

Tinklo apkrovos savastingumo tyrimas realiu laiku

Liudvikas Kaklauskas

Matematikos informatikos instituto
Sistemų analizės skyriaus Operacijų tyrimo
sektorius doktorantas
P. Višinskio g. 19, LT-77156, Šiauliai
El. paštas: liudas@fm.su.lt

Leonidas Sakalauskas

Matematikos informatikos instituto
Sistemų analizės skyriaus Operacijų tyrimo
sektorius vyriausiasis mokslo darbuotojas,
sektorius vadovas, profesorius habilituotas
daktaras
Akademijos g. 4, 319 kab., Vilnius
El. paštas: sakal@ktl.mii.lt

Straipsnyje analizuojami indikatoriai, taikomi tinklo apkrovos savastingumui tirti: Hursto indeksas, stabilumo indeksas, IR (Increment Ratio) statistika. Kompiuteriniu modeliavimu ištirtas šių indikatorių tinkamumas tinklo apkrovos savastingumui vertinti realiu laiku. Sukurta programinių modulių biblioteka SSE (Self-Similarity Estimator), skirta fiksuoti ir agreguoti tinklo duomenų paketus, vertinanti tinklo apkrovos srautų savastingumą realiu laiku. Naudojant SSE programinių modulių biblioteką, suformuotų laiko eilučių Hursto indeksas ir IR statistika apskaičiuotos naudojant analitines formules, o stabilumo indeksas – robustiniu empirinių kvantilių regresijos metodu. Modulių bibliotekos SSE analizės efektyvumas ištirtas kompiuterinio modeliavimo būdu apskaičiuojant savastingumo indikatorius stabilųjų procesų realizacijoms.

Pagrindiniai žodžiai: savastingumas (self-similarity), Hursto indeksas, stabilumo indeksas, IR statistika.

Empiriniai kompiuterių tinklo tyrimai patvirtina, kad šiuolaikinio kompiuterių tinklo parametrus modeliuoti netinka klasikiniai Markovo modeliai, kurie buvo sėkmingai naudoti klasikinių telefoninių tinklų rodikliams įvertinti (Kleinrock, 2002). 1989 m. A. Erramilli'o, O. Narayano ir W. Willingerio atlikti Etherneto tinklo srauto tyrimai patvirtino, kad Etherneto tinklo srautas turi fraktalų savybių bei pasižymi savastingumu¹ (angl.

Self-similarity) su ilgalaikė atmintimi (angl. *Long-range dependence*). Taip pat gali būti vartojamas atitinkamo savipanašumas². Gausūs empiriniai kompiuterių tinklo paketų srauto tyrimai rodo, kad jis pasižymi savastingumu (Beran, 1998; Kaj, 2002; Samorodnitsky, 2006a). Įvertinus šią savybę galima adekvačiai prognozuoti srauto kaitą ir prognozės rezultatus taikyti tinklo pralaidumui didinti bei jo aptarnavimo kokybei QoS (angl. *Quality of Service*) gerinti, reguliuojant paketų užlaikymo,

¹ Kadangi terminas „self-similarity“ neturi visuotinai pripažinto atitikmens lietuvių kalba, straipsnyje vartojamas terminas „savastingumas“. Šį terminą savo darbuose ir disertacijoje vartoja dr. A. Kabašinskas. Terminas taip pat minimas leidinyje „Athena“. Filosofijos studijos, žurnale „Sociologija, mintis ir veiksmai“.

² Kaip *self-similarity* atitinkamo lietuvių kalba vartojamas ir terminas „savipanašumas“. Šį terminą savo darbuose vartoja prof. habil. dr. D. Surgailis, prof. dr. J. Valantinas.

fluktuacijų ribojimo bei paketų praradimą. Kompiuterių tinkle dažnai patiriamas neprognozuojamas perkrovas efektyviai galima valdyti stebint ir matuojant tinklo mazgų apkrovas bei perkrovas. Nustatyta, kad per kompiuterių tinklo mazgą perduodamų duomenų paketų ilgių kaita yra dinaminė, sunkiai prognozuojama ir primena triukšmą (Kaklauskas, Sakalauskas, 2008; Ilnickij, 2004; He, Gao, Hou, Park, 2004; Kleinrock, 2002), kurio neįmanoma aprašyti tradiciniais matematiniais metodais. Šiuolaikinių komunikacinių tinklų srautų modeliai sudaromi naudojant statistinės analizės metodus bei matematinį modeliavimą, modeliuojant tinklo srautą savastinguoju procesu. Tinklo procesams modeliuoti ir aprašyti rekomenduojama taikyti netiesinės analizės metodus, įvertinant didelius tinklo srautų plūpsnius, aprašomus tikimybiniais modeliais su „sunkiomis uodegomis“³ (angl. *Heavy-tail*) (Sheluhin, Smolskiy, Osin, 2007).

Tyrimo tikslas – ištirti indikatorius ir jų matavimo metodus, tinkamus tinklo apkrovos savastingumo realiu laiku matavimo metodikai sudaryti. Tyrimo objektas – kompiuterių tinklo srauto savastingumo analizės ir matavimo metodai. Reikia pažymėti, kad tinklo srauto apkrovos savastingumo savybių tyrimo realiu laiku metodai dar nėra pakankamai ištirti. Dabar kompiuterių tinklo srauto tyrimui galima naudoti nemokamai platinamas programas: Selfis 0.1b bei Fractan 4.4. Jos įvertina jau suformuotas ir agreguotas eilutes, tačiau tinklo apkrovos matavimų realiu laiku neatlieka,

³ Šį terminą savo darbuose vartoja prof. habil. dr. R. Leipus, habil. dr. R. Norvaiša ir kt. Termino apibrėžimą galima rasti straipsnyje: Leipus R., Norvaiša R. Finansų rinkos teorijų pagrindai, *Pinigų studijos* 2003-4, Ekonomikos teorija ir praktika, p. 16.

tam reikia pasitelkti kitus programinius produktus.

Tinklo apkrovos savastingumo indikatoriai

Stochastinis procesas $Y(t)$, $t \geq 0$ yra savastingasis (Samorodnitsky, 2006a), jei galima rasti tokį H , kad visiems $c > 0$, būtų tenkinama lygybė: $(Y(ct), t \geq 0 \stackrel{d}{=} (c^H Y(t), t \geq 0))$, čia $\stackrel{d}{=}$ reiškia, kad ši lygybė galioja visuose funkcijos taškuose, išskyrus trūkio taškus. Jei savastingasis procesas $X = (X_1, X_2, \dots)$ yra stacionarus, tai dalinės sumos $S_n = X_1 + X_2 + \dots + X_n$, čia $n=1, 2, \dots$, kai $n \geq 1$, o $X_n = Y(i) - Y(i-1)$, $i = 1, 2, \dots$ tenkina lygybę $S_n \stackrel{d}{=} n^H S_1$. Eksponentės rodiklis H nusako stacionaraus savastingojo proceso X pasiskirstymo savastingumą ir vadinamas Hursto indeksu. Šio koeficiento reikšmė apibūdina laiko eilutės atminties tipą: $0 \leq H < 0,5$ – laiko eilutė pasižymi trumpalaikėmis atmintimi ir aprašo antipersistentinį procesą, t. y. jei vienu laiko momentu eilutės reikšmės didėjo, tai kitu tolesniu laiko momentu jos būtinai mažės; $0,5 < H < 1,0$ – laiko eilutė pasižymi ilgalaikėmis atmintimi ir aprašo persistentinį procesą, t. y. jei laiko eilutės reikšmės didėja, tai ir ateityje jos išsaugos tą pačią tendenciją. Hursto indekso reikšmei artėjant prie 0,5 reikšmės eilutėje didėja triukšmų, o jos pokyčiai yra sunkiai prognozuojami, todėl 0,5 reikšmė yra laikoma baltuoju triukšmu.

IR (angl. *Increment Ratio*) statistika (Surgailis, Teysiere, Vaičiulis, 2008) yra taikoma nustatyti, ar analizuojamas tinklo srautas yra savastingasis su ilgalaikėmis atmintimi. *IR* statistika skaičiuojama pagal formulę:

$$R = \frac{1}{N-3m} \sum_{k=0}^{N-3m-1} \frac{|\sum_{i=k+1}^{k+m} (X_{i+m} - X_i) + \sum_{i=k+m+1}^{k+2m} (X_{i+m} - X_i)|}{|\sum_{i=k+1}^{k+m} (X_{i+m} - X_i)| + |\sum_{i=k+m+1}^{k+2m} (X_{i+m} - X_i)|}.$$

Pasiklovimo lygmeniu $\alpha = 0,95$ galima teigti, kad agreguota eilutė pasižymi savastingumu su ilgalaikė atmintimi, jei tenkinama lygybė: $IR - \Lambda(0) > z_{\alpha} \sigma(0) \sqrt{\frac{m}{N-3m}}$, čia $\Lambda(0) = 0,5881$, $\sigma(0) = 0,2080$, $z_{0,95} = 1,64$ (Surgailis, Teyssiere, Vaičiulis, 2008).

Modeliuojant tinklo apkrovą, pasižyminčią savastingumu, galima taikyti alfa stabilijų procesų teoriją. Savastingas simetriškas procesas $Y(t)$, pasižymintis begaline dispersija, yra α -stabilusis procesas, jei jo nuokrypių tikimybės tenkina Pareto sąlygą: $P(|Y(t)| > x) \sim cx^{-\alpha}$, čia $x \rightarrow \infty$, o $c > 0$. Nesunku pastebėti, kad, kai $1 < \alpha < 2$, $Y(t)$ vidurkis yra baigtinis, o kai $0 < \alpha \leq 1$ – begalinis (Samorodnitsky, 2006a). α -stabilusis atsitiktinis dydis $S_{\alpha}(\beta, \sigma, \mu)$ yra apibūdinamas keturiais stabilumo parametrais:

- α – stabilumo indeksas $\alpha \in (0, 2]$, dar vadinamas uodegų indeksu, nusakantis proceso pliūpsniškumą,
- β – asimetrijos indeksas $\beta \in [-1, 1]$, nusakantis proceso poslinkį nulio atžvilgiu,
- σ – mastelio indeksas, $\sigma > 0$ ir nusako proceso elementų dydį,
- μ – padėties indeksas $\mu \in R$.

α -stabilijų atsitiktinių dydžių parametrams vertinti rekomenduojama naudoti robastinius metodus, nes jie pasižymi atsparumu klaidoms ir didele skaičiavimo sparta (Kabašinskas, Rachev, Sakalauskas, Sun, Belovas, 2009; Sun, Rachev, Fabozzi, 2007). Darbe pritaikytas empirinių kvantilių regresijos metodas (EKRM), kurį stabilijų dydžių parametrams vertinti pasiūlė Koutrouvel (1981). Šio metodo taikymą

finansinių eilučių analizei 2009 m. aprašė Kabašinskas, Rachev, Sakalauskas, Sun, Belovas. α ir σ skaičiuojami EKRM pagal formules:

$$\alpha = \frac{K_{s_1} - s_2 s_3}{K_{s_4} - s_3^2}, \quad \sigma = \tilde{\sigma} \left(0.5 \exp \left\{ \frac{s_4 s_2 - s_1 s_3}{K_{s_4} - s_3^2} \right\} \right)^{1/\alpha},$$

$$\text{čia } s_1 = \sum_{k=1}^K y_k w_k, \quad s_2 = \sum_{k=1}^K y_k w_k^2, \quad s_3 = \sum_{k=1}^K y_k,$$

$$s_4 = \sum_{k=1}^K w_k, \quad s_5 = \sum_{k=1}^K w_k^2, \quad w_k = \log |t_k|,$$

$$y_k = \log(-\log(|\Phi_n(t_k)|^2)), \quad t_k = \frac{\pi k}{25}, \quad \tilde{\sigma} \text{ abso-}$$

liutinis σ nuokrypis, K rekomenduojama reikšmė 10.

β ir μ skaičiuojami EKRM pagal formules:

$$\beta = \frac{s_5 s_6 - s_7 s_9}{s_5 s_8 - s_7^2}, \quad \mu = \tilde{\mu} + \sigma \left(\frac{s_8 s_9 - s_6 s_7}{s_5 s_8 - s_7^2} \right) - h\pi,$$

$$\text{čia } s_5 = \sum_{i=1}^L u_i^2, \quad s_6 = \sum_{i=1}^L q_i g_n(u_i), \quad s_7 = \sum_{i=1}^L q_i u_i,$$

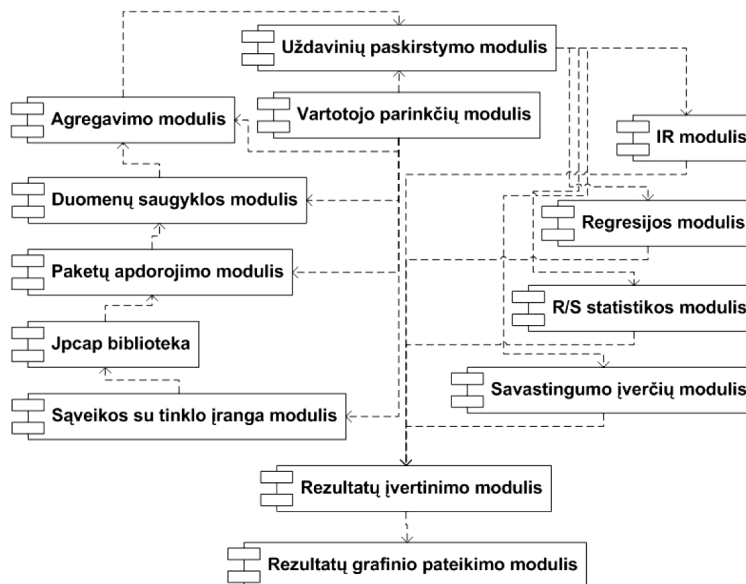
$$s_8 = \sum_{i=1}^L q_i^2, \quad s_9 = \sum_{i=1}^L u_i g_n(u_i), \quad u_i = \frac{\pi i}{50}$$

$$q_i = |\sigma u_i|^{\alpha} \tan\left(\frac{\pi \alpha}{2}\right) \text{sign}(u_i), \quad L=15.$$

Programinių modulių bibliotekoje naudojamas gerai žinomas R/S statistikų metodas, pagrįstas laiko sekų dalinių sumų analize (Samorodnitsky, 2006a, p. 167):

$$R/S = \frac{R(n)}{S(n)} = \frac{\text{Max}(\sum_{i=1}^{\tau} (x_i - \bar{x}_i)) - \text{Min}(\sum_{i=1}^{\tau} (x_i - \bar{x}_i))}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_i)^2}},$$

čia $1 \leq \tau \leq n$, n – sekos narių skaičius, \bar{x}_i – vidutinė eilutės x_i reikšmė, $\sum_{i=1}^{\tau} (x_i - \bar{x}_i)$ – suformuota kumuliatyvi eilutė, aprašanti pokyčių sumą per laiką τ . Pagal R/S sta-



1 pav. SSE programinių modulių bibliotekos sąveika

tistikas skaičiuojamas Hursto indeksas:
 $H = \log(R/S) / \log(n/2)$.

Savastingumo tyrimas realiu laiku

Savastingumo tyrimui realiu laiku Java kalba sudaryta SSE (angl. *Self-Similarity Estimator*) programinių modulių biblioteka, kurioje integruotos šios funkcijos: duomenų paketų atėjimo laiko ir ilgio fiksavimas, gautų laiko eilučių agregavimas, laiko eilučių fraktališkumo įvertinimas ir vertinimo rezultatų pateikimas. Java prietaigai prie fizinių tinklo įrenginių užtikrinti panaudota atviro kodo biblioteka *Jpcap* (2009).

Savastingumo tyrimas, naudojantis programinių modulių biblioteka, skaidomas į kelis etapus. Pirmiausia parenkamas agregavimo intervalas $\Delta t \in [10 \text{ ms}, 100 \text{ ms}, 1000 \text{ ms}, 10\,000 \text{ ms}]$ (*Vartotojo parinkčių modulis* žr. 1 pav.), atsižvelgiant į matuojamo tinklo srauto intensyvumą, ir nurodomas fizinis tinklo įrenginys, iš kurio

bus fiksuojami paketai (*Sąveikos su tinklo įranga modulis*). Toliau nurodomas duomenų saugyklos tipas bei tinklo apkrovos fraktališkumo tyrimo realiu laiku metodai bei analizės rezultatų perteikimo vartotojui būdai (*Vartotojo parinkčių modulis*). Suformuotos laiko eilutės agreguojamos (*Agregavimo modulis*). Paskui atliekama agreguotų laiko eilučių statistinė analizė, pasinaudojant moduliais, realizuojančiais IR metodą, EKSM, R/S statistikos apskaičiavimą. Skaičiavimo rezultatai pateikiami vartotojui grafiniu ir / arba tekstiniu pavidalu (*Rezultatų grafinio pateikimo modulis*).

Savastingumo matavimai ir tyrimas

Programinių modulių bibliotekoje SSE naudojamų skaičiavimo metodų tinkamumas realių srautų savastingumui vertinti iširtas pritaikant Monte Karlo metodą. Atsitiktiniams srautams imituoti buvo sukurtas modulis, generuojantis laiko eilutes,

aprašomas stabiliais dydžiais $S_\alpha(\beta, 1, 0)$, kai $\alpha \neq 1$ (Weron, 2004):

$$X = \mu + \sigma \cdot S_{\alpha\beta} \frac{\sin\{\alpha(V + B_{\alpha\beta})\}}{\{\cos(V)\}^{1/\alpha}} \left[\frac{\cos\{V - \alpha(V + B_{\alpha\beta})\}}{W} \right]^{(1-\alpha)/\alpha},$$

$$\text{čia } B_{\alpha\beta} = \frac{\arctan\left(\beta \tan \frac{\pi\alpha}{2}\right)}{\alpha},$$

$S_{\alpha\beta} = \left\{1 + \beta^2 \tan^2\left(\frac{\pi\alpha}{2}\right)\right\}^{1/(2\alpha)}$, V tolygiai pasiskirstęs intervale $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, W atsitiktinis dydis, generuojamas pagal formulę $W = -\ln(\xi)$, ξ – tolygiai pasiskirstęs intervale $[0, 1]$. Yra žinoma, kad Hursto indeksas ir stabilumo indeksas alfa susiję santykiu $H = 1/\alpha$, kai $1 < \alpha < 2$, $\beta = 0$ (Samorodnitsky, 2006b). Generuotos savastingosios eilutės su šiais parametrais: $\alpha = 1,8$ ($H = 0,56$), $\beta = 0$, $\sigma = 1$, $\mu = 0$, generuojamų eilučių skaičius – 1000, o narių eilutėje kiekis $n = 3000$. Gauti skaičiavimo rezultatai atitiko generuoti parinktas reikšmes, t. y. $1,73 < \alpha < 1,86$, o $0,538 < H < 0,578$. Rezultatų įverčiai pateikiami 1 lentelėje.

Programinių modulių biblioteką taip pat galima naudoti apkrovos savastingumui matuoti tinklo mazge bei kliento kom-

piuteryje. Pagal SSE matavimo rezultatus ir taikomą tinklo modelį galima prognozuoti srauto pliūpsnius tinklo mazguose ir imtis prevencinių priemonių jiems sumažinti.

Naudojant programinių modulių biblioteką SSE atlikti Šiaulių universiteto Nuotolinių studijų centro tinklo mazgo apkrovos matavimai realiu laiku. Apskaičiuotos 309 agreguotų laiko eilučių Hursto ir stabilumo indekso reikšmės. Skaičiavimo rezultatai pateikiami 2 lentelėje. Pagal IR metodą apskaičiuotas koeficientas 87,24 proc. rodė, kad matuojamos laiko eilutės pasižymi ilgalaikę atmintimi ($1 < \alpha < 2$). Vadinasi, agreguotomis eilutėmis aprašomas kompiuterių tinklu perduodamų duomenų paketų srautas yra persistentinis ilgalaikės atminties procesas. Keičiant agregavimo intervalą skaičiavimo rezultatai yra panašūs.

Išvados

Nuotolinių studijų centro tinklo mazgo apkrovos ir tinklo srauto eksperimentiniai tyrimai Monte Karlo metodu parodė, kad sudaryta programinių modulių biblioteka tinka tinklo srauto savastingumui vertinti realiu laiku.

1 lentelė. Savastingumo indikatorių, naudojant kompiuterinį simuliaciją, skaičiavimo rezultatai

Generavimo parametras	Mediana	Standartinis nuokrypis
H=0,56	0,5566	0,0096
$\alpha=1,8$	1,7921	0,1879

2 lentelė. Savastingumo indikatorių realiu laiku skaičiavimo rezultatai

Agregavimo intervalas	Hursto (mediana)	Hursto (standartinis nuokrypis)	α (mediana)	α (standartinis nuokrypis)
10 ms	0,5678	0,0857	1,7542	0,1889
100 ms	0,5611	0,0976	1,7637	0,2133
1000 ms	0,6524	0,1533	1,5328	0,3076

Įvertinus programa SSE gautų rezultatų standartinių nuokrypį, nustatyta, kad Hursto indekso, IR statistikos ir stabilumo indekso vertinimo metodai tinkami savastingumo parametro įverčiui realiu laiku.

Tinklo mazgo apkrovos matavimai parodė, kad perduodamų duomenų paketų srautas yra persistentinis ilgalaikės atminties procesas, nepriklausantis nuo agregavimo intervalo pokyčių.

LITERATŪRA

BERAN, J. (1998). *Statistics for Long-Memory Processes*. Capman & Hall / CRC, USA, 1998.

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: The International Journal of Computer and Telecommunications Networking*, Vol. 45, pp. 743–766.

ILNICKIJ, S. (2004). Research of the Network Server n Self-Similar Traffic Environment. *Telecommunications and Electronics*, Vol. 4.

JPCAP Java library. [interaktyvus]. [Žiūrėta 2009 m. gegužės 05 d.]. Prieiga per internetą: <<http://netresearch.ics.uci.edu/kfujii/jpcap/doc/index.html>>

KABAŠINSKAS, A.; RACHEV, S. T.; SAKALAUSKAS, L.; SUN, W.; BELOVAS, I. (2009). Alpha-stable paradigm in financial markets. *Journal of Computational Analysis and Applications*, 11(4), pp. 641–668.

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

KAKLAUSKAS, L.; SAKALAUSKAS, L. (2008) On network traffic statistical analysis. *Lietuvos matematikos rinkinys LMD darbai*, 48/49, p. 314–319.

KLEINROCK, L. (2002) Creating a Mathematical Theory of Computer Networks. *Operations Research*, Vol. 50, No. 1, pp. 125–131.

KOUTROUVEL, I. A. (1981). An iterative procedure for the estimation of the parameters of the stable law. *Communications in Statistics – Simulation and Computation*, 10, pp. 17–28.

SAMORODNITSKY, G. (2006a) Long Range Dependence. *Foundations and Trends in Stochastic Systems*. Vol. 1, No. 3, pp. 163–257.

SAMORODNITSKY, G. (2006b). Long memory and self-similar processes. *Annales de la faculté des sciences de Toulouse Sér. 6*, 15 No. 1, pp. 107–123.

SHELUHIN, O. I.; SMOLSKIY, S. M.; OSIN A. V. (2007). *Self-similar processes in telecommunications*. John Wiley & Sons, Ltd., 2007.

SUN W.; RACHEV S.; Fabozzi, F. J. (2007). Fractals or I.I.D.: Evidence of Long-Range Dependence and Heavy Tailedness from Modeling German Equity Market Returns. *Journal of Economics and Business*. Vol. 59, No. 6, pp. 575–595.

SURGAILIS, D.; TEYSSIERE, G.; VAIČIULIS, M. (2008). The increment ratio statistics. *Journal of Multivariate Analysis*. Vol. 99, No. 3, March 2008, pp. 510–541.

WERON, R. (2004). Computationally intensive Value at Risk calculations. [interaktyvus]. Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE) 32 [Žiūrėta 2009 m. gegužės 5 d.]. Prieiga per internetą: <http://www.econstor.eu/dspace/bitstream/10419/22205/1/32_rw.pdf>

THE REAL-TIME MODE RESEARCH OF NETWORK TRAFFIC FRACTALITY

Liudvikas Kaklauskas, Leonidas Sakalauskas

S u m m a r y

The article analyses the indicators implemented for investigating the network self-similarity: the Hurst index, stability index, IR (Increment Ratio) statistics. The suitability of these indicators for the on-line estimation of network traffic self-similarity was investigated by applying computer-based modelling. The software SSE (Self-Similarity Estimator) module library was developed; it was designed for the recording and aggregation of network traffic packages as well as for the on-line estimation of

network traffic self-similarity. By using the SSE software module library, the Hurst index and the IR statistics of time series were estimated by applying analytic formulas, and the index of stability was estimated applying the robust method of regression of empirical quantiles. The efficiency of the analysis of the SSE module library was investigated by estimating the self-similarity indicators for realisation of the stable processes while applying the method of computer-based modelling.