ZESZYTY NAUKOWE POLITECHNIKI RZESZOWSKIEJ 296, Elektrotechnika 36
RUTJEE, z. 36 (3/2017), październik-grudzień 2017, s. 43-52

Maksymilian BURDACKI [1]
Paweł DYMORA[2]
Mirosław MAZUREK[3]

ANALIZA RUCHU W SIECI KOMPUTEROWEJ W OPARCIU O MODELE MULTIFRAKTALNE

Celem badań była analiza ruchu w sieci komputerowej z wykorzystaniem wybranych modeli multifraktalnych. W części teoretycznej omówiono podstawowe zagadnienia związane z oprogramowaniem zbierającym dane w sieci komputerowej, klasyfikacją przebiegów czasowych przy użyciu wykładnika Hurst'a. Opisano metody wykorzystane do wyznaczenia widm multifraktalnych. W części badawczej dokonano analizy przepływu ruchu w sieci komputerowej na podstawie liczby pakietów oraz prędkości przesyłania danych. Wykonano analizę wykładnika Hurst'a wyznaczanego dla poszczególnych przebiegów czasowych. Dokonano analizy widm multifraktalnych utworzonych dla badanych rodzajów ruchu sieciowego.

Słowa kluczowe: analiza ruchu sieciowego, sniffing, analiza samopodobieństwa, analiza multifraktalna, wykładnik Hursta.

1. Analiza pakietów w sieci komputerowej

Analiza ruchu sieciowego (znana również jako sniffing, analiza pakietów) jest procesem polegającym na przechwytywaniu i dokładnym badaniu ruchu sieciowego w celu określenia co dzieje się w sieci. Narzędzie analizujące ruch sieciowy dekoduje pakiety danych powszechnych protokołów i wyświetla taki ruch w czytelnym formacie. Tego typu narzędzie nazywane jest snifferem. Nieautoryzowane sniffery są zagrożeniem dla bezpieczeństwa sieci komputerowej, ponieważ są trudne do wykrycia i mogą być umieszczone prawie wszędzie, co sprawia, że są one ulubioną bronią hakerów [1, 2].

Narzędzie analizujące ruch sieciowy może być samodzielnym urządzeniem z wyspecjalizowanym oprogramowaniem lub programem zainstalowanym na

[1] Autor do korespondencji: Maksymilian Burdacki, Politechnika Rzeszowska, adres e-mail: maxb931@gmail.com
[2] Paweł Dymora, Politechnika Rzeszowska, Zakład Systemów Złożonych, pawel.dymora@prz.edu.pl
[3] Mirosław Mazurek, Politechnika Rzeszowska, Zakład Systemów Złożonych, miroslaw.mazurek@prz.edu.pl

komputerze. Sniffery różnią się między sobą funkcjami takimi jak liczba wspieranych protokołów, które mogą być dekodowane, interfejsem użytkownika oraz graficznymi i statystycznymi możliwościami [1].

Sniffer do swojego działania wymaga połączenia sprzętu oraz oprogramowania. Programy analizujące ruch sieciowy różnią się między sobą, ale każdy z nich składa się z następujących części [1]:

- **Sprzęt.** Wiele sniffer'ów sieciowych pracuje ze standardowymi systemami operacyjnymi i kartami interfejsów sieciowych (NIC). Niektóre z nich wspierają tylko karty ethernet'owe lub bezprzewodowe, a inne wspierają różne adaptery i pozwalają użytkownikom na dostosowywanie swojej konfiguracji.
- **Sterownik przechwytujący ruch.** Jest on odpowiedzialny za przechwytywanie surowego ruchu sieciowego przez sniffer. Pozwala na odfiltrowanie ruchu, który ma być zachowany i przechowuje przechwycone dane w buforze. Sterownik przechwytujący ruch jest rdzeniem oprogramowania zbierającego ruch.
- **Bufor.** Ten komponent pozwala na przechowanie przechwyconych danych. Dane mogą być zapisywane w buforze dopóki nie zostanie on zapełniony lub przy użyciu metody rotacyjnej, w której nowe dane zastępują stare.
- **Analiza w czasie rzeczywistym.** Funkcja ta pozwala na analizę danych przesyłanych łączem w danej chwili. Niektóre sniffer'y używają tej funkcji w celu znalezienia przyczyny problemów dotyczących wydajności sieci.
- **Dekodowanie.** Ten komponent wyświetla zawartość ruchu sieciowego w czytelnej postaci. Programy analizujące ruch sieciowy różnią się pod względem liczby dekodowanych przez nie protokołów.

2. Klasyfikacja serii czasowych z wykorzystaniem wykładnika Hurst'a

Wykładnik Hurst'a jest miarą pamięci długoterminowej oraz fraktalności przebiegów czasowych. Na jego podstawie seria czasowa może być sklasyfikowana w trzech kategoriach. Wartość współczynnika $H=0,5$ oznacza losową serię czasową. Jeśli $H<0,5$ to oznacza serię antypersystentą. Natomiast jeśli $H>0,5$ to oznacza serię persystentą. Seria antypersystentna charakteryzuje się tym, że wartości górne są prawdopodobnie poprzedzone wartościami dolnymi i na odwrót. Seria persystentna posiada trend wzmacniający, co oznacza, że następna wartość jest prawdopodobnie taka sama jak obecna. W prognozowaniu serii czasowych pierwszym pytaniem, na które należy udzielić odpowiedzi jest to, czy badana seria jest możliwa do przewidzenia. Serie czasowe charakteryzu-

jące się dużą wartością wykładnika Hurst'a posiadają silny trend, a zatem są one bardziej przewidywalne niż te o wartości H zbliżonej do 0,5 [3-6].

Wykładnik Hurst'a nie jest obliczany, a szacowany. Istnieje wiele różnych sposobów pozwalających na szacowanie wykładnika Hurst'a. W celu oszacowania go należy cofnąć przeskalowany zakres w przedziale czasowym obserwacji. Jest to wykonywane poprzez podzielenie pełnej długości serii czasowej na krótsze serie czasowe i przeskalowany zakres jest obliczany dla każdej z nich. Minimalna długość wynosząca 8 jest zazwyczaj wybierana dla najkrótszych serii czasowych. Przykładowo jeśli seria czasowa posiada 128 obserwacji to jest ona dzielona na:

- 2 części składające się z 64 obserwacji każda,
- 4 części składające się z 32 obserwacji każda,
- 8 części składających się z 16 obserwacji każda,
- 16 części składających się z 8 obserwacji każda [7].

Po podzieleniu serii czasowej na części w celu oszacowania wykładnika Hurst'a dla każdej części obliczane są [8]:

- średnia serii czasowej,
- średnio-wyśrodkowana seria otrzymana poprzez odjęcie średniej od wartości serii,
- łączne odchylenie serii od średniej poprzez zsumowanie średnio-wyśrodkowanych wartości,
- zakres będący różnicą pomiędzy maksymalną wartością łącznego odchylenia i minimalną wartością łącznego odchylenia,
- odchylenie standardowe średnio-wyśrodkowanych wartości,
- przeskalowany zakres otrzymany poprzez podzielenie wcześniej obliczonego zakresu przez odchylenie standardowe.

Końcowym etapem jest uśrednienie przeskalowanego zakresu dla wszystkich części [5].

3. Metody tworzenia widm multifraktalnych

Dekompozycja multifraktalna pozwala na analizowanie procesów w małych skalach czasu. Umożliwia ona rozdzielenie danego procesu na podzbiory punktów, w których otoczeniu ma on zbliżone właściwości geometryczne przedstawiane przy pomocy wykładnika Höldera. Uzyskane podzbiory można następnie zmierzyć poprzez określenie ich wymiaru Hausdorffa. Rezultatem tych działań jest widmo multifraktalne charakteryzujące związek pomiędzy wymiarem Hausdorffa, a wartością wykładnika Höldera [8].

Przedziałowy wykładnik Höldera miary probabilistycznej μ w przedziale I wyraża się zależnością:

$$\alpha_\mu(I) = \frac{\log \mu\,(I)}{\log|I|}$$

W powyższej zależności $|I|$ oznacza miarę Lebesgue'a dla przedziału I.

Niech x będzie punktem z dziedziny miary μ oraz $\{I_k\}$ będzie ciągiem przedziałów takim, że:

$$x \in I_k$$

oraz

$$\lim_{k\to\infty}|I_k| = 0.$$

Wykładnik Höldera miary μ w punkcie x wyraża się wartością następującej granicy:

$$\alpha_\mu(x) = \lim_{k\to\infty}\alpha_\mu(I_k) = \lim_{k\to\infty}\frac{\log\mu(I_k)}{\log|I_k|}.$$

Wymiarem Hausdorffa zbioru **F** określa się następującą granicę:

$$\dim(\boldsymbol{F}) = \lim_{\delta\to 0}\frac{\log N_\delta(\boldsymbol{A})}{-\log\delta},$$

gdzie:

F – podzbiór n-wymiarowej przestrzeni euklidesowej;

A – zbiór n-wymiarowych kul takich, że $\boldsymbol{F \subseteq A}$**;**

$\boldsymbol{\delta}$ – średnica pokrycia A będąca średnicą największej z kul należących do pokrycia;

$N_\delta(\boldsymbol{A})$ - minimalna liczba kul wchodzących w skład pokrycia o średnicy δ.

Widmem multifraktalnym bazującym na dekompozycji multifraktalnej nazywamy związek pomiędzy wymiarem Hausdorffa zbioru punktów miary o określonym wymiarze punktowym, a wymiarem punktowym:

$$f_H(\alpha) = \dim(K_\alpha)\ ;\ K_\alpha = \{x:\ \alpha(x) = \alpha\}.$$

Definicja ta zakłada, że widmo multifraktalne jest obliczane dla miary probabilistycznej. W celu otrzymania widma multifraktalnego procesu stochastycznego należy liniowo przeskalować wartości procesu w taki sposób, aby realizacje przeskalowanego procesu były prawie zawsze miarami probabilistycznymi. W przypadku estymacji widma multifraktalnego dla realizacji procesów natężenia ruchu odnotowanych w pomiarach należy takie procesy przeskalować liniowo, w taki sposób, aby ich wartości spełniały warunek normalizacji. Przeskalowanie tego typu w żaden sposób nie zmniejsza ogólności rozważań, ponieważ widmo multifraktalne jest charakterystyką niezależną od wartości średniej analizowanej próby [4, 5, 8].

Dekompozycja multifraktalna rozdziela analizowany proces na zbiory punktów. Każdy z niech jest tak zwanym zbiorem Cantora. Zbiory te są fraktalami, a ich wymiar fraktalny jest różny od jedności [8].

Istnieją również inne metody pozwalające na wyznaczanie widma multifraktalnego: poprzez funkcję podziału lub poprzez histogram wymiaru punktowego [8].

Relacja określająca funkcję podziału:

$$S_\delta(q) = \sum_{C\in A} \mu(C)^q,$$

gdzie: A – pokrycie dziedziny miary μ o średnicy δ

Transformata Legendre'a pełni ważną rolę w obliczaniu widma multifraktalnego na podstawie funkcji podziału.

Transformatą Legendre'a funkcji $f: \boldsymbol{R} \to \boldsymbol{R}$ można nazwać następujące przekształcenie:

$$f^*(s) = \inf(sx - f(x)).$$

Dla funkcji wklęsłych i różniczkowalnych przekształcenie to przyjmuje postać:

$$f^*\big(s(x)\big) = x \cdot f'(x) - f(x)\ ;\ s(x) = f'(x).$$

Widmem mulitfraktalnym opartym na funkcji podziału określa się transformatę Legendre'a funkcji τ(q):

$$\tau(q) = \lim_{\delta\to 0} \frac{\log S_\delta(q)}{\log \delta},$$

$$f_L(\alpha) = \tau^*(\alpha) = q\tau'(q) - \tau(q)\ ;\ \alpha(q) = \tau'(q).$$

Kolejna metoda uzyskania widma multifraktalnego polega na obliczeniu granicy odpowiednio przeskalowanego histogramu wymiaru punktowego.

Za dziedzinę miary μ należy przyjąć przedział (0; 1). Podprzedział dziedziny miary wynosi:

$$I_k^n = [k \cdot 2^{-n}, (k+1)2^{-n}).$$

Niech:

$$Y_n(\alpha) = \frac{-1}{n \log 2} \|K_\alpha^n\|,$$

$$K_\alpha^n = \{x : x = k \cdot 2^{-n} \wedge \alpha(I_k^n) = \alpha, k \in \{0,1,\dots,2^n - 1\}\}.$$

Widmo multifraktalne f_G oparte na histogramie wymiaru punktowego opisuje zależność:

$$f_G(\alpha) = \lim_{n\to\infty} Y_n(\alpha).$$

Powyższa zależność określa widmo multifraktalne jako granicę histogramów wymiaru punktowego.

Wymienione wcześniej sposoby uzyskania widma multifraktalnego umożliwiają zdefiniowanie tzw. formalizmu multifraktalnego, który określa, że miarę można uznać za multifraktalną, gdy wszystkie sposoby umożliwiają uzyskanie podobnych rezultatów:

$$f_G(\alpha) = f_H(\alpha) = f_L(\alpha).$$

Powyższa zależność nie może być jednak spełniona dla procesów natężenia ruchu występujących w sieciach. Jest to spowodowane ograniczonymi możliwościami obserwacji tego typu procesów, a to prowadzi do ograniczenia dokładności estymacji widma multifraktalnego przy użyciu wcześniej wymienionych metod. Jednakże przybliżona estymacja widm f_G oraz f_L pozwala ustalić, czy istnieje możliwość scharakteryzowania obserwowanego strumienia z wykorzystaniem widma multifraktalnego [6, 8].

4. Badanie ruchu w sieci komputerowej w oparciu o modele multifraktalne

4.1. Metodologia badań

Analizowane na potrzeby artykułu dane pochodzą ze strony www.caida.org. Jest to strona internetowa organizacji CAIDA (ang. *Center for Applied Internet Data Analysis*), która zajmuje się gromadzeniem danych dotyczących przepływu ruchu sieciowego. Monitor zbierający dane internetowe *equininx-chicago* jest zlokalizowany w centrum danych Equinix w Chicago i jest połączony z łączem sieci szkieletowej ISP poziomu 1 pomiędzy Chicago i Seattle. W niniejszej pracy dokonana została analiza ruchu realizowanego w kierunku Seattle – Chicago. Badana seria danych składa się z 1024 pomiarów. Pomiary były wykonywane co 1 sekundę.

Pierwsza część badań polegała na przeanalizowaniu przepływu ruchu sieciowego. Porównano liczbę przesyłanych pakietów na przestrzeni całego badania oraz ich średnią prędkość przesyłania. Dokonanie tej analizy pozwala na wysunięcie odpowiednich wniosków dotyczących charakterystyki ruchu w badanej sieci.

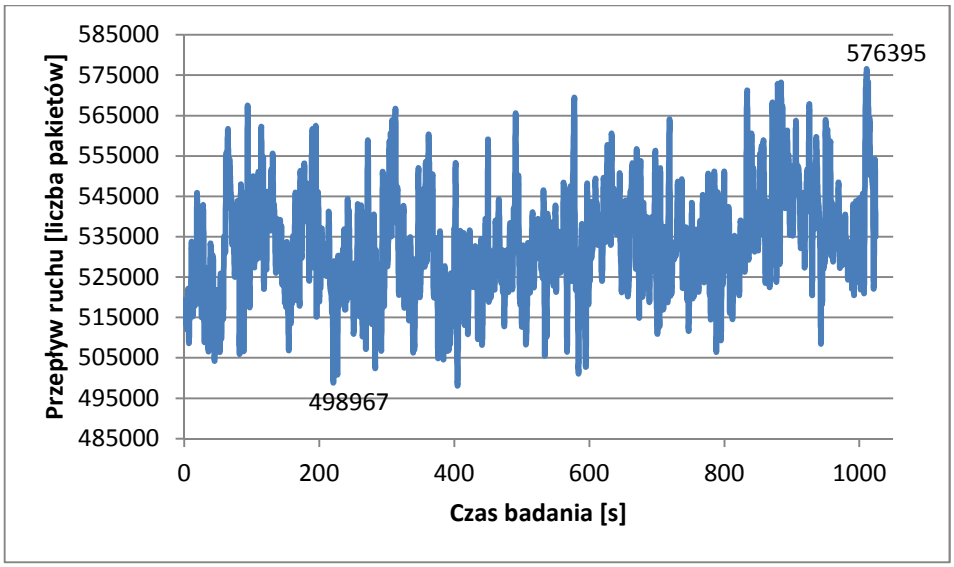
Dane poddano kolejnej analizie, mającej na celu wyznaczenie wykładnika Hurst'a dla poszczególnych rodzajów ruchu przy użyciu programu Selfis. Wartość wykładnika Hurst'a jest szacowana za pomocą czterech metod: wariancji skumulowanej, metody periodogramowej, wariancji szczątkowej oraz estymatora Whittle'a. Dokonanie tej analizy pozwala na określenie czy badana seria czasowa jest przewidywalna i posiada trend wzmacniający.

Trzeci etap badań polegał na wyznaczeniu i porównaniu widm multifraktalnych. Dla każdego rodzaju badanego ruchu wyznaczano dwa widma multifraktalne: widmo Legendre'a oraz widmo dużego odchylenia. Widma te są generowane przy użyciu programu FracLab będącego dodatkiem do oprogramo-

wania Matlab. Dokonanie tej analizy pozwala stwierdzić czy badane typy ruchu sieciowego posiadają własności multifraktalne.

4.2. Analiza przepływu ruchu sieciowego

Na rys. 1 przedstawiono przepływ całego zapisanego i poddanego analizie ruchu sieciowego.



Rys. 1. Całkowity ruch sieciowy

Fig. 1. Total network traffic

Najmniejsza zanotowana liczba pakietów przesłanych w ciągu jednej sekundy wyniosła 498967, a największa 576395. Dla całego ruchu sieciowego zmiana wyniosła 77428 pakietów, a odchylenie standardowe wyniosło 13099 pakietów. Średnia liczba pakietów przesłanych w ciągu jednej sekundy wyniosła 532673.

4.3. Analiza serii czasowych z wykorzystaniem wykładnika Hurst'a

Dla poszczególnych rodzajów ruchu sieciowego obliczono wartości wykładnika Hurst'a. Obliczenia były wykonywane dla okresów czasu wynoszących: 256 s, 512 s oraz 1024 s.

W tabeli 1 przedstawiono wartości wykładnika Hurst'a obliczonych dla całego ruchu sieciowego.

Tabela 1. Wartości wykładnika Hurst'a dla całego ruchu sieciowego

Table 1. Hurst exponent values for the entire network traffic

Metoda estymacji	Wartość H dla okresu czasu 256 s	Wartość H dla okresu czasu 512 s	Wartość H dla okresu czasu 1024 s
Wariancja skumulowana	0,252	0,618	0,834
Periodogram	1,224	1,081	0,919
Wariancja szczątkowa	0,942	0,847	0,887
Estymator Whittle'a	0,805	0,834	0,846

Dla okresu czasu wynoszącego 256 s większość estymatorów uzyskała wartość powyżej 0,5, co oznacza, że badany szereg czasowy jest persystentny i posiada trend wzmacniający. Wartość wariancji skumulowanej znacznie odstaje od reszty uzyskanych wyników.

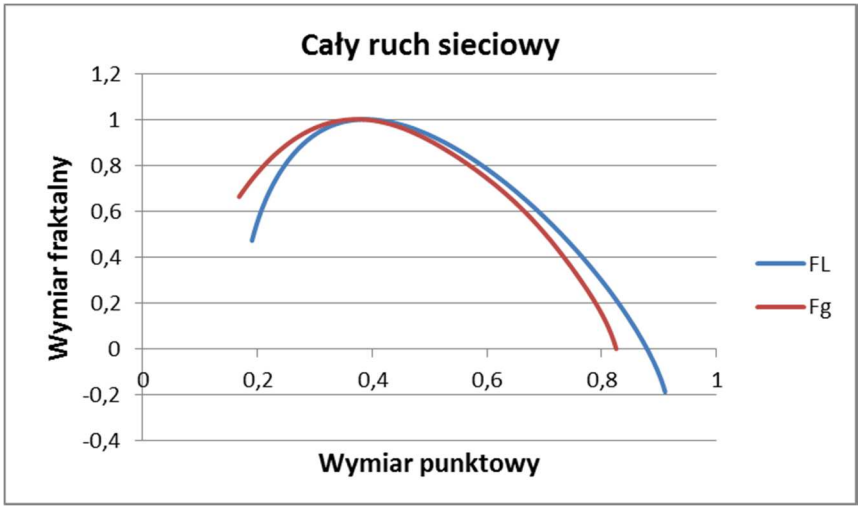
Wartości wszystkich estymatorów są większe od 0,5 dla okresu czasu wynoszącego 512 s, a to świadczy o persystentności badanej serii czasowej. Wartości wariancji skumulowanej oraz estymatora Whittle'a wzrosły w porównaniu do wartości tych estymatorów dla okresu czasu wynoszącego 256 s. Natomiast wartości periodogramu i wariancji szczątkowej zmalały w porównaniu do poprzednio analizowanego okresu czasu.

Dla serii składającej się z 1024 pomiarów wartości wszystkich estymatorów są większe od 0,5. W porównaniu do wartości osiągniętych dla okresu czasu wynoszącego 512 s wartości wszystkich estymatorów poza periodogramem wzrosły. Otrzymane wyniki świadczą o tym, że przepływ całego badanego ruchu sieciowego posiada trend wzmacniający i prawdopodobieństwo jego przewidzenia jest większe niż 50%.

4.4. Analiza widm multifraktalnych

Na rys. 2 przedstawiono widma multifraktalne utworzone dla poszczególnych analizowanych rodzajów ruchu sieciowego.

Widma multifraktalne odnoszące się do całego ruchu sieciowego mają podobny kształt i nieznacznie różnią się w fazie początkowej i końcowej. Przybliżona estymacja widm multifraktalnych świadczy o tym, że badany przepływ ruchu sieciowego posiada własności multifraktalne.



Rys. 2. Widma multifraktalne dla poszczególnych rodzajów ruchu sieciowego

Fig. 2. Multifractal spectra for particular types of network traffic

5. Podsumowanie i wnioski końcowe

Przeprowadzone badania ruchu w sieci komputerowej pozwoliły na określenie cech charakterystycznych dla badanej sieci. Analiza przepływu ruchu sieciowego umożliwiła zaobserwowanie jak kształtuje się zarówno cały ruch sieciowy jak i jego poszczególne rodzaje. W większości analizowanych przypadków wartości estymatorów wykładnika Hurst'a były większe od 0,5, co świadczy o tym, że analizowany ruch sieciowy może być przewidywany z prawdopodobieństwem większym niż 50% oraz posiada naturę samopodobną. W każdym analizowanym przypadku widma multifraktalne posiadały przybliżoną estymację.

Literatura

[1] http://scitechconnect.elsevier.com/wp-content/uploads/2013/09/Introducing-Network-Analysis.pdf, [Dostęp 18.04.2017].

[2] Dymora P., Mazurek M., Zelazny K., *Operating system efficiency evaluation on the base of measurements analysis with the use of non-extensive statistics elements*, Annales UMCS, Informatica. Volume 14, Issue 3, Pages 65–75, ISSN (Online) 2083-3628, 2014.

[3] Qian B., Rasheed K.: *Hurst exponent and financial market predictability*, University of Georgia, 2005.

[4] Dymora P., Mazurek M., *Network Anomaly Detection Based on the Statistical Self-similarity Factor*, Analysis and Simulation of Electrical and Computer Systems Lecture Notes in Electrical Engineering Volume 324, Springer, pp 271-287, 2015.

[5] Mazurek M., Dymora P., *Network anomaly detection based on the statistical self-similarity factor for HTTP protocol*, Przegląd elektrotechniczny, ISSN 0033-2097, R. 90 NR 1/2014, s.127 - 130, 2014.
[6] Brożek B., Dymora P., Mazurek M., *Badanie wydajności systemu operacyjnego zainfekowanego złośliwym oprogramowaniem z wykorzystaniem analizy samopodobieństwa, Zeszyty Naukowe Politechniki Rzeszowskiej 294, Elektrotechnika 35 RUTJEE, t. XXIV, z. 35 (2/16), kwiecień-czerwiec 2016, (p-ISSN 0209-2662, e-ISSN 2300-6358).*
[7] http://analytics-magazine.org/the-hurst-exponent-predictability-of-time-series/, [Dostęp 18.04.2017].
[8] Jędruś S.: *Modele multifraktalne natężenia ruchu sieciowego z uwzględnieniem samopodobieństwa statystycznego*. Telekomunikacja Cyfrowa: technologie i usługi T.4, 2001/2002.

COMPUTER NETWORK TRAFFIC ANALYSIS BASED ON MULTIFRACTAL MODELS

S u m m a r y

The aim of this work was computer network traffic analysis. Theoretical part de-scribes issues referring to network traffic capture software, time-series classification using Hurst exponent and multifractal spectrum creating methods. In research part was made an analysis of network traffic based on a number of packets and data transfer speed. It was also made a Hurst exponent analysis and a multifractal spectrum analysis for each type of analyzed network traffic. After the research it was possible to draw conclusions about characteristic of analyzed network traffic.

Keywords: network traffic analysis, sniffing, self-similarity analysis, multifractal analysis

DOI: 10.7862/re.2017.16

Tekst złożono w redakcji: wrzesień 2017
Przyjęto do druku: październik 2017