VILNIUS UNIVERSITY

Liudvikas Kaklauskas

STUDY AND APPLICATION OF METHODS OF FRACTAL PROCESSES MONITORING IN COMPUTER NETWORKS

Summary of Doctoral Dissertation

Physical Sciences, Informatics (09 P)

Vilnius, 2012

Doctoral dissertation was prepared at Institute of Mathematics and Informatics of Vilnius University in 2008-2012

Scientific Supervisors

Prof. Dr. Habil. Leonidas Sakalauskas (Vilnius University, Physical Sciences, Informatics – 09 P)

The dissertation will be defended at the Council of Scientific Field of Informatics at the Vilnius University:

Chairman

Prof. Dr. Habil. Kazys Kazlauskas (Vilnius University, Physical Sciences, Informatics – 09 P)

Members:

Prof. Dr. Romas Baronas (Vilnius University, Physical Sciences, Informatics – 09 P),
Prof. Dr. Habil. Genadijus Kulvietis (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering – 07 T),
Prof. Dr. Liudas Mažeika (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07 T),
Assoc. Prof. Dr. Rimantas Pupeikis (Vilnius University, Physical Sciences, Informatics – 09 P)

Opponents:

Assoc. Prof. Dr. Igoris Belovas (Vilnius University, Physical sciences, Informatics – 09 P),
Prof. Dr. Dalius Navakauskas (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering – 07 T)

The dissertation will be defended at the public meeting of the Council of Informatics Science in the auditorium number 203 at the Institute of Mathematics and Informatics of Vilnius University, at 1 p.m. on 27 June 2012.
Adress: Akademijos st. 4, LT-08663 Vilnius, Lithuania.

The summary of the doctoral dissertation was distributed on 25th of May 2012.
A copy of the doctoral dissertation is available for review at the Library of the Vilnius University.

VILNIAUS UNIVERSITETAS

Liudvikas Kaklauskas

FRAKTALINIŲ PROCESŲ KOMPIUTERIŲ TINKLUOSE STEBĖSENOS IR VALDYMO METODŲ TYRIMAS

Daktaro disertacijos santrauka

Fiziniai mokslai, informatika (09 P)

Vilnius, 2012

Disertacija rengta 2008–2012 metais Vilniaus universiteto Matematikos ir informatikos institute

Mokslinis vadovas

prof. habil. dr. Leonidas Sakalauskas (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P).

Disertacija ginama Vilniaus universiteto informatikos krypties taryboje:

Pirmininkas

prof. habil. dr. Kazys Kazlauskas (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P)

Nariai:

prof. dr. Romas Baronas (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P),
prof. habil. dr. Genadijus Kulvietis (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – 07 T),
prof. dr. Liudas Mažeika (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07 T),
doc. dr. Rimantas Pupeikis (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P)

Oponentai:

doc. dr. Igoris Belovas (Vilniaus universitetas, fiziniai mokslai, informatika – 09 P).
prof. dr. Dalius Navakauskas (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – 07 T)

Disertacija bus ginama viešame Vilniaus universiteto Informatikos mokslo krypties tarybos posėdyje 2012 m. birželio 27 d. 13 val. Vilniaus universiteto Matematikos ir informatikos instituto 203 auditorijoje.

Adresas: Akademijos g. 4, LT-08663 Vilnius, Lietuva.

Disertacijos santraukos išsiųstos 2012 m. gegužės 25 d.

Disertaciją galima peržiūrėti Vilniaus Universiteto bibliotekoje.

1 Introduction

1.1. Research field

It is known that information of data packages transmitted via a computer network node is dynamic, hardly prognosticated and reminds a noise. Usually, it is impossible to describe this information using traditional mathematical methods. *Ethernet* network traffic research carried out in 1989 by A. Erramilli, O. Narayan and W. Willinger proved that characteristics of this traffic are attributed with fractality and self-similarity with long-range dependence (further – *LRD*). The latter feature allows prognosticating the variance of a computer network and applying the prognosis results for perfection of the quality of service (further – *QoS*) when regulating delay of network traffic packages, traffic fluctuations and package loss. For assessment of self-similarity of the network traffic, statistical methods of time lines analysis, frequency/wave feature estimates, calculated stability parameters of stable random sizes are used as well as the chaos theory is applied. In 2008, D. Surgailis, G Teyssiere and M. Vaičiulis worked out the increment ratio (further – *IR*) statistics of self-similarity analysis of the network traffic with *LRD*. In the present dissertation, *IR* statistics has been applied for self-similarity analysis of the computer network traffic for the first time.

For more thorough research of network processes, mathematical models of network traffic are used. In them, by applying non-linear analysis methods, big-volume network traffic bursts are assessed; they are described by heavy-tail stochastic models. For analysis of contemporary network traffic, *ON-OFF* network traffic models are widely used; they are based on computer network traffic features. For modelling of self-similar traffic, service theory instruments are successfully used. In the dissertation, features of the network traffic have been assessed by modelling it with flexible α–stabile distributions which follow the central limit theorem, describe heavy tails, asymmetry and are attributed with a leptokurtic feature.

By using models of self-similar network traffics, it is possible to assess the impact of network traffic features on network node (switch, router, server etc.) parameters. In such models, synthetic network traffic generated according to parameters indicated by a researcher or traffic data collected in a real computer network is used. When network traffic has no priority data packages, for queue service (node buffer) in network nodes *FIFO* and *LIFO* queue disciplines are used. In 2009, J. Nzouonta, T. Ott and C. Borcea estimated that nodes that serve high speed (approx. 30 Gb/s) and pass computer networks, 15-20 packets buffer is completely sufficient. The feature of network traffic self-similarity can be successfully used for regulation of network packages transportation traffic overloads.

When assessing features of real-time packed-based network traffic, a possibility to dynamically prognosticate the traffic variance appears. Dynamic estimators of network traffic can be applied for perfection of traffic service quality *QoS*, thus ensuring optimal service of arrived packages, minimal packages loss in a network node. It was estimated that network traffic service in a node improved by choosing the optimal queue length, queue service discipline, decreasing delay of packages and changing other node parameters in the way they would meet the dynamic estimators of network traffic features best.

1.2. Relevance of the problem

Computer networks used for distance studies usually undergo non-prognosticated overloads due to specificity of the study process. Aiming at decreasing network overloads and related losses of transmitted information, monitoring of network and its nodes conditions is a must; and, on the ground of the results obtained to prognosticate loads and overloads of network and its nodes. As research of package-based computer network traffic show, it has features of fractals and self-similarity with *LRD*. Empirical research of Šiauliai University e-studies computer network package-based traffic proved this conclusion. Such network traffic can be analysed as a fractal process attributed with the second-line statistical self-similarity characterised by a fractal measure, attractors, auto-correlation function features and a phasic space measure. Many research works on self-similar processes in computer networks show it is possible to prognosticate network variance and to apply the prognosis results for ensuring network traffic service quality.

It should be admitted that there are no extensive research assessing the impact of network node buffer capacity and queue service discipline on self-similar network traffic service quality. Assessing self-similarity of real-time network traffic, it is possible to dynamically prognosticate traffic variance; and it is possible to apply the prognosis results for perfection of network traffic service quality QoS regulating delay of packages and shortening their service time in a network node, decreasing traffic fluctuations.

Scientists of Lithuania usually analyse theoretical models of network traffic self-similarity analysis; however, experimental modelling is selfdom carried out. Indicators of real-time network load self-similarity, applying α-stabile models, are extensively investigated neither in Lithuania nor abroad. It should be admitted that methods of real-time network traffic load self-similarity features still are insufficiently investigated. This problem is relevant and still insufficiently investigated in both theoretical and practical aspects. The research area of this work covers monitoring and analysis of package-based computer network traffic fractal processes, study of methods of assessment of real-time traffic self-similarity features and experimental research of traffic models regarding network traffic features and network node parameters.

1.3. The research object

The object of the dissertation research is the features of computer network packet traffic, the impact of network node features on computer network packet traffic service, methods of real-time network traffic features analysis and their application for dynamic prognostication of network traffic variances.

1.4. The aim and objectives of the research

The aim of work is to investigate fractal processes in computer networks, grounding on the results obtained to select methods suitable for real-time analysis of network traffic and to work out methods for real-time measurement of self-similarity as well as to apply it for perfection of computer networks service quality.

To achieve the aim, the following objectives have been set:

- to investigate methods of research of computer networks' fractality (self-similarity);
- to select methods for real-time analysis of network traffic self-similarity;
- to work out methods for real-time measurement of computer network traffic self-similarity and appropriate tools;

- to create a system for simulation of computer networks information traffics and to apply it to the worked out methods for testing;
- to apply worked out methods for perfection of computer networks traffics service quality QoS.

1.5. Scientific novelty

The following results have been obtained:

1. The package of analysis of network traffic features has been worked out, used for analysis, assessment and comparison of computer networks fractality and self-similarity research methods.
2. Methods of computer packet network traffic self-similarity analysis have been investigated:
   a. time lines analysis methods when gradual statistical dependences of selected sequences and m-size blocks selected from the sequence are investigated,
   b. methods for assessment of frequency/wave features based on frequency wave ripples features,
   c. estimators of time line stability parameters.
3. The package of real-time network traffic self-similarity analysis (further – SRLAP) has been worked out.
4. By using SRLAP for real-time measurement of self-similarity the robust *regression* method and *IR* statistics have been selected.
5. Methods for computer network traffic self-similarity real-time measurement have been worked out; they recurrently form a queue, calculate its statistical parameters and stored results in a temporal memory. These methods differ from others by real-time aggregation of network traffic by applying the method of smoothing moving means or the sum of transmitted data traffic during a time interval, the formed queue is stored in a temporal memory, recurrently calculates self-similarity parameters of a dynamically formed queue calculating α-stability index, *IR* estimator and *Hurst* coefficient, stores the calculations results in a temporal or/and permanent memory for further

analysis. The method has been patented at the State Patent Bureau of the Republic of Lithuania, patent No. LT20011099.

6. The system of simulation of computer network self-similar and *Poisson* traffics have been worked out and, using it, the impact of network node features on quality of network packet traffic service has been investigated.
7. The methods worked out have been applied for perfection of computer network traffic service quality QoS.

1.6. Practical significance of the work

The following practical results have been obtained in the work:

1. Methods and tools for real-time assessment of computer network data traffic self-similarity have been worked out;
2. Methods and tools for simulation of computer network self-similar traffics with heavy tails have been worked out;
3. Solutions on self-similar heavy-tail computer network traffic service quality QoS obtained after estimation of self-similarity of information traffics have been suggested and, grounded on them, the model of real-time traffic control have been projected.

1.7. Defended statements

Main dissertation statements to be defended:

1. For assessment of computer network traffic self-similarity, groups of methods based on time sequence analysis, assessment of frequency/wave processes features and estimators of queue stability parameters are used. Newly applied estimators of time line parameters for real-time assessment of network traffic self-similarity obtained by calculating the following: tail index according to the robust *regression* method and *IR* statistics estimators.
2. Fractal processes of computer network packet traffic make impact on network traffic service quality *QoS* in a network node and depend on network node parameters. According to estimators of network traffic features, when changing network node characteristics, buffer length and buffer queue service

discipline, probability of package loss, the duration of average package time in a network node buffer queue, time of full buffer queue filling and delay decrease. Such variances of node features ensure higher quality of network traffic service in network node.

3. A new method for real-time assessment of computer network traffic self-similarity allows to dynamically change features of a network node, thus ensuring better traffic service in a network node.

1.8. Approval of the results

Research results have been presented and discussed in two international conferences, one international seminar, two national conferences and three national seminars. Two articles have been published in foreign scholarly publications included into the Institute's of Scientific Information list of main journals with the citation index, one published in international conference proceedings included into the Institute's of Scientific Information list and four appeared in reviewed Lithuanian and foreign publications.

1.9. Structure of the dissertation

Chapter 1 denotes the object, aim, objectives of the dissertation. Scientific novelty and practical significance of the work are discussed. Presentations and publications on the dissertation topic are introduced.

Chapter 2 describes the mathematical model of computer networks. Technological solutions of contemporary package-based networks, technological solutions of network equipment and information traffics transmission, structure of packages. Possibilities for mathematical modelling of network components, computer network packet traffic models and models using service theory instruments are described.

Chapter 3 presents technologies for measurement of fractal traffics in computer networks. Technological solutions for network traffic measurement are discussed. Estimation of network traffic parameters according to the chaos theory, calculating Hurst statistics, obtaining estimators of stable parameters and applying *IR* statistics are dealt with.

Chapter 4 discusses means of network traffic analysis, testing of network analysis methods, formation of network traffic time lines, and the problem of time lines aggregation and generalises the results of computer network traffic time lines analysis.
Chapter 5 analyses software for simulation of fractal processes, discusses results of one-channel network node research. After applying a stochastically limited network communication model and algebra plus means, the GI/G/10//N computer network node is defined. Results of research on the impact of Poisson and self-similar network traffic features on its service in a network node are generalised.
Chapter 6 presents the method of real-time measurement of self-similarity and the model of computer network control is investigated. Here, world-wide used analogous measurement methods, their strengths and weaknesses are discussed, the presented invention – the method of real-time analysis of network traffic – is described. The open network node control model with feedback is analysed; the invention is applied in it for dynamic network node control.

2 Modelling of the computer network

A computer network is a group of computers and other devices joined by connection lines to ensure data exchange among them. A computer network consists of the following: network nodes, network customers and connection lines. Network nodes distribute data traffic transmitted via a computer network. In network nodes, routers, switches, hubs and bridges are usually used for distribution of data packages. A router is a device ensuring connectivity between two or more networks, intended for formation of big networks, description of information transmission network routes, joining networks of different kinds and purposes into one joint network. A switch is a network node device joining many connection lines coming from various network nodes or network customers. It distributes data packages in a network with regard to sender's and receiver's addresses, usually used for network sub-networks' or segments' service. A simpler device is called a hub. It multiplies a received package as many times as it has links, except the sender. Such multiplication of packages increases network load; therefore, these devices are seldom used in computer networks. A bridge is a network node device joining and converting networks of different standards and technological solutions. A network customer is a computer, printer, mobile or other device used by a computer network user using network services. It is usually joined to a network node and receives network data packages from a service server which receives a request. Connection lines are a physical connection channel used for transmission of data packages via a computer network.

2.1. Modelling of computer network information traffic

In computer networks, process of information transmission on the functional level are described on the ground of the OSI (*Open System Interconnection*) model which consists of seven layers. For collection of information on a network traffic, four lowest layers are enough: physical – provides information on physical addresses of network devices; data or connection channel – provides information on data distortion, package length and package arrival time; network – provides information on sender’s and receiver’s addresses; transport – provides information on a used service type. Computer network equipment parameters are described on the ground of set standards. *Ethernet Type II* standard packages were analysed; according to the Project 802 standard issued in February 1980 they meet the category 802.3. Also, packages of categories 802.1, 802.2, 802.8 and 802.11 were analysed.

When modelling computer network traffic, two main parameters generated by applying computer network models are available: packages length distributor and times between arrived packages distributor. The selection of a packages length distributor with regard to features of an information traffic transport packages, is not a difficult task. It is more difficult to select a proper distributor of times between arrived packages and to apply a network traffic model to it. Earlier network traffic models, such as *Poisson* with an exponent distributor of packages arrival time, were intensively until estimation of a traffic self-similarity feature. For modelling self-similar heavy-tail computer network traffic, more complex traffic models are applied. Aiming at more subtle modelling results, it is possible to regard features of network components influencing results of information traffic service. For a more thorough analysis of computer network traffic, mathematical network traffic models are applied. In 2008, C. Grimm and G.

Schlüchtermann suggested to describe computer network packet traffic by applying three models:

1. Incoming traffic model. Here incoming traffic is formed out of data packages of data units sent via computer network. For process modelling, package arrival time t_i and time intervals between two successfully arrived packages τ_i are used.
2. Service model. Here packages service in a network device is modelled, and every arrived package is recorded as a request for service. For process modelling, a discrete number of requests for services and a gradual service time for every package are used.
3. Model of processes condition. When implementing the model, two conditions of processes are regarded: 1 – discrete number of arrived packages in *k* system, 2 – gradual time of the package in the system T_i which is equal to the sum of the waiting for a request time w_i and its service time x_i.

2.2. Models of package-based computer network traffic

Poisson mathematical model is the oldest and most often used tool for modelling telephone network traffic. A combined *Poisson* model expands a standard model by performing analysis of groups of packages. Like the combined *Poisson* model, the *MAIG* (*Maximum Allowed Inter-Car Gap*) model also implements transmission of groups of packages. Classical Markov and Markov chains methods model computer network traffic when amount of its states is limited. The *MMRP* (*Markov Modulated Rate Processes*) model is based on matrix proper meaning's estimation which becomes more complex when the amount of described variables increases. Markov chains can be applied when modelling self-similar network traffic. *AR* (*Autoregressive*) network traffic models are best suitable for modelling of a network where transmitted data packages' priorities, i.e. sound transmission, view transmission, are important and analogous networks. *ARMA* (*Autoregressive Moving Average*) models, also called Box-Jenkins models, for prognostication of a outgoing traffic use net only incoming traffic but also parameters of a previously outcome traffic. The essence of autoregressive integrated moving average method *ARIMA* lies in the fact that it unites possibilities of autoregression, differentiation and moving averages methods and it is used for modelling of linear time lines. The *ON-OFF* network model is more often used when a structure of computer network traffic is analysed when severe network traffic fluctuations occur. By using the model, it is possible to describe network traffic from the physical layer to the application *OSI* layer. On the base of the *ON-OFF* model, the *IPP* (*Interrupted Poisson Process*) model has been worked out; it describes a network channel as a system with two different conditions: 1 – when it is loaded and operates, 2 – when there are no requests in the network channel and it is free and idle.

For modelling of bursting traffics, α–stable models are usually applied. They follow the generalised central limit theorem and well describe heavy tails, asymmetry. According to the definition proposed by G. Samorodnitsky and M. S. Taqqu in 1994, self-similar symmetric process described by formulas and attributed with infinite dispersion is α-stable process, when $0 < \alpha < 2$, if for every stochastic process *Y(t)* heavy tails can be described by a formula $P(|Y(t)| > x) \sim cx^{-\alpha}$, when $x \to \infty$, o $c > 0$. Assessing any stable stochastic distributor $S_{\alpha}(\beta, \sigma, \mu)$, it is recommended to assess four stability parameters:

- α – stability parameter $\alpha \in (0,2]$, also called tail parameter, defining burstiness of a process,
- β – asymmetry parameter $\beta \in [-1,1]$, defining shift of a process in regard to zero,
- σ – scale parameter, σ>0 and define a size of elements,
- μ – position parameter $\mu \in R$.

The service theory worked out in 1909 by A. K. Erlong and later developed by A. Khinchin, T. Engset, D. Kendall, A. Kolmogorov and many other scientists integrated a few concepts characteristic to connection technologies: incoming traffic intensity – λ, service traffic intensity – μ, service duration – *x*, load coefficient – ρ, channel load coefficient – *q*, amount of service channels – *m*, average duration of one request service – $\overline{X}$, average amount of request in a system – $\overline{N}$ etc. According to the service theory, when describing the modelled traffics in a computer network, specialised marks, are used, and network data packages are called applications. Traffic of events is called determined if events follow one another in strictly set moments that are known in advance. A determined traffic is regular if events follow one another in equal periods of time. The regular determined traffic is marked by a symbol ***D***. Computer network packet traffics are usually stochastic with undefined time of package arrival. Such traffics are described by stochastic distributors. The network traffic the moments of packages arrivals t_n are randomly distributed are marked by a symbol ***G*** (*General*). In a recurrent traffic ***GI*** (*General Independent*), periods between packages arrivals are independently equally distributed. A stochastic traffic of packages is called ordinary if a probability that for a short interval of time two or more packages will occur decreases faster, when the interval shortens, than the duration of this period. The traffic of packages is considered to be without the interaction if in case of any two disjoint time intervals the amount of events occurring in one of them does not depend on the amount of events in the other interval. A traffic which is stationary, ordinary and without interaction is called a simple traffic. For modelling of network traffic, the simple traffic is usually used; it is marked by **M**. The amount of events in the simple traffic occurring in a certain length interval is distributed according to the *Poisson* or exponent law. A regularly rarefied simple traffic is *Erlango* traffic and is marked by a symbol $\boldsymbol{E_r}$. The hyperexponent traffic is marked by $\boldsymbol{H_r}$ and described by the sum of several exponent functions.

In a common case, a network traffic model can be written down using Kendall's notation A/B/C/D/E/F, which describes and classify the queueing model. Here, A – standard of incoming traffic description, B – standard describing service time distributor, C – amount of service channels sometimes called the amount of servers in a modelled system and is a natural number equal to 1 or bigger that 1, *D* – amount of places in the system encompassing capacity of the system buffer and currently serviced requests and which can be infinite, E – size of applications population expressed by a natural number, if this argument is skipped then it is supposed that the amount of applications is infinite, *F* – discipline of device queue service, if this argument is skipped it is usually supposed that the *FIFO* discipline is used.

3 Technologies and methods of measurement of fractal traffics in computer networks

For assessment of fractality, information about computer network traffic package headings is needed; it can be directly recorded in a computer network node or customer's computer. For accumulation of needed information for the research, the event-calculation-based strategy was employed. Šiauliai University LitNet network node was chosen for measurement; it is characteristic with the highest load and traffic intensity (1 Gbps traffic). Measurements carried out in one microsecond exactness. The data base accumulated more than three billion records (from 04-01-2008 13:30:35 to 16-04-2008 12:00:00) which corresponded 8936965 seconds or 103 days 10 hours 29 minutes and 25 seconds. For assessment of fractality features, with regard to network traffic intensity during the week and the day, 309 queues of one hour time were selected, they had from 500,000 to 1,500,000 elements and corresponded measurement of one hour.

3.1. Assessment of network traffic parameters by applying the chaos theory

The chaos theory is based on two statements: it is impossible to define the future in detail because of measurement biases and not knowing all initial conditions; reliability of worked out prognoses rapidly decrease with time. Major instruments of this theory – attractors and fractals – are used aiming to describe dynamic systems not obeying classical theories. An attractor is a geometrical structure defining behaviour of a system in a phasic space when time approaches infinity. A phasic space is an abstract space the coordinates of which define freedom degrees of a system under investigation. When analysing by a computer a system defined in three freedom degrees, consisting of three simple differential equations with three constants and three initial conditions, E. Lorenz described the first chaotic or strange attractor. The strange attractor has certain maximum limits characterised by attractor's dimension, and a fractal is its geometrical expression. The usage of a fractal measure measuring coastline of England was described by B. Mandelbrot.

By using attractors and fractals, it is possible to describe dynamic systems nor obeying classical theories. In 1983, P. Grassberger and I. Procaccia suggested to assess a fractal measure by assessing a correlative measure, using a correlate integral which characterises probability that two points of an attractor are in a distance *R* from each other. Suppose, x_t is an aggregated line of the process *X* with gradually fixes time intervals Δt_n, where $i \in [1, n]$. It is stated that the process *X* is attributed with *LRD*, if the equation is satisfied: $r(k) \approx k^{\beta} \cdot L_i(k), k \to \infty$. *Here, r(k)* is an autocorrelate 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 line X, $\sigma^2(X)$ – dispersion, $k \in Z_+ = \{1,2,...\}$, $0 < \beta < 1$, L_i – slowly in infinity changing function, i.e. $\lim_{t\to\infty} \frac{L_i(tx)}{L_i(t)} = 1$, for all $x > 0$. An aggregated line x_t and non-aggregated process *X* are characterised by an autocorrelate function which hyperbolically decreases when extending the time interval. It is stated that the process *X* is attributed with decreasing dependence *SRD*, if x_t and *X* are defined by a rapidly decreasing exponent formula: $r(k) \approx \rho^k, k \to \infty, 0 < \rho < 1$.

3.2. Methods of self-similarity analysis in computer networks

The phenomenon of self-similarity is explained by the character of network services usage attributed with burstiness. In contemporary university studies, computer networks are widely used; they often undergo non-prognosticated overloads. For effective network control, monitoring of network nodes must be carried out in order to prognosticate network loads and overloads. According to the definition proposed by G. Samorodnitsky in 2006, a stochastic process $Y(t), t \geq 0$ is self-similar if it is possible to find such *H* that all c>0, to satisfy the equation: $(Y(ct), t \geq 0) \overset{d}{=} (c^H Y(t), t \geq 0)$. $\overset{d}{=}$ means that this equation is valid in all points of the function, except break points. Here, the exponent *H* defines significance of distribution of a stationary process *X* and is called *Hurst* coefficient. If *H=0.5*, then sequence members are random and every following its member does not depend from previous line members. When $0 \leq H < 0.5$, the process defined by a time line is antipersistent, i.e. it can be stated that if at one time period an increase is observed then the following period will necessarily by decrease. If $0.5 < H < 1.0$, then the persistent process with *LRD*, also called Markov dependence, i.e. if in past the process had a quality to increase, then in future it will retain this quality with higher probability when *H* is closer to 1.

For assessment of *Hurst* coefficient, time sequence-based *R/S*, absolute meanings, sometimes called absolute moments, dispersion, residuals dispersion methods or periodograms-based on estimators of frequency/waves features, decreasing, *Abby-Veich* methods are usually used. According to G. Samorodnitsky, a self-similar symmetric process described by formulas and is attributed with infinite dispersion is α-stable process. When assessing α-stable process $S_\alpha(\beta, \sigma, \mu)$, four stability parameters must be calculated and assessed. The dissertation analysed *Fama-Roll*, *McCulloch*, regression and moments methods and applied them for assessment of network traffic self-similarity features.

For the first time, for assessment of computer network self-similarity, the IR (*Incerment ratio*) statistics proposed by D. Surgailis, G. Teyssiere and M. Vaičiulis in 2008 was employed. It is based on calculation of partial sums of the process $S_n = \sum_{t=1}^{n} X_t$ line applying second-line intervals or differences. Statistics is calculated by the formula:

$$IR = \frac{1}{N-3m} \sum_{k=0}^{N-3m-1} \frac{\left| \sum_{t=k+1}^{k+m} (X_{t+m} - X_t) + \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t) \right|}{\left| \sum_{t=k+1}^{k+m} (X_{t+m} - X_t) \right| + \left| \sum_{t=k+m+1}^{k+2m} (X_{t+m} - X_t) \right|}.$$

Here, $X_1, X_2, ..., X_N$ is a time line with length *N*, $m = 1,2,...$ – parameter of pass. According to the definition, *IR* statistics is limited and changes $0 \le IR \le 1$. The scale of estimators does not change when X_t is changed by a linear combination $aX_t + b$, where $a \ne 0$, o b is a constant. Empirical simulations of the process carried out by the authors showed that *IR* statistics was quite insensitive to noises, local non-stationarities and heavy tails. Stability of statistics is relevant when analysing computer network traffic because it is attributed with high burstiness.

4 Analysis of fractality in computer network traffic

Formed computer network package traffic time lines remind of a determined chaos which depends on the following: channel pass, efficiency of network nodes, data transmission protocol, error correction methods, node protection rules and other objective and subjective factors. To achieve better calculation results, it is recommended to aggregate such lines, i.e. restructure them in the way they would have stationary variances. In 2007 O. I. Sheluhin, S. M. Smolskiy and A. V. Osin proved that any discrete stochastic process $\{X_k, k \in Z\}$ is strictly self-similar in a broader sense (exact second-line self-similarity) with a self-similarity parameter *H*, when $0.5 < H < 1$, if $R(k) = \frac{\sigma^2}{2}\left[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right]$, for any $k \geq 1$. In case of strict self-similarity $R(k)$ it exactly corresponds to self-similarity of a non-aggregated line. Aiming at more exact estimators of package-based network traffic self-similarity, for data aggregation two aggregation strategies were used: smoothing of moving averages $x_k^{\Delta} = \frac{\sum_{\tau_i \in [t_k, t_{k-1}]} x_i}{\Delta t}$ and calculation of a sum of transmitted data traffic in a time interval Δt $x_k^{\Sigma} = \sum_{\tau_i \in [t_k, t_{k-1}]} x_i$, where $t_k = k \cdot \Delta t + \tau_1$. Aggregated 1854 lines were divided into 6 groups: $\{X_k^{\min \Delta}\}$, $\{X_k^{\min \Sigma}\}$, $\{X_k^{avg\Delta}\}$, $\{X_k^{avg\Sigma}\}$, $\{X_k^{\max \Delta}\}$, $\{X_k^{\max \Sigma}\}$. Aiming at assessment of traffic fractality, every time line was aggregated and assessed by applying three different aggregation intervals $\Delta t \in [100ms, 500ms, 1000ms]$. Thus, every group was divided into three sub-groups. 5,562 lines were formed for assessment of self-similarity.

When forming an aggregated line real-time, a line formed for the first time is updated recurrently, i.e. a new line member is added instead of a member that stayed in the line for the longest time. The new line member is usually included in the end of the line and the old one is deleted from its beginning. Traditionally, after such substitution, a newly formed line must be re-written; because of that the system waists quite much time. For making the process of formation and aggregation faster, I have created the algorithm of re-indexation of an aggregated line which changes indices of line members only, this way saving line formation time.

For analysis of formed time lines, two freeware programs have been used: V. Sychyov *Fractan* 4.4 of 2003 and T. Karagiannis *Selfis* 0.1b of 2002, and for assessment of parameters of α-stable distributors the *SSE* (*Self-similarity estimator*) library of program modules was worked out. In SSE, four self-similarity assessment methods have been applied: *Fama-Roll*, *McCulloch*, *Regression*, *Moments*. Suitability of the used methods for analysis was tested by means of computer-assisted simulation, by applying synthetic traffic simulation formulas suggested by A. Janicki and A. Weron in 2000.

Research of simulated time lines showed that the variance of the asymmetry coefficient β for time lines attributed with self-similarity had minimum impact on *Hurst* coefficient and stability parameter α meaning. Comparative analysis of simulation results obtained by *Fractan*, *Selfis* and *SSE* showed self-similarity estimators of a synthetic traffic were calculated most exactly when applying the *regression* method. Results of research of real network traffic aggregated lines showed that the process described by time lines was a persistent process with *LRD*, when *Hurst* coefficient meaning changed from 0.61 to 0.79. Moreover, the investigated computer network traffic was attributed with fractality.

5 Simulation of fractal processes

It should be admitted that there are no extensive research works assessing the impact of network node buffer capacity and queue service discipline on network traffic service with regard to features of that traffic. Such assessment would help select the most suitable network node parameters in order network customers were successfully services even when high overloads and self-similar heavy-tail traffic occur. This would ensure stable work of computer network what is especially relevant in educational institutions when distance studies become more popular in the study process.

In 1998, W. Willinger, V. Paxson and M. S. Taqqu showed that when simulating fractal processes a stochastic system model changing dependence of analytical system condition and results into analogous stochastic characteristics of interrelations of network traffic features and network nodes parameters is described. For simulation of fractal processes mathematical models of service theory are widely used. For description of network traffic calculable network models *Network Calculus* are used. For description of processes taking place in the network *Min-Plus* and *Max-Plus* algebras are employed. For simulation modelling of fractal processes, the system for network traffic modelling *MulNodSimSys* (*Multichannel Node Simulation System*, system was created by L. Kaklauskas) has been designed and created; it performs the following functions:

- generates *Poisson* or self-similar with *LRD* and heavy tails network traffic,
- features of network node are described by using dynamic library which stores a set of node features variances,
- generated network traffic in a prepared network model is serviced for several times, dynamically changing network node features according to records of variances formed in a dynamic library,

- results of network traffic service and variances of network objects features are kept in data storages for further analysis of results.

Using *MulNodSimSys*, one-channel service system with a recurrent arrival traffic, general character service traffic, limits for a number of packages in a buffer and service waiting time GI/G/1//N was investigated. For simulation of the Poisson process formulas were applied: $\tau_i = -\frac{\ln(\xi_i)}{\lambda}$, $x_i = -\frac{\ln(\zeta_i)}{\mu}$, ξ_i, ζ_i are independent evenly distributed in a unit interval random sizes. For simulation of a self-similar process formulas were used: $\tau_i = \frac{\sqrt[\alpha]{\ln(\xi_i)/\ln(\omega_i)}}{\lambda}$, $x_i = \frac{\sqrt[\beta]{\ln(\zeta_i)/\ln(\upsilon_i)}}{\mu}$, ξ_i, ζ_i, ω_i, υ_i are independent evenly distributed in a unit interval random sizes. In *MulNodSimSys* traffic sub-system, time lines with 1,000,000 requests having incoming/services traffic have been generated changing and variously combining the incoming/serviced traffic type, incoming traffic intensity, serviced traffic intensity, parameter of self-similar line stability. In a servicing network node, queue length and queue service discipline were changed. 6,776 generated incoming traffic time lines and 74,536 results of their service in network traffic were analysed. It was estimated that when network node service traffic is Poisson traffic, then when intensity of incoming traffic increased the service probability decreased and probability of package loss increased. Queue service discipline *LIFO*, when incoming traffic was Poisson traffic and service was attributed with self-similarity, ensures higher service quality.

Using *MulNodSimSys* a multi-channel service system (ten channels), recurrent incoming traffic, general character service traffic, limitations for the amount of packages in a buffer and service waiting time in GI/G/m//N network node have been investigated. This network model is based on the stochastic network model (*Stochastic Network Calculus*) which allows analysing *end-to-end* computer network *QoS* systems with stochastic incoming network traffic and stochastic network nodes. When working out the model, the generalised stochastically bounded bursty traffic for communication network (further *gSBB*) described by J. Jiang, Q. Yin, Y. Liu and S. Jiang in 2009 was applied. For analytical description of this model *min-plus* algebra does not suit because analysis of processes is based on the total time of virtual processes; therefore, *max-plus* algebra means were chosen for analysis of network processes. Let us indicate as $A(k)$ a stationary and ergodic line which describes times of incoming applications traffic into a modelled network. $x_j(k)$ – k^{th} time of application service in a j node, when $t \geq 0$. $\sigma_j(k)$ is indicated as k^{th} request service beginning in a *j* line. A modelled system is open and consists of 10 nodes. System development is described by a *j+1* length vector $x(k) = (x_0(k),...,x_j(k))$ and a homogenous equation following from it $x(k+1) = A(k) \otimes x(k)$. On the ground of proofs provided by J. Jiang, Q. Yin, Y. Liu and S. Jiang in 2009, *gSBB* self-similar process is modelled according to the formula: $f^{self-similar}(x) = C_\alpha \left(\frac{\rho - m}{\delta} \right)^{-\alpha}$ when inequality $P\left\{ \hat{A}(t;\rho) > x \right\} \leq f^{self-similar}(x)$ is satisfied. Here C_α and *m* are calculated as $S_\alpha(\beta, \sigma, \mu)$ and *x* parameters. Poisson traffic is modelled according to the formula: $f^{Poisson}(x) = 1 - (1-\eta) \cdot \sum_{i=0}^{k} \left[\frac{[\eta(i-k)]^i}{i!} e^{-\eta(i-k)} \right]$, when inequality $P\left\{ \hat{A}(t;\rho) > x \right\} \leq f^{Poisson}(x)$ is satisfied, Here, $\eta = \frac{\lambda \bar{S}}{\rho}$, and $k = \left[\frac{x}{\bar{S}} \right]$, where $\bar{S}$ is average length of packages.

67760 queues were analysed combining network node and negated traffic parameters. Applied linear regressive analysis showed that when λ more than 10 times surpasses μ *Poisson* traffic applications were serviced better when network node queue discipline was *LIFO*. Self-similar incoming traffic requests waited in a queue shorter for 14 times on average and delay was 1.4 times lower in a network node than for incoming *Poisson* traffic.

6 Method of real-time self-similarity measurement

It should be admitted that methods of real-time study of network traffic load self-similarity features still have not been sufficiently explored. Dynamic models for changing network node features regarding network traffic parameters were analysed in particular, narrow aspects. In 2008 F. Mario, G. Nuno, M. Hajduczenia, P. Inacio, P. Monteiro and H. Silva patented (patent EP1983687A1) a device for network traffic analysis which automatically formed a time line of 1,000 members, later supplemented by new sequence members not changing the overall line length. A line is formed in a buffer, here R/S statistics, time dispersion, residual dispersion, time absolute moments, highest probability of narrowing are calculated, signal waves are assessed, methods of included split and trendless fluctuation analysis are applied. Results are classified and stored in the output buffer. Drawback of this invention lies in fixed length of a line and only one aggregation method used for measurements, also *Hurst* coefficients are not stored for further analysis.

Worked out and tested method of real-time analysis of computer network traffic (patent LT20011099) recurrently forms a line, calculates its statistical parameters and stores results in temporal memory, differs from others by the following:

- real-time aggregates network traffic by applying the method of smoothing moving averages or a sum of transmitted data traffic during time interval,
- stores a formed line in temporal memory,
- recurrently calculates parameters of a dynamically formed line self-similarity,
- stores calculation results for further analysis in temporal or/and permanent memory,
- calculates self-similarity parameters by applying *robust regression* method, *IR* statistics and *R/S* statistic,
- for assessment of self-similarity calculates α stability index, *IR* estimator and *Hurst* coefficient.

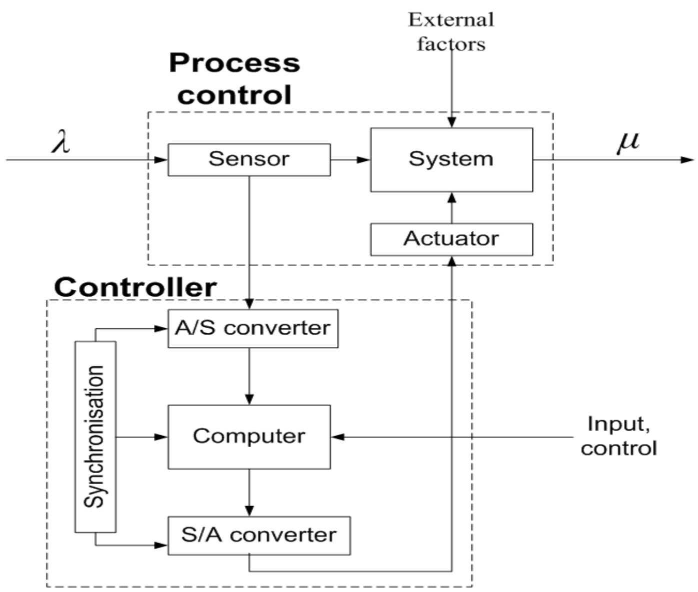

Figure 1 Open model of network node control including feedback

On the ground of real-time analysis of computer network traffic, the model of computer network node control has been worked out (see Figure 1); it ensures dynamic control over computer network traffic. Here, a computer node sends copies of headlines of received packages to a computer and performs the role of a system sensory device. The computer real-time measures characteristics of network packet traffic and with regard to the obtained results changes the length of node queue, queue service discipline through variables (see Figure 1 Actuator). Processes synchronisation device (see Figure 1 *Synchronisation*) ensures interaction of projected model's components and orderly information exchange. Converters ensure matching of the transmitted information. The described model was tested in a computer-assisted simulation way. For testing the library of program modules *SSE-OL* (*Self-Similarity Estimator On-Line*) had been created. The library can be integrated into customer's computer or network node which has network operating system and to perform real-time measurements. Testing was carried out in Šiauliai University Distance Study Centre's network node where network node server of average power (according to Tom's Hardware processor index 38677) with Windows 2003 Server Standard Edition operating system is used. Testing results showed that processor's load changed from 1% to 5%, and processing memory was used from 44 MB to 50 MB. Calculation of self-similarity estimators according all three methods for one line lasts 14.1329 ms on the average and lien formation time depends on a chosen line aggregation interval Δt.

7 Conclusions

Research of simulated time lines with *Fractan*, *Selfis* and *SSE* showed that the variation of the asymmetry coefficient β for time lines attributed with self-similarity had minimal impact on *Hurst* coefficient and stability parameter α meaning. Comparative analysis of results of simulation and real network traffic lines obtained by *Fractan*, *Selfis* and *SSE* showed that estimators of synthetic traffic self-similarity were calculated most exactly when applying the regression method. Results of research of real-time network traffic aggregated lines show that the process described by time lines is a persistent process with *LRD* when *Hurst* coefficient meaning changes from 0.61 to 0.79, moreover, computer network traffic under investigation is attributed with fractality.

By using the created package for modelling network traffic *MulNodSimSys* the impact of self-similarity of incoming and service traffics on network node operation was investigated with regard to the service queue discipline and buffer capacity, load and service intensities and self-similarity. It was estimated that when network traffic intensity increased then the traffic attributed with self-similarity was serviced stably. Servicing *Poisson* traffic, probability of package loss increases when intensity of incoming traffic increases. The queue service discipline LIFO when the traffic is *Poisson* traffic and service is attributed with self-similarity, ensures higher service quality. When network node buffer capacity increases, traffic service probability increases if it is attributed with self-similarity. When stability parameter of a self-similar line increases then service probability increases. Stronger dependence is observed when the incoming traffic is self-similar and service traffic is attributed with self-similarity. On the ground of linear regression results and determination coefficient r^2 very strong dependence was estimated between queue length average package time in network node queue, average delay and queue completion time, and average dependence was estimated between queue length and service as well as loss probability, and Fisher statistics F and P_F show that the link between dependent and independent variables is not stochastic. It was estimated that average delay was higher when the incoming traffic was Poisson traffic and service was attributed with self-similarity when queue service discipline was *FIFO*. Results of analysis of obtained meanings showed that the incoming traffic was Poisson traffic and $(\lambda / \mu) \geq 10$, then intensity of the incoming traffic more impacted the probability of loss, average time of request in a buffer and average delay, and queue service was better when its service discipline was *LIFO*. It was estimated that when incoming traffic is self-similar then the average time of a request in a buffer and the average delay independently from the queue service discipline was lower, when all other parameters of the generated traffic were the same. After analysing calculation results it was obtained that when self-similarity coefficient increased then the average amount of requests in the system slightly increased. The queue service discipline completely does not impact other variables and α dependences. After assessment of standard deviation of requests serviced in channels it was obtained that when incoming traffic intensity increased channels were loaded more evenly. It should be admitted that at the same incoming traffic intensity the efficiency of self-similar network node traffic is more than twice higher than that of *Poisson* traffic.

The method for analysis of computer network packet traffic self-similarity was created and patented in the Patent Bureau of Lithuania. The most suitable methods of real-time calculation of self-similarity: *robust regression* method and *IR* statistics. The recurrent algorithm of real-time formation of lines using the method of line re-indexing has been worked out. By using the method of real-time analysis of self-similarity of package-based network traffic, an open model for network node control including feedback has been worked out.

Publications

1. L. Kaklauskas, L. Sakalauskas (2011). Study of the Impact of Self-Similarity on the Network Node Traffic. Electronics and electrical engineering, No. 5(111), p. 27-32.
2. L. Kaklauskas, L. Sakalauskas (2011). Study of on-line measurement of traffic self-similarity. Central European Journal of Operations Research, DOI: 10.1007/s10100-011-0216-5. Accepted on 16-06-2011. http://www.springerlink.com/content/0036x207656pq211/
3. L. Kaklauskas, L. Sakalauskas (2009). Application of Chaos Theory to Analysis of Computer Network Traffic. Proceedings of XIII International Conference "Applied Stochastic Models and Data Analysis", Vilnius Lithuania, 407-412. (ISI Proceedings).
4. L. Kaklauskas, L. Sakalauskas (2008). On network traffic statistical analysis. Lietuvos matematikos rinkinys. LMD darbai, 48/49, 2008, pp. 314–319.
5. L. Kaklauskas, L. Sakalauskas (2010). Tinklo apkrovos savastingumo tyrimas realiu laiku. „Informacijos mokslai", 53 tomas, pp. 100-105.

6. L. Kaklauskas, L. Sakalauskas (2010). Savastingumo įtakos tinklo mazgo darbui tyrimas. Jaunųjų mokslininkų darbai, Nr. 3, 100-107.
7. L. Kaklauskas, M. Vaičiulis (2011). Skirstinio uodegos indekso MAX-DPR įvertis. Jaunųjų mokslininkų darbai, Nr. 1, p 126-130.

Briefly about the author

Liudvikas Kaklauskas in 1975 graduated from Radviliškis district secondary school No. 3. In 1980 graduated from Šiauliai Pedagogical Institute and was awarded the qualification of a teacher of mathematics and physics. In 2002 graduated from Šiauliai University In-service Training Institute and was awarded the qualification of a teacher of informatics. In 2007 passed Cisco CCNA examinations, certificate No. CSCO11291846. Doctoral student at the Institute of Mathematics and Informatics from 2008 to 2012. Works as a lecturer at the Department of Informatics at Šiauliai University Faculty of Mathematics and Informatics. E-mail: liudas@fm.su.lt.

REZIUMĖ

Tyrimų sritis

Žinoma, kad per kompiuterių tinklo mazgą perduodamų duomenų paketų informacija yra dinaminė, sunkiai prognozuojama ir primena triukšmą. Šios informacijos tradiciniais matematiniais metodais dažniausiai neįmanoma aprašyti. *Ethernet* tinklo paketinio srauto tyrimai, atlikti 1989 m. A. Erramilli, O. Narayan ir W. Willinger, patvirtino, kad šio srauto charakteristikos pasižymi fraktališkumu bei savastingumu (angl. *Self-Similarity*) su ilgalaike atmintimi (angl. *Long-Range Dependence* – toliau *LRD*). Pastaroji savybė leidžia prognozuoti kompiuterių tinklo srauto kaitą, ir prognozės rezultatus taikyti srauto aptarnavimo kokybei *QoS* (angl. *Quality of Service* – toliau *QoS*) gerinti, reguliuojant tinklo srauto paketų vėlinimą, srauto fluktuacijas bei paketų praradimą. Paketinio tinklo srauto savastingumo vertinimui naudojami statistiniai laiko eilučių analizės metodai, dažninių/banginių savybių įverčiai, skaičiuojami stabiliųjų atsitiktinių dydžių stabilumo parametrai bei taikoma chaoso teorija. 2008 metais D. Surgailis, G Teyssiere ir M. Vaičiulis sukūrė tinklo srauto su LRD savastingumo analizės *IR* (angl. *Increment Ratio* – toliau IR) statistiką. IR statistika disertacijoje pirmą kartą pritaikytą kompiuterių tinklo srauto savastingumo analizei.

Išsamesniam tinklo procesų tyrimui naudojami tinklo srauto matematiniai modeliai. Juose, taikant netiesinės analizės metodus, vertinami dideli tinklo srauto pliūpsniai, aprašomi tikimybiniais modeliais su „sunkiomis uodegomis“ (angl. *Heavy-Tail*). Šiuolaikinio tinklo srauto analizei plačiai naudojami *ON-OFF* tinklo srauto modeliai, kurie pagrįsti kompiuterių tinklo srauto savybėmis. Savastingojo srauto modeliavimui sėkmingai taikomi aptarnavimo teorijos instrumentai. Disertacijoje tinklo srauto savybės vertintos jį modeliuojant lanksčiais *α*–stabiliaisiais skirtiniais, kurie paklūsta centrinei ribinei teoremai, aprašo sunkias uodegas, asimetriją ir pasižymintys leptokurtiškumu.

Naudojant savastingųjų tinklo srautų modelius galima vertinti tinklo srauto savybių įtaką tinklo mazgo (komutatorius, maršrutizatorius, serveris ir kt.) parametrams. Tokiuose modeliuose naudojamas sintetinis tinklo srautas, sugeneruotas pagal tyrėjo nurodytus parametrus, arba realiame kompiuterių tinkle surinkti srauto duomenys. Kai tinklo srautas neturi prioritetinių duomenų paketų, tinklo mazguose eilių (mazgo buferio) aptarnavimui naudojamos *FIFO* ir *LIFO* eilių disciplinos. 2009 metais J. Nzouonta, T. Ott ir C. Borcea nustatė, kad mazguose, aptarnaujančiuose didelės spartos (apie 30Gb/s) ir pralaidos kompiuterių tinklus, pilnai pakanka 15-20 paketų buferio. Tinklo srauto savastingumo savybė gali būti sėkmingai naudojama tinklo paketų transporto srauto perkrovų reguliavimui.

Vertinant paketinio tinklo srauto savybes realiu laiku, atsiranda galimybė dinamiškai prognozuoti srauto kaitą. Tinklo srauto dinaminius įverčius galima taikyti srauto aptarnavimo kokybės *QoS* gerinimui, taip užtikrinant optimalų atėjusių paketų aptarnavimą, minimalius paketų praradimus tinklo mazge. Nustatyta, kad tinklo srauto aptarnavimas mazge gerėja parenkant optimalų eilės ilgį, eilės aptarnavimo discipliną, mažinant paketų vėlinimą bei keičiant kitus mazgo parametrus taip, kad jie geriausiai atitiktų dinaminius tinklo srauto savybių įverčius.

Problemos aktualumas

Kompiuterių tinklai, naudojami nuotolinėms studijoms, dažnai patiria neprognozuojamas perkrovas dėl studijų proceso specifiškumo. Siekiant sumažinti tinklo perkrovas ir su tuo susijusius perduodamos informacijos praradimus reikia stebėti tinklo ir jo mazgų būseną ir, remiantis gautais rezultatais, prognozuoti tinklo bei jo mazgų apkrovas bei perkrovas. Kompiuterių tinklo paketinio srauto tyrimai rodo, kad jis turi fraktalų savybių bei pasižymi savastingumu su *LRD*. Empiriniai Šiaulių universiteto e. studijų kompiuterių tinklo paketinio srauto tyrimai patvirtino šią išvadą. Tokį tinklo srautą galima analizuoti kaip fraktalinį procesą, pasižymintį antros eilės statistiniu savastingumu, charakterizuojamu fraktaliniu matu, atraktoriais, autokoreliacijos funkcijos savybėmis bei fazinės erdvės matu. Gausūs savastingųjų procesų kompiuterių tinkluose tyrimai rodo, kad galima prognozuoti srauto kaitą ir prognozės rezultatus taikyti tinklo srauto aptarnavimo kokybei užtikrinti.

Reikia pažymėti, kad nėra išsamių tyrimų, kuriuose būtų įvertinta tinklo mazgo buferio talpos ir eilės aptarnavimo disciplinos įtaka savastingojo tinklo srauto aptarnavimo kokybei. Vertinant tinklo srauto savastingumą realiu laiku galima dinamiškai prognozuoti srauto kaitą, o prognozės rezultatus galima taikyti tinklo srauto aptarnavimo kokybės *QoS* gerinimui, reguliuojant paketų užlaikymą ir trumpinant jų aptarnavimo laiką tinklo mazge, mažinant srauto fluktuacijas.

Lietuvos mokslininkai dažniausiai analizuoja teorinius tinklo srauto savastingumo analizės modelius, tačiau eksperimentinis modeliavimas retai atliekamas. Tinklo apkrovos savastingumo indikatoriai realiu laiku, taikant α-stabiliuosius modelius, nėra išsamiai tirti nei Lietuvoje, nei užsienyje. Reikia pažymėti, kad tinklo srauto apkrovos savastingumo savybių tyrimo realiu laiku metodai dar nėra pakankamai ištirti. Ši problema yra aktuali ir dar nepakankamai išnagrinėta teoriniu bei praktiniu aspektais. Šio darbo tyrimų sritis apima paketinio kompiuterių tinklo srauto fraktalinių procesų stebėjimą ir analizę, srauto savastingumo savybių realiu laiku įvertinimo metodų tyrimą bei srauto modelių eksperimentini tyrimą, atsižvelgiant į tinklo srauto savybes bei tinklo mazgo parametrus.

Tyrimų objektas

Disertacijos tyrimų objektas yra kompiuterių tinklo srauto savybės, tinklo mazgo savybių įtaka kompiuterių tinklo paketinio srauto aptarnavimui, realaus laiko tinklo srauto savybių analizės metodai ir jų taikymas tinklo srauto kaitos dinaminiam prognozavimui.

Tyrimų tikslas ir uždaviniai

Darbo tikslas ištirti fraktalinius procesus kompiuterių tinkluose, remiantis gautais rezultatais parinkti metodus tinkamus tinklo srauto analizei realiu laiku ir sukurti savastingumo matavimo realiu laiku metodiką bei ją pritaikyti kompiuterių tinklų aptarnavimo kokybei gerinti.

Siekiant šio tikslo, darbe sprendžiami šie uždaviniai:

- išnagrinėti kompiuterių tinklų fraktališkumo (savastingumo) tyrimo metodus;
- parinkti metodus tinklo srauto savastingumo analizei realiu laiku;
- sukurti kompiuterių tinklo srauto savastingumo matavimo realiu laiku metodiką bei atitinkamus įrankius;
- sukurti kompiuterių tinklų informacinių srautų imitavimo sistemą ir pritaikyti ją sukurtai metodikai testuoti;

- pritaikyti sukurtą metodiką kompiuterių tinklų aptarnavimo kokybei *QoS* gerinti.

Mokslinis naujumas

Darbe gauti tokie rezultatai:

1. Parengtas tinklo srauto savybių analizės paketas, panaudotas analizuoti, vertinti ir palyginti kompiuterių tinklų fraktališkumo ir savastingumo tyrimo metodus;
2. Ištirti kompiuterių tinklo paketinio srauto savastingumo analizės metodai:
    a. laiko analizės metodai, kai tiriama laipsninės pasirinktų sekų ir m dydžio blokų, atrinktų iš sekos, statistinės priklausomybės,
    b. dažninių/banginių savybių įvertinimo metodai, grindžiami dažninėmis bangos vilnelių savybėmis,
    c. laiko eilutės stabilumo parametrų įverčiai.
3. Sudarytas tinklo srauto savastingumo realiu laiku analizės paketas (toliau SRLAP).
4. Naudojant *SRLAP* savastingumo matavimui realiu laiku atrinktas robastinis *regresijos* metodas ir *IR* statistika.
5. Sukurta kompiuterių tinklo srauto savastingumo matavimo realiu laiku metodika, kuri rekurentiškai formuoja eilutę, skaičiuoja statistinius jos parametrus ir rezultatus saugo laikinojoje atmintyje. Ši metodika nuo kitų skiriasi tuo, kad realiu laiku agreguoja tinklo srautą taikant judančių vidurkių glotninimo arba perduotų duomenų srauto sumos per laiko intervalą metodą, suformuotą eilutę saugo laikinoje atmintyje, rekurentiškai skaičiuoja dinamiškai formuojamos eilutės savastingumo parametrus, apskaičiuojant α stabilumo indeksą, *IR* įvertį ir *Hurst* koeficientą, skaičiavimo rezultatus tolesnei analizei saugo laikinojoje arba/ir pastoviojoje atmintyje. Metodas užpatentuotas Lietuvos Respublikos valstybiniame patentų biure, patento numeris LT20011099.

6. Sukurta kompiuterių tinklo savastingojo ir *Puasono* srautų imitavimo sistema ir, ja naudojantis, iširta tinklo mazgo savybių įtaka paketinio tinklo srauto aptarnavimo kokybei.
7. Sukurta metodika pritaikyta kompiuterių tinklo aptarnavimo kokybei *QoS* gerinti.

Praktinė darbo reikšmė

Darbe gauti tokie praktiniai rezultatai:

1. Sukurta kompiuterių tinklo duomenų srauto savastingumo įvertinimo realiu laiku metodika bei įrankiai;
2. Sukurta kompiuterių tinklo savastingųjų srautų su sunkiomis uodegomis imitavimo metodika bei įrankiai;
3. Pasiūlyti savastingojo kompiuterių tinklo srauto su sunkiomis uodegomis aptarnavimo kokybės *QoS* sprendimai, gaunami įvertinus informacinių srautų savastingumą ir jų pagrindu suprojektuotas tinklo srauto valdymo realiu laiku modelis.

Ginamieji teiginiai

Pagrindiniai darbo teiginiai, pateikiami gynimui, yra šie:

1. Kompiuterių tinklo srauto savastingumo vertinimui naudojamos laiko sekų analize, dažninių/banginių procesų savybių vertinimu bei eilutės stabilumo parametrų įverčiais grįstos metodų grupės. Naujai pritaikyti laiko eilutės stabilumo parametrų įverčiai tinklo srauto savastingumo vertinimui realiu laiku, kurie gaunami apskaičiuojant: uodegų indeksą pagal robastinį *regresijos* metodą ir *IR* statistikos įverčius.
2. Paketinio kompiuterių tinklo srauto fraktaliniai procesai įtakoja tinklo srauto aptarnavimo kokybę *QoS* tinklo mazge ir priklauso nuo tinklo mazgo parametrų. Pagal tinklo srauto savybių įverčius, keičiant tinklo mazgo charakteristikas, eilės ilgį ir eilės aptarnavimo discipliną, mažėja paketų praradimo tikimybė, vidutinis paketo buvimo laikas tinklo mazgo eilėje ir

mazge, visiško eilės užpildymo laikas bei vėlinimas. Tokie mazgo savybių pokyčiai užtikrina geresnę tinklo srauto aptarnavimo kokybę tinklo mazge.

3. Naujas kompiuterių tinklo srauto savastingumo vertinimo realiu laiku metodas leidžia dinamiškai keisti tinklo mazgo savybes, taip užtikrinant geresnį srauto aptarnavimą tinklo mazge.

Rezultatų aprobavimas

Tyrimų rezultatai buvo pristatyti ir aptarti dviejose tarptautinėse konferencijose, viename tarptautiniame seminare, dviejose respublikinėse konferencijose bei trijuose respublikiniuose seminaruose. Publikuoti du straipsniai užsienio mokslo leidiniuose, įtrauktuose į Mokslinės informacijos instituto pagrindinių žurnalų sąrašą su citavimo indeksu, vienas – tarptautinių konferencijų darbuose, įtrauktuose į Mokslinės informacijos instituto sąrašą, ir keturi recenzuojamuose Lietuvos ir užsienio leidiniuose.

Disertacijos struktūra

1 skyriuje įvardinamas disertacijos objektas, tikslas, uždaviniai. Aptariamas darbo mokslinis naujumas ir jo praktinė reikšmė. Įvardinami pranešimai ir publikacijos paskelbtos disertacijos tematika.

2 skyriuje aprašomas kompiuterių tinklų matematinis modelis. Šiuolaikinių paketinių tinklų technologiniai sprendimai, tinklo įrangos ir informacinių srautų perdavimo technologiniai sprendimai, paketų struktūra. Aprašomos tinklo komponentų matematinio modeliavimo galimybės, kompiuterių tinklo paketinio srauto modeliai ir modeliai naudojantys aptarnavimo teorijos instrumentus.

3 skyriuje pateikiamos fraktalinių srautų paketiniuose kompiuterių tinkluose matavimo technologijos. Aptariami technologiniai tinklo srauto matavimo sprendimai. Analizuojamas tinklo srauto parametrų įvertinimas pagal chaoso teoriją, skaičiuojant *Hurst'o* statistikas, gaunant stabiliųjų parametrų įverčius ir taikant IR statistiką.

4 skyriuje aptariamos tinklo srauto analizės priemonės, srauto analizės metodų testavimas, tinklo srauto laiko eilučių formavimas, laiko eilučių agregavimo problematika bei apibendrinami kompiuterių tinklo srauto laiko eilučių analizės rezultatai.

5 skyriuje analizuojama fraktalinių procesų imitavimo programinė įranga, aptariami vieno kanalo tinklo mazgo tyrimo rezultatai. Pritaikius stochastiškai apribotą tinklo komunikavimo modelį ir *algebra plus* priemones, aprašomas GI/G/10//N kompiuterių tinklo mazgas. Apibendrinami Puasono ir savastingojo tinklo srautų savybių įtakos jo aptarnavimui tinklo mazge tyrimų rezultatai.

6 skyriuje pateikiamas savastingumo matavimo realiu laiku metodas bei nagrinėjamas paketinio kompiuterių tinklo valdymo modelis. Čia aptariami pasaulyje naudojami analogiški matavimo metodai, jų privalumai ir trūkumai, aprašomas pateiktas išradimas – tinklo srauto analizės realiu laiku metodas. Analizuojamas atviras tinklo mazgo valdymo modelis su grįžtamuoju ryšiu, kuriame išradimas pritaikytas dinaminiam tinklo mazgo valdymui.

Išvados

Simuliuotų laiko eilučių tyrimai su *Fractan*, *Selfis* ir *SSE* parodė, kad asimetrijos koeficiento β pokytis laiko eilutėms pasižyminčioms savastingumu turi minimalią įtaką *Hurst'o* koeficiento bei stabilumo parametro α reikšmei. Imitavimo ir realaus tinklo srauto eilučių rezultatų, gautų su *Fractan*, *Selfis* ir *SSE*, lyginamoji analizė parodė, kad sintetinio srauto savastingumo įverčiai tiksliausiai apskaičiuojami taikant regresijos metodą. Realaus tinklo srauto agreguotų eilučių tyrimo rezultatai rodo, kad procesas aprašomas laiko eilutėmis yra persistentinis procesas su *LRD*, kai *Hurst'o* koeficiento reikšmė kinta nuo 0,61 iki 0,79, be to, analizuotas kompiuterių tinklo srautas pasižymi fraktališkumu.

Panaudojant sukurtą tinklo srauto modeliavimo paketą *MulNodSimSys* ištirta įėjimo bei aptarnavimo srautų savastingumo įtaka tinklo mazgo darbui, atsižvelgiant į aptarnavimo eilės disciplįną ir buferio talpą, apkrovos ir aptarnavimo intensyvumus bei savastingumą. Nustatyta, kad, didėjant tinklo srauto intensyvumui, savastingumu pasižymintis srautas yra aptarnaujamas stabiliai. Aptarnaujant *Puasono* srautą didėja paketų praradimo tikimybė, didėjant įeinančio srauto intensyvumui. Eilės aptarnavimo disciplina *LIFO*, kai įeinantis srautas yra Puasono, o aptarnavimo pasižymi savastingumu, užtikrina geresnę aptarnavimo kokybę. Didėjant tinklo mazgo buferio talpai, srauto aptarnavimo tikimybė didėja, jei jis pasižymi savastingumu. Didėjant savastingosios eilutės stabilumo parametrui, aptarnavimo tikimybė didėja. Stipresnė priklausomybė stebima, kai įeinantis srautas yra savastingasis ir aptarnavimo srautas pasižymi savastingumu. Remiantis tiesinės regresijos rezultatais ir determinacijos koeficientu r^2 tarp eilės ilgio ir vidutinio paketo buvimo laiko tinklo mazgo eilėje ir mazge, vidutinio vėlinimo ir visiško eilės užpildymo laiko nustatyta labai stipri, o tarp eilės ilgio ir aptarnavimo bei praradimo tikimybės vidutiniška priklausomybė, o Fišerio statistika F ir P_F rodo, kad stebimas ryšys tarp priklausomų ir nepriklausomų kintamųjų nėra atsitiktinis. Nustatyta, kad vidutinis vėlinimas yra didesnis, kai įeinantis srautas yra Puasono, o aptarnavimo pasižymi savastingumu, kai eilės aptarnavimo disciplina *FIFO*. Gautų reikšmių analizės rezultatai parodė, kad kai įeinantis srautas yra Puasono ir $(\lambda / \mu) \geq 10$, tai įeinančio srauto intensyvumas daugiausia įtakoja praradimo tikimybę, vidutinį paraiškos buvimo buferyje laiką ir vidutinį vėlinimą, o eilės aptarnavimas yra geresnis, kai jos aptarnavimo disciplina *LIFO*. Nustatyta, kad kai įeinantis srautas yra savastingasis, tai vidutinis paraiškos buvimo buferyje laikas ir vidutinis vėlinimas nepriklausomai nuo eilės aptarnavimo disciplinos yra mažesnis, kai visi kiti generuojamo srauto parametrai yra vienodi. Išanalizavus skaičiavimo rezultatus gauta, kad didėjant savastingumo koeficientui nežymiai didėja vidutinis paraiškų skaičius sistemoje. Kitų kintamųjų ir α priklausomybės visiškai neįtakoja eilės aptarnavimo disciplina. Įvertinus kanaluose aptarnautų paraiškų standartinį nuokrypį gauta, kad didėjant įeinančio srauto intensyvumui kanalai apkraunami tolygiau. Pažymėtina, kad esant tam pačiam įeinančio srauto intensyvumui savastingojo tinklo mazgo srauto aptarnavimo našumas yra daugiau nei dvigubai didesnis nei tinklo mazgo, kurio srautas yra *Puasono*.

Sukurtas ir Lietuvos patentų biure užpatentuotas paketinio kompiuterių tinklo srauto savastingumo analizės metodas realiu laiku. Įvertintos ir parinktos tinkamiausios savastingumo skaičiavimo realiu laiku metodikos: *robastinis regresijos* metodas ir *IR* statistika. Sukurtas rekurentinis eilučių formavimo realiu laiku algoritmas, naudojantis eilutės perindeksavimo metodą. Panaudojant paketinio kompiuterių tinklo srauto savastingumo analizės realiu laiku metodą parengtas atviras tinklo mazgo valdymo modelis su grįžtamuoju ryšiu.

Glaustai apie autorių

Liudvikas Kaklauskas 1975 metais baigė Radviliškio rajono trečią vidurinę mokyklą. 1980 metais baigė studijas Šiaulių pedagoginiame institute ir jam suteikta matematikos-fizikos mokytojo kvalifikacija. 2002 metais baigė studijas Šiaulių universiteto Kvalifikacijos kėlimo institute ir įgijo informatikos mokytojo kvalifikaciją. 2007 metais išlaikė Cisco CCNA egzaminus, sertifikato Nr. CSCO11291846. Nuo 2008 iki 2012 metų Matematikos ir informatikos instituto doktorantas. Dirba Šiaulių universiteto Matematikos ir informatikos fakulteto Informatikos katedros lektoriumi.

El. paštas: liudas@fm.su.lt.