.

6. Fig.ic estimation of Hurst coefficient of network load time lines

Hurst coefficient can be estimated Fig.ically as well, while using the formula: $\log\left(M\left[\frac{R(n)}{S(n)}\right]\right) = H(\log(n) + \log(c)), n \to \infty$. In *X*-axis, values *log(n)* are indicated, and, in *Y*-axis – $\log\left(M\left[\frac{R(n)}{S(n)}\right]\right)$. While analysing the charts obtained, it was noticed that when Hurst coefficient draws closer to 1, noises decrease in the chart, i.e. the curve becomes continuous. In 1988, Feder proved by one's research that when *H>0.7* Hurst coefficient has a tendency for increment, and when *H<0.3* – a tendency for decrement (Feder, 2013). It was estimated that when Hurst coefficient increases, the curve that indicates the time queue becomes continuous. This theory is proved in short queues where regression is impossible. Dependence of Hurst coefficient from time (amount of members of the aggregated queues) is presented in Figure 10.

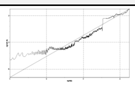
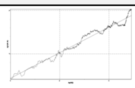
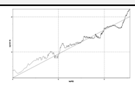
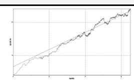
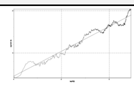
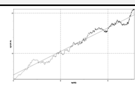
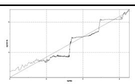
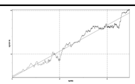
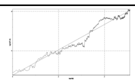
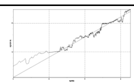
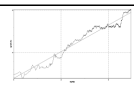
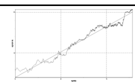
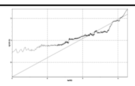
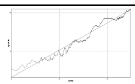
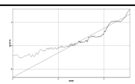
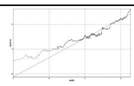
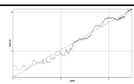
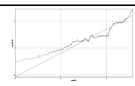
x_t^{Σ}	Δt^{Σ} =100ms	Δt^{Σ} =500ms	Δt^{Σ} =1s	Δt^{Δ} =100ms	Δt^{Δ} =500ms	Δt^{Δ} =1s
$x_t^{\Theta_{\min}}$	 H=0.7870	 H=0.5891	 H=0.6794	 H=0.5190	 H=0.5888	 H=0.5820
$x_t^{\Theta_{\Delta}}$	 H=0.6699	 H=0.5993	 H=0.7093	 H=0.5989	 H=0.6306	 H=0.6232
$x_t^{\Theta_{\max}}$	 H=0.8736	 H=0.7079	 H=1.0524	 H=0.7819	 H=0.7059	 H=0.8885

Fig. 10. Characteristic Hurst exponent charts

While analysing Fig.ical expressions of Hurst coefficient, it was estimated that absolute majority of the charts becomes continuous in relation with increment of the time interval; it is shown by research described by Freder and proves once more that aggregated time queues describe a dynamic process sith long-term memory.

7. Conclusions

1. fter carrying out research of smoothed and summed up network time queues, it was estimated that their attractors belong to the group of the strange attractors, thus, the time queues investigated are attributed with self-similarity.
2. Analysis of auto-correlation function proved that aggregated time queues are attributed with long-range dependence.
3. Reliability measures for calculation results (standard deviation, dependent interval with 95% reliability) show that the phasic space measure is finite and, while aggregating time queues, by applying both summing up and smoothing of averages, are in the interval from 6.5 to 8, analogically, the correlation measure is in the interval from 10 to 11, thus, these queues describe determined chaos.
4. Measures of Hurst coefficient reliability proved that the aggregated queues describe a persistent process with long-term memory. It was proved by analysis of Hurst coefficient charts.
5. The data traffic of Šiauliai University LitNet network node is attributed with self-similarity with long-term memory.

References

1. Bárány, B. (2009). On the Hausdorff dimension of a family of self-similar sets with complicated overlaps. *Fundamenta Mathematicae*, 206, 49.
2. Beran, J. (1994). *Statistics for long-memory processes*, (Vol. 61). CRC Press.
3. Бельков, Д. В., Едемская, Е. Н., & Незамова, Л. В. (2011). Статистический анализ сетевого трафика. *Наукові праці Донецького національного технічного університету. Сер.: Інформатика, кібернетика та обчислювальна техніка*, (13), 66-75.
4. Богутский, В. Б., & Незамова, Л. В. (2010). *Использование программы fractan 4.4 для анализа сетевого трафика*.
5. Cenedese, A., Michielan, M., Tramarin, F., & Vitturi, S. (2015). *An Energy Efficient Ethernet Strategy Based on Traffic Prediction and Shaping*. arXiv preprint arXiv:1503.02843.

6. Domańska, J., Domański, A., & Czachórski, T. (2015). Estimating the Intensity of Long-Range Dependence in Real and Synthetic Traffic Traces. In *Computer Networks*, 11-22. Springer International Publishing.

7. Erramilli, A., Narayan, O., & Willinger, W. (1996). Experimental queueing analysis with long-range dependent packet traffic. *IEEE/ACM Transactions on Networking (TON),* 4(2), 209-223.

8. Feder, J. (2013). Fractals. Springer Science & Business Media.

9. Gallos, L. K., Song, C., & Makse, H. A. (2007). A review of fractality and self-similarity in complex networks. *Physica A: Statistical Mechanics and its Applications*, 386(2), 686-691.

10. Gebali, F. (2015). *Analysis of Computer Networks*. Springer.

11. Grassberger, P., & Procaccia, I. (1983). Characterization of strange attractors. *Physical review letters,* 50(5), 346.

12. He, G., Gao, Y., Hou, J. C., & Park, K. (2004). A case for exploiting self-similarity of network traffic in TCP congestion control. *Computer Networks,* 45(6), 743-766.

13. Kaj I. (2002). *Stochastic Modeling in Broadband Communications Systems*. SIAM, Philadelphia, USA.

14. Kaklauskas, L., & Sakalauskas, L. (2008). On network traffic statistical analysis. *Liet. mat. rink. LMD darbai,* 48(49), 314-319.

15. Kaklauskas, L., & Sakalauskas, L. (2011). Study of the Impact of Self-Similarity on the Network Node Traffic. *Elektronika ir Elektrotechnika,* 111(5), 27-32.

16. Leland E., Taqqu S., Willinger W., Wilson D. V. (1994). On the Self-Similar Nature of Ethernet Traffic. *IEEE/ACM Transactions on networking*, Vol. 2, No 1.

17. Leland, W. E., & Wilson, D. V. (1991, April). High time-resolution measurement and analysis of LAN traffic: Implications for LAN interconnection. In *INFOCOM'91. Proceedings. Tenth Annual Joint Conference of the IEEE Computer and Communications Societies. Networking in the 90s*., IEEE, 1360-1366.

18. Liu, G., Li, Y., Guo, J., & Li, Z. (2015). Maximum transport capacity of a network. *Physica A: Statistical Mechanics and its Applications,* 432, 315-320.

19. Miesowicz, K., Staszewski, W. J., & Korbiel, T. (2015). Analysis of Barkhausen Noise Using Wavelet-Based Fractal Signal Processing for Fatigue Crack Detection. *International Journal of Fatigue*.

20. Olsen L. (2000). Review of Integral, probability, and fractal measures, by G Edgar. New York, Bull. *Amer. Math. Soc.*, 37 (4), 481-498.

21. Park, K., & Willinger, W. (Eds.). (2000). *Self-similar network traffic and performance evaluation*, 94-95. New York: Wiley.

22. Петров, В. В. (2003). *То, что вы хотели знать о самоподобном телетрафике, но стеснялись спросить*. М.: МЭИ, ИРЭ.

23. Петров, В. В., & Богатырев, Е. А. (2003). Статистический анализ сетевого трафика. In *Радиоэлектроника, электротехника и энергетика: Тез. докл. Десятой Междунар. научно–техн. конференции студентов и аспирантов,* (Vol. 1).

24. Riddle, L. (2014). *Koch Snowflake*. [interactive]. [Accessed on 2015-09-11]. Accessed via the Internet: <http://ecademy.agnesscott.edu/~lriddle/ifs/ksnow/ksnow.htm>.

25. Ruelle, D., Takens, F. (1994). "On the nature of turbulence". *Communications of Mathematical Physics,* 20: 167-192.

26. Sakalauskas, L., & Kaklauskas, L. (2010). Tinklo apkrovos savastingumo tyrimas realiu laiku. *Informacijos mokslai*, (53), 100-105.

27. Welte, H., Ayuso, P. N. (2015) *Userspace logging daemon*. [interacttive]. [Accessed on 2015-09-10]. Accessed via the Internet: http://www.netfilter.org.

28. Welte, H. (2000). The netfilter framework in Linux 2.4. In *Proceedings of Linux Kongress*.

29. Василенко, С. Л. (2004). *Случайность и" золотая" пропорция в системе «хаос–порядок»*. Академия Тринитаризма.–М.: Эл, (77-6567).

Received: 15 October 2015
Accepted: 29 February 2016