

Paweł DYMORA¹
Mirosław MAZUREK²
Sławomir JASKÓŁKA³

ANALIZA STATYSTYCZNA PRACY SIECI KOMPUTEROWEJ W ŚRODOWISKU LABVIEW

Analiza ruchu sieciowego oraz metod wykrywania nieprawidłowości pracy urządzeń sieciowych stanowi ciekawy problem dla analityków badających sieci komputerowe. Poprawna interpretacja anomalii oraz odpowiednia reakcja na nie może poprawić jakość działania sieci, zapobiec awarii lub skrócić jej czas. W artykule przedstawiono oryginalną aplikację utworzoną w środowisku LabVIEW, z zaimplementowanymi algorytmami do wyznaczania współczynnika Hursta, będącego miarą samopodobieństwa i określenia zależności długoterminowych oraz multifraktalności ruchu sieciowego. Celem aplikacji było zaimplementowanie znanych metod wyznaczania współczynnika Hursta m.in. metody statystyki R/S, metody wartości bezwzględnej oraz zagregowanej wariancji - jako aparatu statystycznego do określenia cech ruchu sieciowego. W badaniach wykorzystano wirtualną sieć testową, której model utworzono w środowisku OPNET Modeler. Przeprowadzona w programie analiza statystyczna wskazała, że poziom samopodobieństwa ruchu sieciowego zawiera się w przedziale od 0,5 do 1, zaś jego wartość przybiera wyższe wartości wraz z rosnącym wypełnieniem pasma sieci. Nieprzerwany ruch sieciowy o małym natężeniu (np. ruch typu VoIP) posiada samopodobieństwo porównywalne do szumu białego równe 0,5 co zaprezentowano w artykule.

Słowa kluczowe: współczynnik Hursta, wykrywanie anomalii sieci komputerowych, samopodobieństwo, zależności długoterminowe, systemy złożone.

1. Wprowadzenie

Ruch w sieci komputerowej bez względu na charakter usługi charakteryzuje się zarówno zależnościami długoterminowymi (dalekosięznymi), jak i multifrak-

¹ Autor do korespondencji: Paweł Dymora, Politechnika Rzeszowska, Zakład Systemów Rozproszonych, ul. W. Pola 2, 35-959 Rzeszów, tel. 17 865 1106, Pawel.Dymora@prz.edu.pl.

² Mirosław Mazurek, Politechnika Rzeszowska, Zakład Systemów Rozproszonych, ul. W. Pola 2, 35-959 Rzeszów, Mirosław.Mazurek@prz.edu.pl.

³ Sławomir Jaskółka, Politechnika Rzeszowska, Zakład Systemów Rozproszonych, ul. W. Pola 2, 35-959 Rzeszów, Sławomir.Jaskolka@prz.edu.pl.

talnością. Dowodem tego będzie wyznaczona wartość współczynnika Hursta zachowująca podobny poziom niezależnie od szerokości badanego okna pomiarowego (liczby próbek i czasu pomiaru). Analiza ruchu sieciowego oraz metody wykrywania nieprawidłowości i anomalii stanowi ciekawy problem dla analityków badających sieci komputerowe. Poprawna interpretacja anomalii oraz odpowiednia reakcja na nie może polepszyć jakość działania sieci jak również zapobiec lub skrócić czas awarii. W artykule zaprezentowano możliwości wykorzystanie współczynnika Hursta, który jest miarą samopodobieństwa do analizy wielu typów ruchu sieciowego w celu znalezienia zależności długoterminowych oraz multifraktalności. Odchylenia od wartości standardowych samopodobieństwa obliczone na podstawie pomiaru zużycia pasma oraz opóźnień sieciowych mogą być pomocne podczas identyfikacji typu ruchu sieciowego, zmian w topologii sieci, lub wczesnego wykrywania i zapobiegania awarii sieciowych.

2. Wybrane metody wyznaczania współczynnika Hursta

Najstarszą metodą do wyznaczania współczynnika Hursta jest metoda wykorzystująca statystykę R/S. Metodę R/S dla ciągu wartości o skończonej długości zarejestrowanych odstępów czasu pomiędzy zdarzeniami t_1, t_2, \dots, t_n definiujemy w następujący sposób [2, 4, 6, 8].

W pierwszym kroku wprowadza się średni odstęp czasowy dla zarejestrowanych ciągów odstępów pomiędzy zdarzeniami:

$$\hat{t}(n) = \frac{1}{n} \sum_{i=1}^n t_i \quad (1)$$

Następnie należy obliczyć wariancję tego ciągu:

$$S^2(n) = \frac{1}{n} \sum_{i=1}^n [t_i - \hat{t}(n)]^2 \quad (2)$$

oraz moment pojawienia się k -tego zdarzenia:

$$T_k = \sum_{i=1}^k t_i, \quad \text{dla } k = 1, 2, \dots, n \quad (3)$$

Należy także uwzględnić odchyłkę rzeczywistą T_k chwili pojawienia się k -tego zdarzenia w chwili uśrednionej $k\hat{t}(n)$, gdzie $k = 1, 2, \dots, n$:

$$U_k = T_k - k\hat{t}(n), \quad \text{dla } k = 1, 2, \dots, n \quad (4)$$

Statystykę R/S definiuje się w następujący sposób:

$$\frac{R(n)}{S(n)} = \frac{\max(0, U_1, U_2, \dots, U_n) - \min(0, U_1, U_2, \dots, U_n)}{\sqrt{S^2(n)}} \quad (5)$$

Dla zjawisk wskazujących samopodobieństwo można zaobserwować zależność przedstawioną jako:

$$\frac{R(n)}{S(n)} \approx n^H \quad (6)$$

gdzie zmienna H jest wykładnikiem Hursta. Tworząc wykres statystyki R/S w funkcji skali czasu należy przyjąć podwójnie logarytmiczny układ współrzędnych, gdzie otrzymamy zależność [1, 7]:

$$\log \frac{R(n)}{S(n)} \approx H \log(n) + const \quad (7)$$

która jest linią prostą o współczynniku kierunkowym H .

Kolejną metodą do wyznaczania współczynnika Hursta jest metoda wartości bezwzględnej. W celu wyznaczenia współczynnika Hursta tą metodą utworzono zagregowany szereg $X(m)$, dzieląc szereg badanych obserwacji o długości N na bloki o długości m i uśredniając każdy blok [1, 9].

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad \text{dla } k = 1, 2, \dots, N/m \quad (8)$$

Rozważając n -ty moment szeregu:

$$AM_n(m) = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \bar{X}|^n \quad (9)$$

gdzie \bar{X} jest średnią szeregów. Zagregowany szereg $X^{(m)}$ dla dużych wartości m asymptotycznie dąży do $Cm^{n(H-1)}$ i $AM_n(m)$ jest asymptotycznie proporcjonalne do $m^{n(H-1)}$. Nanosząc wyliczone wartości $AM_n(m)$ na wykres w podwójnej skali logarytmicznej i aproksymując uzyskane punkty metodą najmniejszych kwadratów, otrzymujemy prostą o nachyleniu do osi X równym $H-1$ [3, 7].

Ostatnią analizowaną metodą wyznaczenia współczynnika Hursta jest metoda zagregowanej wariancji. W metodzie tej utworzono zagregowany szereg

$X(m)$, a następnie podzielono szereg badanych obserwacji o długości N na bloki o długości m i uśredniono każdy blok [4, 8].

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad \text{dla } k = 1, 2, \dots, N/m \quad (10)$$

dla każdego m z zakresu od 2 do $N/2$ obliczana jest wariancja próbki $X^{(m)}(k)$:

$$\text{Var}X^{(m)} = \frac{1}{([N/m]-1)} \sum_{k=1}^{[N/m]} (X^{(m)}(k) - \bar{X})^2 \quad (11)$$



Na wykresie zaznaczane są punkty $\log(S^2 m)$ oraz $\log(m)$, które dla wystarczająco dużej wartości m , układają się w linię prostą o ujemnym współczynniku kierunkowym równym $2H - 2$ [3, 4, 9].






3. Model analizowanej sieci komputerowej

Symulację ruchu sieciowego w sieci komputerowej zamodelowano w programie OPNET Modeler. Testowaną sieć podzielono na dwie podsieci: podsieć stacji roboczych generujących ruch sieciowy (opisany w definicjach profili) oraz serwer znajdujący się w drugiej podsieci. Komunikacja między stacjami roboczymi a serwerem znajdującym się w innej podsieci wymaga przetworzenia ruchu sieciowego przez szereg urządzeń sieciowych, które analizując go generują opóźnienia. Obciążenie serwera oraz czasy odpowiedzi na żądania stacji roboczych dotyczących poszczególnych rodzajów ruchu sieciowego są mierzone i zapisywane podczas symulacji sieci. Testowa sieć wykorzystuje modele urządzeń podane w tabeli 1.

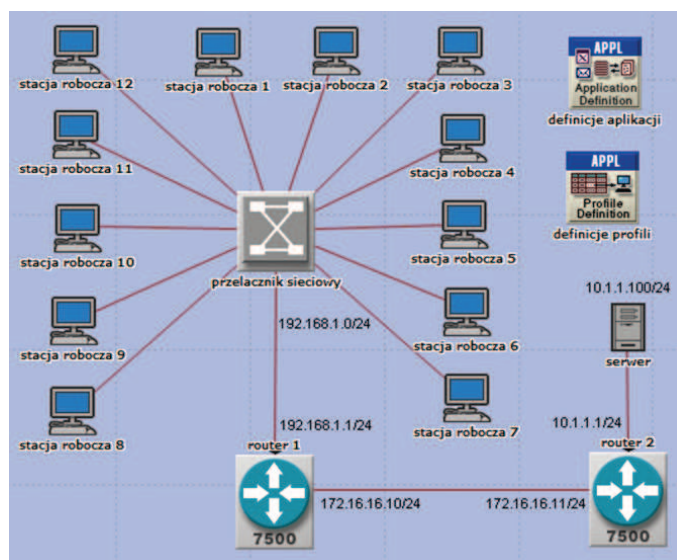
Tabela (tablica) 1. Opis elementów modelu sieci użytych w symulacji ruchu sieciowego

Table 1. Description of the network model elements used in the simulation of network traffic

| Ikona | Ilość | Opis |
|---|-------|--|
|  | 12 | Stacja robocza: urządzenie przeznaczone do bezpośredniej pracy inicjalizujące ruch sieciowy. |
|  | 2 | Router Cisco (model 7507): urządzenie pełniące rolę węzła komunikacyjnego łączące podsieci ze stacjami roboczymi i serwerem. |

| | | |
|---|----|--|
|  | 1 | Przełącznik Sieciowy: urządzenie łączące segment sieci komputerowej złożonej z 12 stacji roboczych do jednego z ruterów marki Cisco. |
|  | 1 | Serwer: obsługuje usługi sieciowe określone w definicjach aplikacji, które odpowiadają na zapytania stacji roboczych. |
|  | 1 | Definicje profili: określają profile, które wykorzystują aplikacje wyznaczone w definicjach aplikacji. |
|  | 1 | Definicje aplikacji: określają rodzaj, intensywność oraz czas trwania ruchu sieciowego. |
|  | 15 | 10Gbps Ethernet: łącze o przepustowości 10 gigabitów na sekundę. |

Schemat wzajemnych połączeń pomiędzy urządzeniami w symulowanej sieci komputerowej pokazano na rysunku 1.

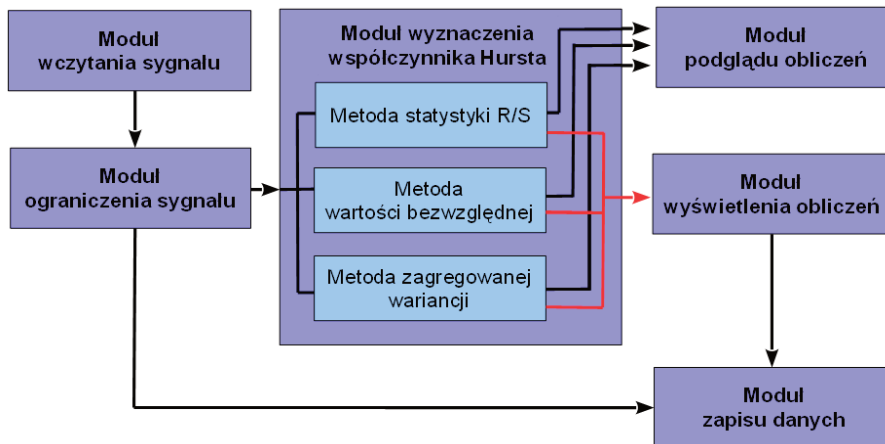


Rys. 1. Model symulowanej struktury sieciowej

Fig. 1. The model of the simulated network structure

4. Aplikacja wyznaczająca współczynnik Hursta

Do analizy ruchu sieciowego i wyznaczania wartości współczynnika Hursta w oparciu o przedstawione metody wykorzystano oryginalny program stworzony w środowisku LabVIEW. Schemat blokowy aplikacji przedstawiono na rysunku 2.



Rys. 2. Bloki funkcjonalne aplikacji

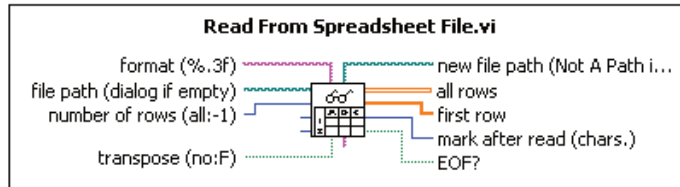
Fig. 2. Application functional blocks

W aplikacji wydzielono moduły, które działają niezależnie od siebie przekazując dane w sposób pokazany na rysunku 2. Dane wejściowe są wczytywane przez moduł wczytania sygnału. Następnie moduł ograniczenia sygnału umożliwia w prosty sposób edycję danych wejściowych poprzez wycięcie, powtórzenie zaznaczonej części sygnału lub nałożenie dwóch sygnałów na siebie. Przetworzony sygnał według preferencji użytkownika może być zapisany przez moduł zapisu danych i ponownie wczytany przez moduł wczytania sygnału jako podstawowe dane wejściowe.

Moduł wyznaczenia współczynnika Hursta pobiera dane z modułu ograniczenia sygnału i wyznacza trzema oddzielnymi metodami wyznacznika Hursta uwzględniając parametry zadane przez użytkownika. Wynik w postaci trzech wykresów (po jednym dla każdej z metod) przekazuje do modułu podglądu obliczeń, który wyświetla poszczególne kroki podjęte w celu wyznaczenia współczynnika Hursta. Moduł ten przekazuje także wynik jednej z zaznaczonych przez użytkownika metod do modułu wyświetlenia obliczeń, który prezentuje wykres wyznaczonego współczynnika Hursta. Program pozwala również na zapisanie do pliku tekstowego wykresu wyznaczonych współczynników. Odbywa się to poprzez ponowne wykorzystanie modułu zapisu danych.

4.1. Wczytywanie, edycja i prezentacja danych pomiarowych

W programie proces wczytywania danych wykonuje funkcja biblioteczna o nazwie „*Read From Spreadsheet File.vi*”. Po wybraniu ścieżki dostępu funkcja (rysunek 3) zwraca wynik w postaci tablicy liczb, które następnie są ograniczane zakresem podanym przez użytkownika i wyświetlane na wykresie w całości lub jako segment wybrany przez użytkownika.

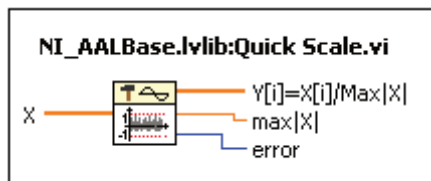


Rys. 3. Funkcja wczytująca plik tekstowy z danymi wejściowymi

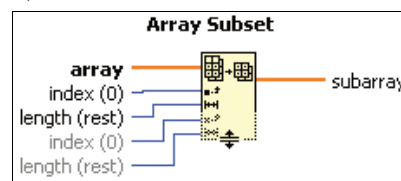
Fig. 3. Application's function loading the text file with the input data

Wczytany sygnał jest poddawany operacji normowania amplitudy, która odbywa się za pomocą funkcji bibliotecznej „*Quick Scale*” (rysunek 4a). Dzięki temu przekształceniu próbki analizowanego sygnału przyjmują wartości z przedziału $[-1; +1]$.

a)



b)



Rys. 4. a) Funkcja *Quick Scale*; b) Funkcja *Array Subset*

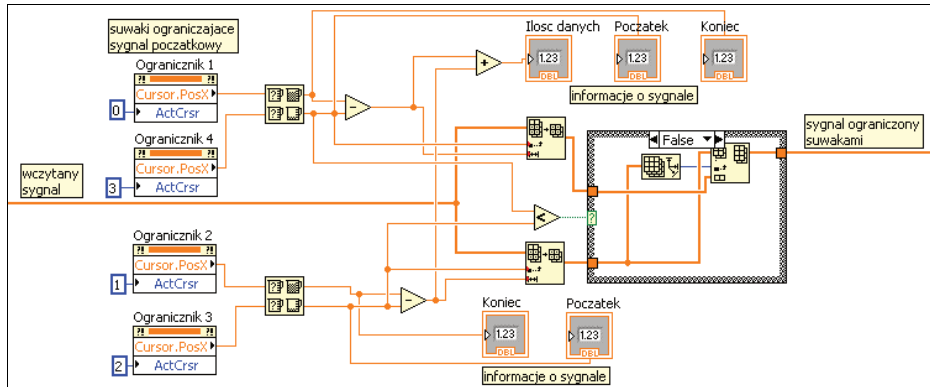
Fig. 4. a) Function *Quick Scale*; b) Function *Array Subset*

Sygnał wyświetlany jest na wykresie, który umożliwia wybranie interesującej części początkowo wczytanego sygnału. Odbywa się to z wykorzystaniem funkcji bibliotecznej „*Array Subset*” (rysunek 4b), która oddziela zadaną tablicę danych o podanej długości, na której będą prowadzone obliczenia.

Moduł wczytania sygnału wyświetla również czas potrzebny na obliczenie współczynnika Hursta. Czas ten wyznaczany jest przez obliczenie długości trwania jednej iteracji pętli „for”, co umożliwia to funkcja „*Tick Count*”.

Moduł ograniczania sygnału pozwala na podstawową edycję wczytanego sygnału. Umożliwiają to jeden lub dwa suwaki, które odpowiednio ustawione przez użytkownika na sygnale początkowym wycinają niepotrzebną część sy-

gnału, powielają lub nakładają dwa sygnały na siebie. Kod zaimplementowanych suwaków przedstawiono na rysunku 5.



Rys. 5. Kod modułu ograniczającego sygnał

Fig. 5. Limiting signal module code

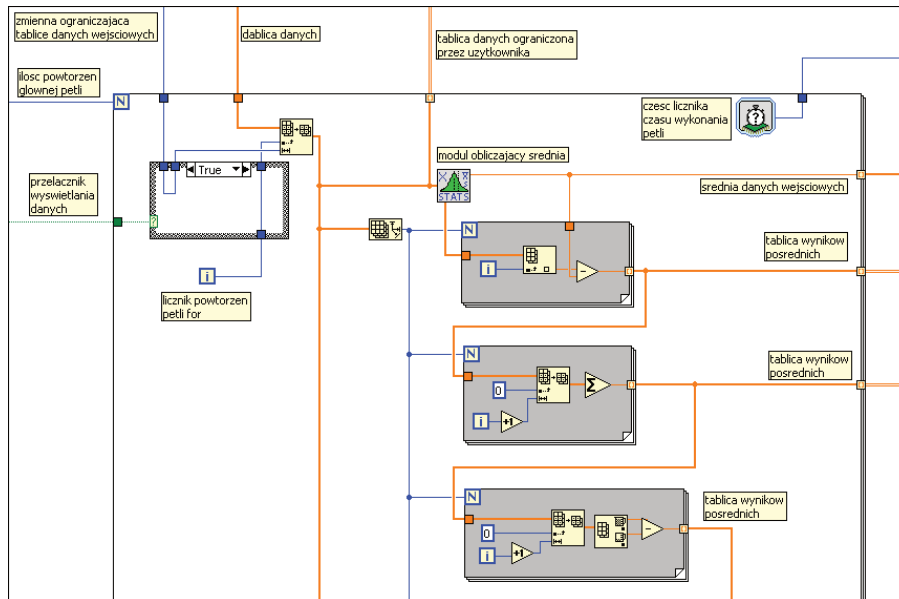
Wczytany sygnał jest dzielony na dwie tablice danych o początku i długości zadanej suwakami przez użytkownika. Następnie program sprawdza, która z dwóch nowo utworzonych tablic powinna być początkiem sygnału wynikowego wykorzystując do tego numery indeksów pozycji suwaków. W ostatnim kroku dwie nowe tablice są łączone w pętli „case” w zależności od kolejności ustalonej suwakami. Do połączenia tablic wykorzystano funkcję biblioteczną „*Inster Into Array*”. Moduł wyświetlania obliczeń przedstawia wynik w postaci wykresu, który jest zbiorem wskaźników Hursta dla każdego miejsca na wykresie danych wejściowych. Dodatkowo w module tym pokazano wartość średnią wszystkich wyznaczonych wskaźników oraz końcowy współczynnik Hursta, który uwzględnia cały zakres danych wejściowych po uwzględnieniu poprawek naniesionych w module ograniczenia sygnału. Moduł wyświetlania obliczeń posiada także interfejs, który pozwala na wybór metody obliczeń współczynnika Hursta oraz aktywacji okna obliczeń pozwalającego na określenie ilości próbek sygnału do wyznaczenia wykładnika Hursta.

Istotną funkcjonalnością programu jest możliwość zapisu wykorzystywanych danych jak i uzyskiwanych wyników analizy. Moduł zapisu danych umożliwia zapis do pliku tekstowego oryginalnych wartości mierzonych sygnałów, jak również ich wartości po edycji w module ograniczenia sygnału. Ponadto istnieje możliwość zapisu danych z wykresów prezentujących obliczone współczynniki Hursta. Odbywa się to przy wykorzystaniu funkcji bibliotecznego „*Write To Spreadsheet File*”. Jednowymiarowa tablica danych zapisywana jest w pliku tekstowym, w którym wartości występują w oddzielnych liniach i zapisane są za pomocą systemu kodowania UTF-8. Umożliwia to łatwy eksport danych do in-

nych programów, jak również potraktowanie wyniku ponownie jako dane wejściowe (w celu dodatkowej obróbki przez moduł ograniczenia sygnału).

4.2. Moduł wyznaczania współczynnika Hursta

Moduł obliczający współczynnik Hursta to główna część programu odpowiadająca za obliczanie i wyświetlenie współczynnika Hursta na wykresie dla wczytanego sygnału. Jego znaczna część znajduje się w pętli „for”, której główna zmienna odpowiada za przeskok okna danych o długości określonej przez użytkownika. Dzięki temu obliczenie współczynnika Hursta odbywa się w każdym miejscu wczytanego sygnału na podstawie historii zawartej w oknie danych przesuwanym się o jedną próbkę sygnału z każdą iteracją. Część głównej pętli „for” oraz pierwsze trzy etapy wyznaczania współczynnika Hursta pokazano na rysunku 6.

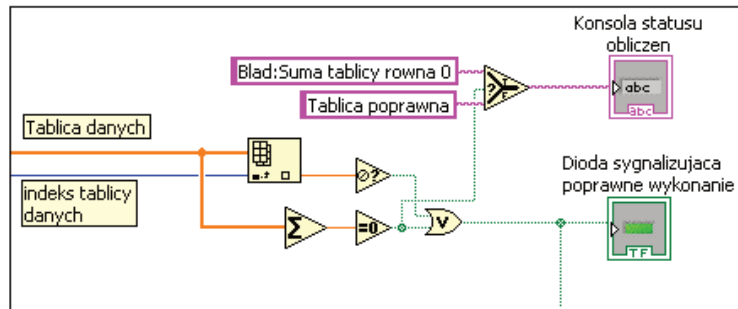


Rys. 6. Kod modułu wyznaczania współczynnika Hursta

Fig. 6. The Hurst coefficient determination module code

Moduł główny pozwala również na określenie ilości próbek sygnału zawartych w oknie ograniczającym zakres danych, na podstawie których będzie obliczany współczynnik Hursta. Ostatnią funkcją modułu jest obliczenie i wyświetlenie średniej arytmetycznej wartości współczynnika Hursta przy uwzględnieniu tylko tych miejsc na wykresie, gdzie współczynnik mógł być wyznaczony. Miejsca na wykresie gdzie współczynniki Hursta są wyznaczone niepoprawnie (ze względu na niewystarczającą ilość danych, lub próbę dzielenia przez zero)

nie są brane pod uwagę i nie mają wpływu na średnią. Funkcja odpowiedzialna za obliczanie średniej arytmetycznej to „*Sample statistics*”.



Rys. 7. Fragment kodu sprawdzającego warunki obliczenia współczynnika Hursta dla metody statystyki R/S

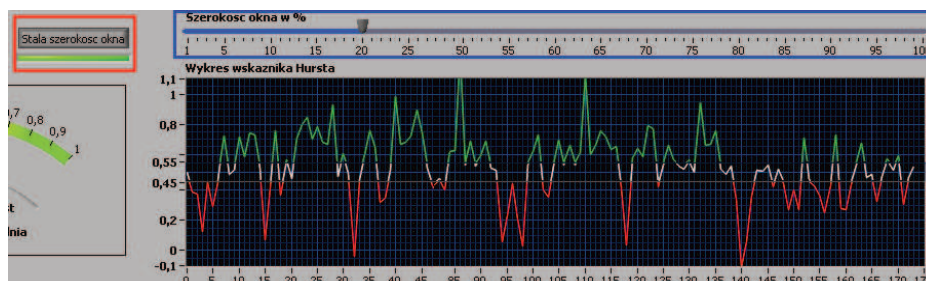
Fig. 7. The part of the program code checking the calculation conditions for the Hurst coefficient obtained with the method of R/S statistics

Istotną częścią modułu głównego jest kod pozwalający na podglądu obliczeń, a w szczególności na wyświetlenie poszczególnych kroków, które są podejmowane w celu obliczenia współczynnika Hursta. Moduł jest podzielony na szereg części, które reprezentują poszczególne kroki wymagane do obliczenia współczynnika Hursta. Każda z części wyświetla wyniki częściowe zarówno w postaci danych liczbowych jak i wykresu. Moduł ten zawiera także zbiór warunków, które sprawdzają na każdym etapie obliczeń czy współczynnik może być wyznaczony. Pozwala to na szybkie określenie miejsca i powodu nieprawidłowego wykonania algorytmu. Przykładowy kod sprawdzający warunki obliczenia współczynnika pokazano na rysunku 7.

4.3. Interfejs aplikacji i przykładowe uruchomienie

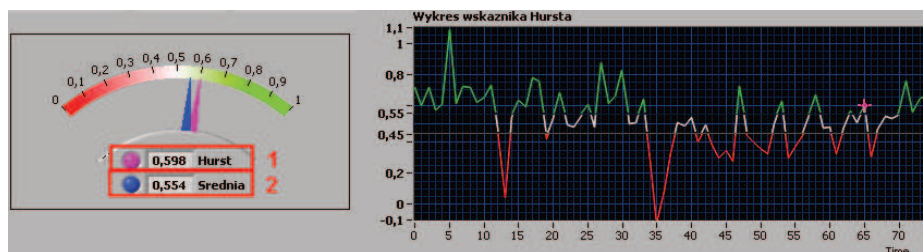
W programie współczynnik Hursta jest domyślnie obliczany przy uwzględnieniu całego zakresu danych (określonego przez użytkownika granicami) występujących przed punktem, dla którego współczynnik jest obliczany. Oznacza to, że każda następna wartość w licznym sygnale będzie miała coraz mniejszy wpływ na wartość współczynnika Hursta ponieważ ilość próbek branych pod uwagę się zwiększa. Aby określić stałą ilość próbek sygnału branych pod uwagę podczas liczenia współczynnika Hursta, należy kliknąć na przycisk „*Stać szerokość okna*”. Spowoduje to pojawienie się suwaka, który pozwala na określenie stałej ilości danych, na podstawie których będzie obliczany współczynnik Hursta dla dowolnego miejsca na wykresie. Oznacza to, że punkty nie znajdujące się w oknie ograniczenia danych (oddalone od miejsca wyznaczenia współczynnika bardziej niż zakłada to użytkownik programu) nie będą miały wpływu na obliczenia. Używanie okna danych pozwala na zwiększenie czułości zmiany współ-

czynnika Hursta w sytuacji gdy analizowany sygnał posiada wiele próbek. Pod uwagę brane są najbardziej aktualne próbki sygnału podczas gdy starsze punkty są odrzucane. Przycisk ustawiający stałą długość okna danych (zaznaczony czerwoną ramką) i suwak, którym można określić ilość danych w oknie (zaznaczony niebieską ramką) pokazany jest na rysunku 8.



Rys. 8. Wykres wskaźnika Hursta oraz suwak zmieniający szerokość okna danych wejściowych
Fig. 8. The Hurst coefficient chart and a slider changing the width of the input data range

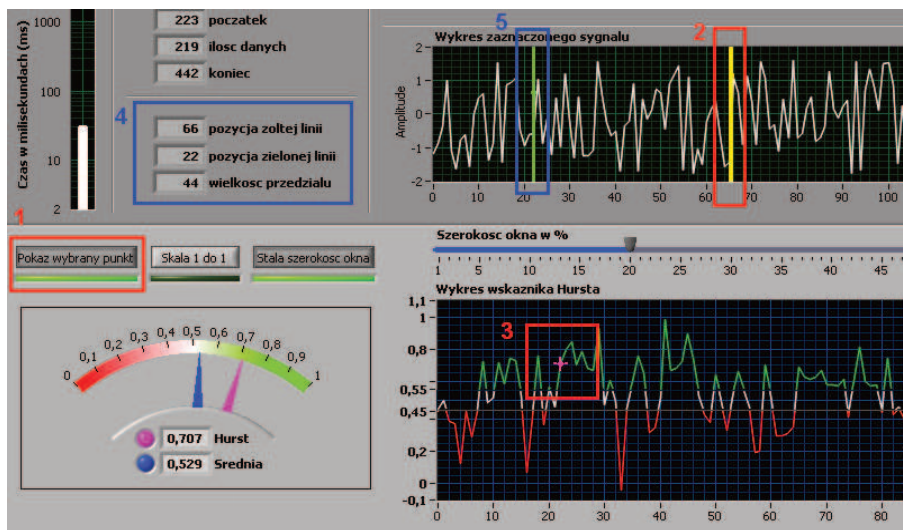
Wyniki obliczeń programu są wyświetlane w jego drugim module. Wyniki przedstawione są w postaci liczbowej (jeden współczynnik Hursta wyznaczony w miejscu wybranym przez użytkownika) oraz w postaci wykresu wyznaczonego na podstawie wszystkich miejsc w jakich współczynnik Hursta może zostać wyznaczony. Przykładowe wyniki pokazano na rysunku 9. W czerwonej ramce nr 1 zaznaczony jest współczynnik Hursta dla miejsca wyznaczonego przez użytkownika, ramka nr 2 pokazuje średnią wszystkich współczynników wyznaczonych przez program dla wybranych danych wejściowych.



Rys. 9. Wykres wskaźnika Hursta
Fig. 9. The Hurst coefficient chart

Program umożliwia śledzenie poszczególnych kroków obliczeń współczynnika Hursta dla dowolnego punktu danych pomiarowych. Gdy funkcja podglądu zostanie aktywowana na wykresie wczytanego sygnału pojawi się żółta linia pozwalająca na zaznaczenie miejsca na wykresie, dla którego będą pokazane

poszczególne kroki wyznaczenia współczynnika Hursta. Przycisk aktywujący podgląd i żółta linia umożliwiająca wybranie próbki sygnału są zaznaczone czerwonymi ramkami o numerach 1 i 2 na rysunku 10. Po wybraniu określonego miejsca na wykresie zaznaczonego sygnału użytkownik może odczytać wybraną pozycję w tabeli danych zaznaczonej na rysunku 10 niebieską ramką o numerze 4. Dodatkowo jeżeli opcja stałej szerokości okna jest aktywna na wykresie pojawia się zielona linia (zaznaczona na rysunku 10 niebieską ramką nr 5) oznaczająca początek zakresu danych, jakie brane są pod uwagę przy obliczeniu współczynnika Hursta dla wybranego miejsca na wykresie. W takiej sytuacji okno danych będzie miało szerokość równą przedziałowi ograniczonemu przez zieloną linię z lewej strony i żółtą linię z prawej strony.

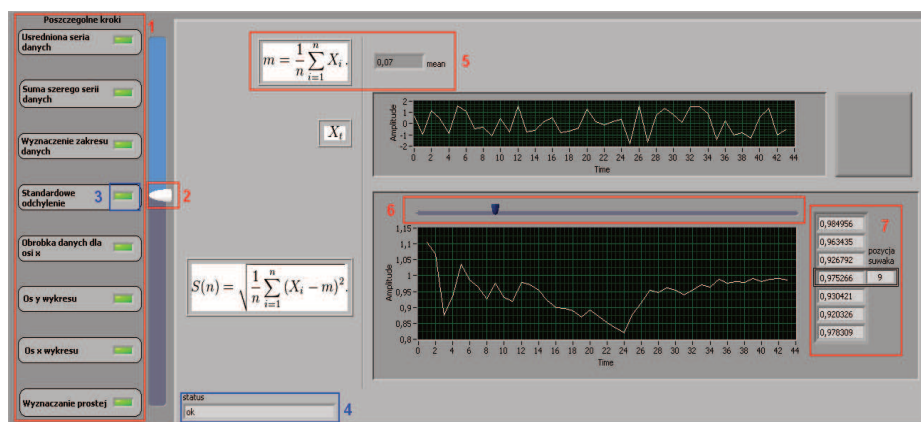


Rys. 10. Główna część interfejsu modułu analizy pojedynczego punktu

Fig. 10. The main part of the single point analysis module interface

Żółta linia umożliwia zaznaczenie dowolnego miejsca na wykresie analizowanego sygnału i odczytanie współczynnika Hursta jeżeli może zostać on wyznaczony. Wyznaczony współczynnik dla miejsca określonego żółtą linią jest zaznaczony różowym krzyżykiem na wykresie współczynników Hursta pokazanym na rysunku 10 w ramce nr 3. W celu analizy poszczególnych kroków użytych do wyznaczenia współczynnika Hursta należy aktywować przycisk „Pokaż wybrany punkt” oraz zaznaczyć żółtą linią interesujące miejsce na wykresie. Analizę poszczególnych kroków umożliwia trzeci moduł programu znajdujący pokazany na rysunku 11. Lista przekształceń matematycznych wykonywanych jeden po drugim znajduje się po lewej stronie modułu i jest zaznaczona na ry-

sunku 11 czerwoną ramką nr 1. Suwak zaznaczony czerwoną ramką nr 2 umożliwia przełączanie głównego okna modułu pomiędzy poszczególnymi przekształceniami i podgląd dowolnego z nich. Przy suwaku znajduje się osiem diod przypisanych pojedynczo do każdego z ośmiu przekształceń matematycznych. Przykładowa dioda zaznaczona jest niebieską ramką nr 3. Jeżeli poszczególny krok z przekształceniem matematycznym nie będzie mógł zostać obliczony dioda przypisana do tego kroku zaświeci się w kolorze czerwonym, który oznacza błąd obliczeń, w przeciwnym wypadku będzie to kolor zielony. Aby poznać powód błędu należy suwakiem zaznaczyć krok, w którym dioda sygnalizuje niepoprawność obliczeń i odczytać okienko „Status” znajdujące się w głównej części modułu i zaznaczone na rysunku 11 niebieską ramką nr 4. Najczęstsze powody wystąpienia błędu obliczeń to niewystarczająca ilość danych lub próba dzielenia przez zero.



Rys. 11. Moduł analizy pojedynczego punktu

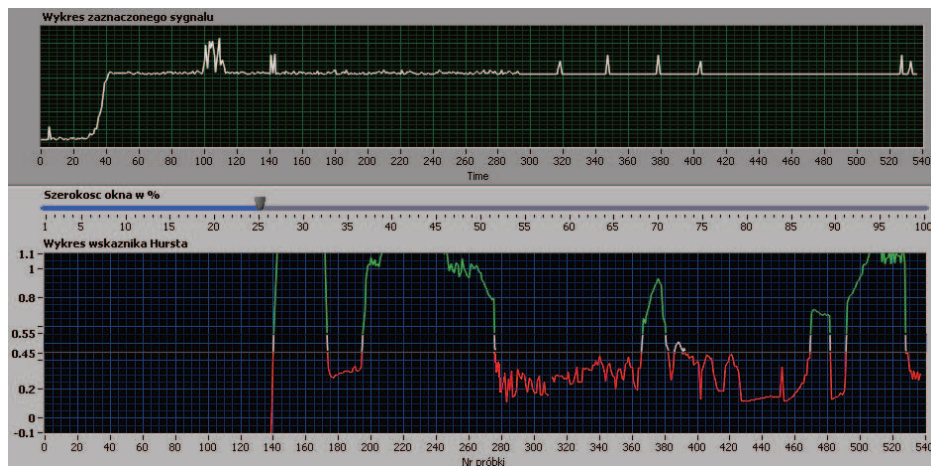
Fig. 11. The single point analysis module interface

Główna część modułu do podglądu obliczeń zmienia się w zależności od tego, które z ośmiu przekształceń matematycznych jest zaznaczone suwakiem. Ogólnie okno podzielone jest na część lewą, gdzie pokazane są wzory, na podstawie których są wykonywane obliczenia oraz na część prawą gdzie pokazane są wykresy danych na jakich prowadzone są obliczenia. Każdy wykres po prawej stronie modułu odpowiada wzorowi po jego lewej stronie. Przykładowy wzór pokazano w ramce nr 5 na rysunku 11. Aby umożliwić dokładne przeanalizowanie zmian, każde z ośmiu przekształceń posiada wykres wyniku poprzedniego przekształcenia, który teraz traktowany jest jako dane wejściowe. Wykres z wynikiem posiada dodatkowo suwak połączony z tabelą, gdzie można odczytać dokładną wartość punktu na wykresie zarówno w miejscu przesunięcia suwaka jak i trzy punkty znajdujące się przed i za suwakiem. Eliminuje to problem

precyzyjnego ustawienia suwaka na interesującym indeksie co w przypadku dużej ilości danych może być kłopotliwe. Suwak głównego wykresu i tabela z danymi jest zaznaczono czerwonymi ramkami odpowiednio o numerach 6 i 7. Ostatnie z ośmiu przekształceń wyznacza współczynnik Hursta, który następnie jest przedstawiony na wykresie wszystkich współczynników jakie mogą być wyznaczone dla danych określonych przez użytkownika, jest też zaznaczony różowym krzyżykiem (zaznaczony czerwoną ramką nr 3 na rysunku 11) do momentu wybrania przez użytkownika innego miejsca, dla którego ma być wyznaczony nowy współczynnik Hursta oraz pokazane poszczególne kroki jego wyznaczenia.

5. Wyznaczanie współczynnika Hursta dla różnych typów ruchu sieciowego

W rozdziale tym zaprezentowano wyniki obliczeń współczynnika Hursta dla danych otrzymanych z symulacji ruchu sieciowego o charakterze wideokonferencyjnym. Symulacja trwała ok. 10 minut podczas której zebrano 600 próbek danych dla dwóch parametrów: obciążenia serwera i opóźnień sieciowych. Całość symulacji przeprowadzona była w programie OPNET Modeler, a otrzymane wyniki wyeksportowano do stworzonej w środowisku LabVIEW aplikacji.

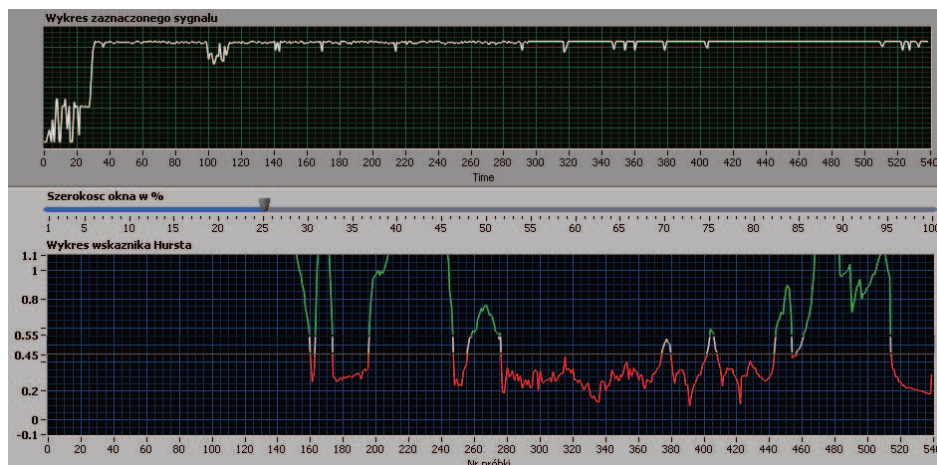


Rys. 12. Wykres ruchu sieciowego zamodelowanych usług oraz współczynnika Hursta wyznaczony metodą wartości bezwzględnej przy ograniczeniu okna pomiarowego do 25%

Fig. 12. The network traffic chart of modeled services and the Hurst coefficient determined by the absolute value method with the limited measuring window up to 25%

Ustawienie okna pomiarowego ograniczającego zakres danych do ostatnich 25% próbek całego wykresu spowodowało podwyższenie wrażliwości na

niewielkie zmiany w sygnale. Pomimo, że większość wykresu współczynnika Hursta znajduje się po stronie szeregów antypersystentnych to ekstremalne wartości samopodobieństwa znajdujące się na początku i końcu wykresu, wychodzące poza widoczny zakres 1 powodują, że średnia wartość wszystkich współczynników dla metody wartości bezwzględnej wynosi od 0,48 do 0,62 w zależności od szerokości ustawionego okna danych co zaprezentowano na rysunku 12.



Rys. 13. Wykres opóźnień sieciowych dla wideokonferencji oraz współczynnik Hursta wyznaczonego metodą wartości bezwzględnej przy ograniczeniu okna pomiarowego do 25%

Fig. 13. The videoconference network traffic delays chart and the Hurst coefficient determined by the absolute value method with the limited measuring window up to 25%

Rysunek 13 przedstawia wykres będący wynikiem obliczeń współczynnika Hursta dla danych określających opóźnienia ruchu sieciowego typu wideokonferencja. W przypadkach innych typów ruchu wartości przekraczające poziom 1,1 występują na początku i końcu badanego sygnału, a przez większą część symulacji współczynnik Hursta przyjmuje wartości bliskie 0,3. Ekstrema przekraczające poziom 1,1 na wykresie opóźnień sieciowych nie są tak wielkie jak w przypadku wykresu obciążenia sieciowego, co powoduje niższe wartości średnie samopodobieństwa, które wyznaczone są w zakresie od 0,35 do 0,56 w zależności od użytej metody i wielkości okna pomiarowego.

Podsumowanie obliczeń dla analizowanego przypadku ruchu sieciowego i mierzonych parametrów, uwzględniające wszystkie typy przeprowadzonych badań, w szczególności porównanie trzech metod wyznaczania współczynnika Hursta dla różnych wielkości okna pomiarowego pokazano w tabeli 2. Przeprowadzona w programie analiza statystyczna wskazuje, że w większości przypadków poziom samopodobieństwa ruchu sieciowego niezależnie od jego typu zawiera się w prze-

dziale od 0,5 do 1 na skali współczynnika Hursta. Można zauważyć, że współczynnik Hursta przybiera wyższe wartości wraz z rosnącym wypełnieniem pasma sieci, a nieprzerwany ruch sieciowy o małym natężeniu (np. ruch typu VoIP) posiada samopodobieństwo porównywalne do szumu białego równe 0,5.

Tabela (tablica) 2. Wartości współczynnika Hursta dla wideokonferencji wyznaczone trzema metodami dla różnych szerokości okna pomiarowego

Table 2. The videoconference Hurst coefficient values determined by the three methods with the different values of the measuring window

| | | Wartość średnia okien o szerokości: | | | Całość (100%) |
|-------------------------------------|---------------------|-------------------------------------|------|------|---------------|
| | | 15% | 20% | 25% | |
| Metoda Statystyki R/S | Aktywność Serwera | 0,38 | 0,54 | 0,46 | 0,57 |
| | Opóźnienia sieciowe | 0,35 | 0,42 | 0,42 | 0,52 |
| Metoda Zagregowanej Wariacji | Aktywność Serwera | 0,42 | 0,46 | 0,61 | 0,49 |
| | Opóźnienia sieciowe | 0,44 | 0,5 | 0,56 | 0,41 |
| Metoda Wartości Bezwzględnej | Aktywność Serwera | 0,56 | 0,48 | 0,58 | 0,62 |
| | Opóźnienia sieciowe | 0,41 | 0,47 | 0,46 | 0,56 |

Ograniczając ilość zakresu danych do obliczeń poprzez ustawienie stałej ilości próbek występującej przed miejscem obliczeń (zastosowanie okna danych) powoduje, że wskaźnik samopodobieństwa częściej przybiera wartości skrajne.

6. Podsumowanie

Celem prezentowanej pracy było zbadanie możliwości wykorzystania wydajnego klasyfikatora do oszacowania parametru Hurst w oparciu o oryginalną aplikację, który wykorzystywany będzie do pomiarów samopodobieństwa ruchu w sieci komputerowej. Otrzymane wyniki potwierdziły, że usługi VoIP ma właściwości samopodobieństwa w zakresie od 0,5 do 1. W przypadku niewielkiego ruchu osiągającym wartość wykładnika równą 0,5 (szum biały), charakteryzuje się on kompletną losowością i brakiem korelacji pomiędzy pakietami. Wartość średnia wszystkich współczynników Hursta z ograniczonym zakresem przybiera podobną wartość co współczynnik Hursta obliczony dla całego zakresu danych. Wyniki obliczeń dla poszczególnych metod zazwyczaj nie różnią się od siebie więcej niż o 15-20%. Różnice w wyniku zmniejszają się wraz ze zbliżaniem do ekstremów współczynnika. Niskie wartości współczynnika Hursta przy dużym obciążeniu sieciowym mogą wskazywać na częstą zmianę typów transmitowanego ruchu sieciowego. Można stwierdzić także, że metoda statystyki R/S posiada najmniejszą złożoność obliczeniową z trzech zaimplementowanych metod do wyznaczania współczynnika Hursta i jest najmniej podatna na spowolnienie obliczeń spowodowane przyrostem danych wejściowych.

Literatura

- [1] Thottan M., Ji Ch.: Anomaly detection in IP networks, *IEEE Transactions on signal processing*, Vol. 51, No.8, 2003, pp. 2191 – 2204.
- [2] A. J. Field, U. Harder, P. G. Harrison, Measurement and modeling of self-similar traffic in computer network, *IEE Proc. Commun.*, Vol. 151, No. 4, August 2004.
- [3] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, E. Vazquez, Anomaly-based network intrusion detection: techniques, systems and challenges, *Comput. Secur.*, 28, pp. 18–28, 2009.
- [4] H.-D. Jeong, J.-S. Lee, D. McNickle, K. Pawlikowski, Self-similar properties of malicious teletraffic, *Int. J. Comput. Syst. Sci. Eng.*, 28 (1), pp. 1–7, 2012.
- [6] P. Dymora, M. Mazurek, D. Strzałka, Computer network traffic analysis with the use of statistical self-similarity factor, *Annales UMCS Informatica* Vol.13 (2), 69–81, 2013.
- [7] M. Mazurek, P. Dymora, 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.
- [8] P. Dymora, M. Mazurek, K. Żelazny, 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.
- [9] P. Dymora, M. Mazurek, 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.

STATISTICAL ANALYSIS OF THE COMPUTER NETWORK OPERATION IN THE LABVIEW ENVIRONMENT

Summary

Network traffic analysis and the network devices working anomaly detection methods is an interesting problem for analysts researching computer networks. Proper interpretation of the anomalies and appropriate response to it can improve the quality of the network, to prevent a failure or shorten. The paper presents an original application created in the LabVIEW environment, with implemented algorithms to determine the Hurst coefficient, which is a measure of self-similarity and determine the long-range dependencies and multifractal traffic. The aim of the application was to implement known methods of determining the Hurst coefficient, e.g. the R/S statistics method, the absolute value method and the aggregate variance method, as a statistical apparatus to determine the characteristics of network traffic. The study used a virtual test network which model was created in the OPNET Modeler environment. Carried out in the application the statistical analysis indicated that the level of network traffic self-similarity is in the range from 0.5 to 1, and it's value becomes higher with increasing fulfillment of the network bandwidth. Uninterrupted network traffic with a low intensity (e.g. VoIP traffic type) has a self-similarity comparable to the white noise equal to 0.5 which is presented in the article.

Keywords: Hurst coefficient, computer networks traffic anomaly detection, self-similarity, long-range dependencies, complex systems

DOI: 10.7862/re.2015.24

Tekst złożono w redakcji: luty 2015

Przyjęto do druku: kwiecień 2015