STUDIA INFORMATICA 2005
Volume 26 Number 3 (64)

Arkadiusz BIERNACKI
Politechnika Śląska, Instytut Informatyki

ANALIZA PRZESTRZENI FAZOWEJ NATĘŻENIA RUCHU RAMEK W SIECI ETHERNET

Streszczenie. W niniejszej pracy zrekonstruowano przestrzeń fazową szeregu czasowego utworzonego z pomiarów natężeń ruchu ramek w dziesięciogigabajtowej sieci Ethernet. Wykorzystane do tego celu zostały elementy teorii systemów dynamicznych. Oszacowano opóźnienie oraz wymiar zanurzenia dla szeregu czasowego. W odtworzonej przestrzeni fazowej odnaleziono atraktory. Za pomocą metody Analizy Komponentów Głównych zredukowano wymiar atraktorów i wykreślono je w dwuwymiarowym układzie współrzędnych.

Słowa kluczowe: modelowanie natężenia ruchu sieciowego, system dynamiczny, atraktor

PHASE SPACE ANALYSIS OF FRAMES TRAFFIC INTENSITY IN AN ETHERNET NETWORK

Summary. In this paper, a phase space of time series, which was constructed from measurement of frames traffic intensity in a ten-gigabyte Ethernet network, was reconstructed. For this purpose elements of dynamic systems theory were used. An embedding delay and embedding dimension for time series were estimated. In the reconstructed phase space attractors were found. Using Principle Component Analysis, dimension of the attractors was reduced and the attractors were drawn in a two-dimensional coordinate system.

Keywords: network traffic intensity modelling, dynamic system, attractor

1. Wprowadzenie

Badania natężenia ruchu danych w sieci Ethernet przeprowadzone w laboratorium Bellcore [11] wykazały, że jest ono daleko bardziej skomplikowane od przyjętego powszechnie modelu Poissona. Wyniki tych badań zintensyfikowały prace nad poszukiwaniem nowych

modeli natężeń ruchu danych w sieciach pakietowych. Badano dynamikę natężeń ruchu danych, w tym także dynamikę ruchu pakietów, nie tylko dla protokołu Ethernet, ale również dla protokołów wyższych warstw modelu ISO/OSI. Wynikiem tych badań było opracowanie wielu różnorodnych modeli natężeń ruchu danych, na temat których przeglądowe informacje można znaleźć m.in. w [1, 4].

Jednym ze sposobów wyjaśnienia wielkości natężenia ruchu sieciowego jest próba jego modelowania przy wykorzystaniu do tego celu teorii systemów dynamicznych. Zakłada się, że wartości natężenia ruchu danych w pewnych okresach czasu są wartościami wyjściowymi deterministycznego systemu dynamicznego. Jednym z pierwszych zastosowań tej teorii do wyjaśnienia wielkości natężeń ruchu danych w sieci komputerowej był model pracy pojedynczej stacji roboczej emitującej dane. Czasy przebywania stacji emitującej pakiety, będącej w stanie włączonym lub wyłączonym, określone były poprzez deterministyczną mapę chaotyczną [13]. Szereg czasowy składający się z sum pakietów, obliczonych w kolejnych odstępach czasu, generowanych przez pewną ilość takich stacji, był bardziej zbliżony do wartości rzeczywistych natężeń ruchu w porównaniu z szeregiem czasowym stworzonym przez wykorzystanie tradycyjnego modelu Poissona. Mapy chaotyczne stosowano także w [6] do modelowania wielkości okna w protokole TCP. Symulacja komputerowa przeprowadzona przez Veresa w [18] ujawniła, iż zależności pomiędzy wielkościami okien przeciążeniowych dwóch strumieni danych TCP, dzielących wspólne łącze, mogą tworzyć deterministyczny ciąg wartości. Badania Veresa kontynuowano w [7], postawiono hipotezę, że również wielkości strat pakietów w połączeniu używającym protokołu TCP podlegają deterministycznej dynamice nieliniowej. W [19] zrekonstruowano i zbadano przestrzeń fazową strumieni danych pochodzących z protokołów TCP i UPD. Badania prowadzono dla danych zagregowanych w skalach czasu od 2 [ms] do 400 [ms]. W zrekonstruowanej przestrzeni fazowej stwierdzono występowanie atraktorów przyjmujących różne kształty geometryczne. Badania prowadzone były na zbiorach danych zebranych w roku 1995, pochodzących z archiwum ITA [16].

Celem przeprowadzonych badań, opisanych w niniejszej pracy, było określenie czy atraktory można znaleźć również w szeregu czasowym pochodzącym z pomiarów natężeń ruchu ramek w sieci Ethernet. Ponieważ odpowiedź na powyższe pytanie jest pozytywna, dodatkowo zbadano, jak zmienia się geometryczny kształt atraktora wraz z upływem czasu. Niniejsza praca korzysta z danych zebranych w roku 2004. Wybór danych pochodzących z aktualnych pomiarów jest istotny dla przeprowadzanych eksperymentów. Według badań opublikowanych w [9] w ciągu ostatnich piętnastu lat statystyczne właściwości ruchu danych w sieci Internet podlegają nieustannym zmianom.

Dalsza część niniejszej pracy zorganizowana została następująco. W rozdziale 2 omówiono metody rekonstrukcji przestrzeni fazowej szeregu czasowego. Opis rezultatów zasto-

sowania przedstawionych metod do analizy natężenia ruchu ramek w sieci komputerowej znajduje się w rozdziale 3. Podsumowanie pracy stanowi rozdział 4.

2. Rekonstrukcja przestrzeni fazowej szeregu czasowego

Jedną z metod nieliniowej analizy szeregów czasowych jest rekonstrukcja ich przestrzeni fazowej. Rekonstrukcja taka ma na celu ustalenie, czy szereg czasowy jest generowany przez deterministyczny system dynamiczny.

Ewolucja deterministycznego systemu dynamicznego w przestrzeni R^m może być opisana poprzez równanie różnicowe $\mathbf{x_{n+1}} = f(\mathbf{x_n})$. Wektory $\mathbf{x_n}$ tworzą trajektorię systemu dynamicznego. Trajektorie mogą dążyć do nieskończoności lub pozostawać w ograniczonym obszarze. Wraz z upływem czasu trajektorie pozostające w ograniczonym obszarze mogą tworzyć pewien regularny wzór geometryczny. Wzór ten nazywany jest atraktorem systemu dynamicznego.

Jednakże dane analizowane w eksperymencie, będące wartościami natężeń ruchu ramek $\{s_n\}$, nie są wektorami, ale wartościami skalarnymi i muszą zostać zamienione na wektory. Zamiana ta jest nazywana rekonstrukcją przestrzeni fazowej. Do rekonstrukcji przestrzeni fazowej stosuje się zazwyczaj metodę opóźnień. Zakłada się, że szereg czasowy $\{s_n\}$ jest sekwencją obserwacji trajektorii tworzonych przez układ dynamiczny $\{s_n = s(\mathbf{x_n})\}$, do których została zastosowana pewna funkcja mierząca $s(\cdot)$. Następnie szereg czasowy o elementach $\{s_n\}$ jest przekształcany na postać wektorową. Wektory te tworzą nową przestrzeń, zwaną przestrzenią fazową. Przekształcenie to określone jest wzorem:

$$\mathbf{y_n} = \lfloor s_{n-(m-1)\tau}, s_{n-(m-2)\tau}, \ldots, s_n \rfloor. \tag{1}$$

Wartość τ nazywana jest opóźnieniem (ang. delay), ilość elementów m jest nazywana wymiarem zanurzenia (ang. embedding dimension). Do rekonstrukcji przestrzeni fazowej konieczne jest oszacowanie dwóch powyższych wartości. Jeżeli źródłem szeregu czasowego jest deterministyczny układ dynamiczny, rekonstrukcja przestrzeni fazowej jest krokiem na drodze do odnalezienia atraktora, który otrzymywany jest poprzez umieszczenie w m-wymiarowej przestrzeni wektorów $\{y_n\}$.

Według teorii zaproponowanej przez Takensa [15] wystarczającą wielkością wymiaru zanurzenia m do tego, aby $\{y_n\}$ odwzorowywał dynamikę systemu wraz z jego atraktorem, jest taka wartość m, aby $m > 2D^F$, gdzie D^F oznacza wymiar fraktalny atraktora w oryginalnej przestrzeni fazowej. Wybór optymalnej wartości wymiaru zanurzenia pozostaje jak na razie otwartym problemem. W niniejszej pracy wartość tę oszacowano za pomocą

metody zaproponowanej przez Cao w [2]. Do oszacowania wartości τ zastosowano metodę informacji wzajemnej. Metody te zostały opisane w dalszej części pracy.

2.1.1. Szacowanie opóźnienia

Spośród szeregu metod stosowanych do szacowania opóźnienia τ, których omówienie można znaleźć m.in. w [14], w niniejszej pracy zastosowano metodę informacji wzajemnej (MIW) (ang. time delayed mutual information) [8]. Metoda ta jest podobna do metody obliczania funkcji autokorelacji, z tym że brane pod uwagę zostają również korelacje nieliniowe. Wartość τ_0, dla której MIW przyjmuje minimum, jest optymalną wartością opóźnienia dla szeregu czasowego. MIW jest obliczana według wzoru:

$$S = -\sum_{i,j} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j}, \tag{2}$$

gdzie p_i (lub p_j) oznacza prawdopodobieństwo wystąpienia wartości i-tej (lub j-tej) w zbiorze danych. Natomiast $p_{ij}(\tau)$ oznacza łączne prawdopodobieństwo, że obie wartości i-ta i j-ta wystąpią jednocześnie. Argument τ określa przesunięcie elementów szeregu czasowego względem siebie. Jeżeli τ_0 jest zbyt małe w porównaniu ze swoją rzeczywistą wartością, wtedy kolejne elementy wektorów $\mathbf{y_n}$ będą skorelowane. Wektory będą sklasteryzowane wzdłuż pewnej linii w przestrzeni R^m. Jeżeli τ_0 będzie zbyt duże, kolejne elementy wektora $\mathbf{y_n}$ będą niezależne i powstanie tzw. chmura w przestrzeni R^m, która nie będzie posiadała deterministycznych komponentów [3]. W praktyce, ze względu na to, że zbiory składające się z wartości i i j mogą mieć dosyć dużo elementów, zbiory te dzieli się na przedziały i dla nich oblicza się informację wzajemną.

2.1.2. Szacowanie wymiaru zanurzenia

Kennel [10] zaproponował metodę fałszywych najbliższych sąsiadów (ang. false nearest neighbours) jako metodę szacowania minimalnego wymiaru zanurzenia. Zakładając, że minimalny wymiar zanurzenia dla szeregu czasowego $\{s_i\}$ wynosi m_0, to w m_0-wymiarowej przestrzeni odtworzony atraktor układu dynamicznego, którego obserwacją jest szereg $\{s_i\}$, jest dokładnym obrazem oryginalnego atraktora układu dynamicznego. Zachowane zostają w szczególności topologiczne właściwości atraktora, m.in. punkty sąsiednie (ang. nearest neighbours) należące do atraktora w oryginalnej przestrzeni fazowej, które są mapowane na punkty sąsiednie w m_0-wymiarowej przestrzeni rekonstruowanego atraktora. Jeśli oryginalny atraktor zostanie odtworzony w m-wymiarowej przestrzeni, gdzie $m < m_0$, topologiczna struktura atraktora nie zostanie zachowana. Dwa punkty należące do atraktora, które nie są

sąsiednie w m_0-wymiarowej przestrzeni, mogą stać się sąsiednie w zrekonstruowanej przestrzeni o niższym wymiarze.

Przedstawione spostrzeżenia są podstawą do oszacowania wymiaru zanurzenia dla zbioru wektorów. Obliczane w tym celu zostaje wyrażenie:

$$a(i,d) = \frac{\left\| y_i(d+1) - y_{n(i,d)}(d+1) \right\|}{\left\| y_i(d) - y_{n(i,d)}(d) \right\|} \quad i = 1, 2, \ldots, N - d\tau, \tag{3}$$

gdzie $\|\cdot\|$ oznacza pewną miarę Euklidesową, $y_i(d)$ jest i-tym zrekonstruowanym wektorem o wymiarze d, $n(i,d)$ ($1 \le n(i,d) \le N - d\tau$) jest liczbą całkowitą taką, że punkt $y_{n(i,d)}$ jest najbliższym sąsiadem $y_i(d)$ w zrekonstruowanej d-wymiarowej przestrzeni fazowej z miarą odległości $\|\cdot\|$. Jeżeli $a(i,d)$ przekracza pewną heurystycznie ustaloną wartość progową, najbliższy sąsiad punktu $y_i(d)$ jest „fałszywy". Problemem jest znalezienie odpowiedniej wartości progowej, niezależnej od wymiaru d przestrzeni i wyboru punktu $y_i(d)$. Cao [2] zaproponował obliczenie dodatkowego wyrażenia

$$E1(d) = E(d+1) / E(d), \tag{4}$$

gdzie

$$E(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} a(i,d).$$

Jeśli $E1(d)$ przestanie się zmieniać dla pewnej wartości d, gdzie $d > d_0$, oznacza to, że badane wektory mogą tworzyć w $d_0 + 1$- wymiarowej przestrzeni atraktor.

2.2. Analiza Komponentów Głównych

Wzorce występujące w wielowymiarowych danych mogą być trudne do odnalezienia oraz wizualizacji. Do redukcji rozmiaru przestrzeni stosowana jest Analiza Komponentów Głównych (AKG) (ang. Principal Components Analysis). AKG wprowadza nową bazę ortonormalnych wektorów w n-wymiarowej podprzestrzeni przestrzeni m-wymiarowej, $n \le m$, w ten sposób, aby zmaksymalizować wariancję pomiędzy zrzutowanymi danymi w nowej bazie. Innymi słowy błąd rzutowania na przestrzeń n-wymiarową jest minimalizowany dla wyznaczonych wcześniej wektorów rzutowania. Wektory te są wyznaczane jako wektory własne symetrycznej macierzy kowariancji wektorów w przestrzeni m-wymiarowej. Nową bazę stanowi te n wektorów, z którymi jest związanych n największych wartości własnych macierzy kowariancji.

AKG określona jest wzorem $\mathbf{u} = \mathbf{W}^T\mathbf{y}$, gdzie $\mathbf{u}$ jest rzutem danych na n-wymiarową podprzestrzeń, $\mathbf{y}$ - wektorem o wymiarze m o współrzędnych w oryginalnej przestrzeni, $\mathbf{W}$ jest macierzą wektorów własnych macierzy kowariancji o wymiarach $m \times n$.

W niniejszej pracy AKG została wykorzystana do wizualizacji wielowymiarowych atraktorów. Ponieważ AKG jest transformacją liniową, nie zmienia ona własności topologicznych uzyskanego atraktora.

3. Analiza natężenia ruchu ramek w Ethernecie

3.1. Opis eksperymentu

Dane wykorzystane w eksperymencie są publicznie dostępne i pochodzą z National Laboratory for Applied Network Research [12]. Wybrano archiwum Teragrid-I, będące zapisem natężeń ruchu ramek w 10GB sieci Ethernet. Dane znajdujące się w archiwum pochodzą z pomiarów wykonanych za pomocą pary specjalnych kart sieciowych DAG 6.1, dedykowanych do pomiarów wartości natężeń ruchu pakietów lub ramek w sieci komputerowej [5]. Do przeprowadzenia eksperymentów wykorzystano darmowy pakiet narzędziowy TSTOOL [17] przeznaczony do środowiska Matlab, służący do nieliniowej analizy szeregów czasowych.

Dane badano w skalach czasowych w zakresie od 1 [ms] do 70 [ms]. Dla każdej ze skal czasowych ze zbioru danych wybrano 20 podzbiorów o rozmiarach od 700 do 1200 elementów. W przypadku skal czasowych mniejszych od 10 [ms] w szeregach czasowych występowało wiele wartości zerowych. Wybrano te podzbiory, w których stosunek ilości zer do rozmiaru danych nie przekraczał 5%. Dla uzyskanych tą drogą szeregów czasowych:

1. Obliczono opóźnienia metodą informacji wzajemnej według wzoru (2).
2. Obliczono wymiar zanurzenia metodą Cao według wzorów (3) i (4).
3. Zrekonstruowano przestrzeń fazową korzystając z rezultatów powyższych dwóch metod według wzoru (1).

Spośród zrekonstruowanych przestrzeni fazowych wyodrębniono te, w których dało się odnaleźć atraktory. Przeanalizowano komponenty główne odnalezionych atraktorów, a następnie wykreślono je w dwuwymiarowym układzie współrzędnych.

3.2. Wyniki eksperymentu

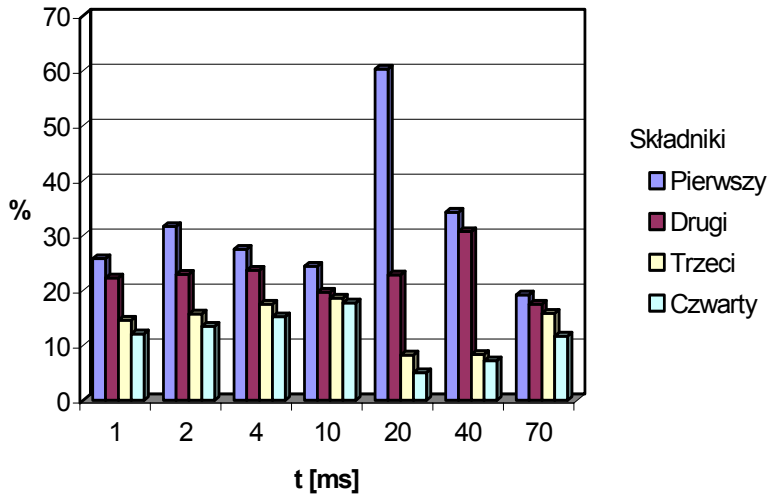
W 45 procentach zbiorów danych poddanych eksperymentowi udało się odnaleźć atraktory, co świadczy o deterministycznym charakterze szeregu czasowego utworzonego z pomiarów natężeń ruchu ramek w sieci Ethernet.

Czasy opóźnienia dla szeregu czasowego tworzącego atraktor oraz wymiary zanurzenia tego szeregu zmieniały się wraz z czasem oraz skalą pomiarów. W tabeli 1 przedstawiono uśrednione wartości powyższych atrybutów dla poszczególnych skal czasowych. Średnia wartość opóźnienia kształtowała się na poziomie 4-5 wyrazów szeregu, zwiększając się wraz ze wzrostem skali do 8 wyrazów. Średni wymiar zanurzenia zawierał się w przedziale od 5 do 9 wyrazów i nie wydawał się być skorelowany liniowo ze skalą czasową dokonywanych pomiarów.

Tabela 1

Wartości opóźnienia oraz wymiaru zanurzenia dla natężenia ruchu ramek mierzonego w różnych skalach czasowych

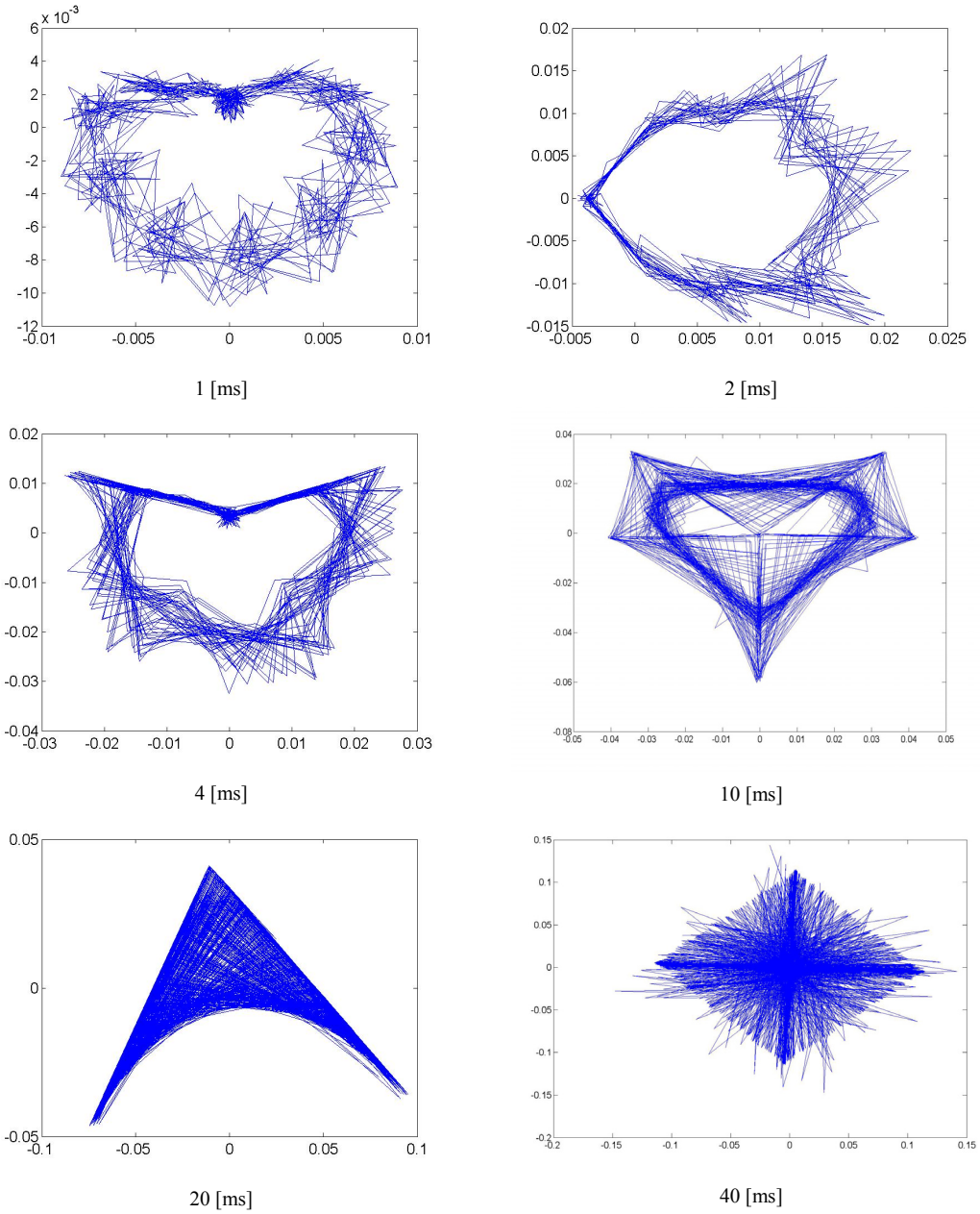
	Skala czasowa [ms]						
	1	2	4	10	20	40	70
Opóźnienie	4	4	5	4	4	8	8
Wymiar zanurzenia	8	5	5	8	7	9	8



Rys. 1. Analiza pierwszych czterech komponentów głównych dla natężeń ruchu ramek mierzonych w różnych skalach czasowych

Fig. 1. Analysis of first four main component for frames traffic intensity measured at different time scales

Analiza Komponentów Głównych wektorów, z których zbudowane były atraktory, wskazywała, że większość informacji zawartych było w czterech zmiennych stanu układu dynamicznego tworzącego atraktor, z wyjątkiem skal 20 [ms] i 40 [ms], gdzie większość informacji zawarta była w dwóch zmiennych (rys. 1).



Rys. 2. Atraktory dla natężeń ruchu ramek mierzonych w różnych skalach czasowych
Fig. 2. Attractors for frames traffic intensity measured at different time scales

Na podstawie informacji zawartych w dwóch głównych komponentach wykreślono atraktory dla wybranych skal czasowych (rys. 2). Kształt atraktora zmieniał się wraz z czasem oraz skalą pomiarów.

Wraz ze wzrostem ziarnistości skali czasu zaobserwowano również, że:

- okresy, kiedy natężenie ruchu ramek nie podlegało żadnym wzorcom, nasilały się,
- okresy, kiedy natężenie ruchu ramek tworzyło atraktor, skracały się, stąd też zbiory danych były odpowiednio mniej liczne dla wyższych skal,
- nasilała się niestacjonarność natężenia ruchu ramek.

4. Podsumowanie

Wyniki uzyskanych badań wskazują, że natężenie ruchu danych w 10GB sieci Ethernet w pewnych okresach czasu i w różnych skalach czasowych podlega prawom dynamiki nieliniowej. Ciągi wartości natężeń ruchu ramek mogą tworzyć atraktory. W niniejszej pracy atraktory te zostały odkryte poprzez rekonstrukcję przestrzeni fazowej szeregów czasowych utworzonych z natężeń ruchu ramek. W celu wizualizacji odnalezionych w zrekonstruowanej przestrzeni fazowej atraktorów ich wymiar został zredukowany za pomocą Analizy Komponentów Głównych. Dla różnych skal czasowych oraz różnych podzbiorów dla tej samej skali czasowej atraktory te przyjmowały różne kształty geometryczne.

LITERATURA

1. Abry P., Baraniuk R., Flandrin P., Riedi R., Veitch D.: Multiscale Nature of Network Traffic, Signal Processing Magazine, 2002.
2. Cao L.: Practical method for determining the minimum embedding dimension of a scalar time series, Physica D, 1997, pp. 43–50.
3. Casdagli M., Eubank S., Farmer J. D., Gibson J.: State space reconstruction in the presence of noise, Physica D, No 51, 1991, pp. 52-98.
4. Dua A.: Modeling of Network Traffic. Technical Report. http://www.stanford.edu/~dua/files/network_model.html
5. Endance Measurement Systems. http://www.endace.com.
6. Erramilli A., Broughan, M. Veitch D., Willinger W.: Selfsimilar traffic and network dynamics, Proceedings of the IEEE, 2002.
7. Fekete A., Marodi M., Vattay G.: On the Prospects of Chaos Aware Traffic Modeling, ArXiv Condensed Matter e-prints, 2002.
8. Hegger R., Kantz H.: Practical implementation of nonlinear time series methods. The TISEAN software package, http://www.mpiipks-dresden.mpg.de/~tisean. Online documentation, 1998.
9. Karagiannis T., Molle M., Faloutsos M, Broido A.: A Nonstationary Poisson View of Internet Traffic. IEEE INFOCOM, 2004.
10. Kennel M. B., Brown R., Abarbanel H.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A 45, 3403, 1992.
11. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V.: On the Self- Similar Nature of Ethernet Traffic. ACM SIGComm, San Francisco, CA, USA, 1993.

12. National Laboratory For Applied Networking Research. Nlanr network analysis infrastructure. http://moat.nlanr.net. NLANR PMA and AMP datasets are provided by the National Laboratory for Applied Networking Research under NSF Cooperative Agreement ANI-9807579.
13. Pruthi P.: An Application of Chaotic Maps to Packet Traffic Modeling. Ph.D. Dissertation, Royal Institute of Technology, Stockholm 1995.
14. Rosenstein M. T., Collins J. J., DeLuca C. J.: Reconstruction expansion as a geometry-based framework for choosing proper delay times, Physica D, No 73, pp. 82-98.
15. Takens F.: Detecting Strange Attractors in Turbulence. Lecture Notes in Math., Vol. 898, Springer, 1981.
16. The Internet traffic archive, http://www.acm.org/sigcomm/ITA/, sponsored by ACM SIGCOMM.
17. The nonlinear time series analysis tool package TSTOOL and its documents, http://www.physik3.gwdg.de/tstool/index.html, 2001.
18. Veres A., Boda M.: The chaotic nature of TCP congestion control, Proc. IEEE INFOCOM 2000, pp. 1715–1723.
19. Zhang W. a.o: Chaotic Network Attractor in Packet Traffic Series. Comput. Phys. Commun, Vol. 161, No 3, 2004, pp.129-142.

Recenzent: Dr inż. Arkadiusz Sochan

Wpłynęło do Redakcji 7 czerwca 2005 r.

Abstract

In this paper, a phase space of time series, which was constructed from measurement of frames traffic intensity in a ten-gigabyte Ethernet network, was reconstructed. For this purpose elements of dynamic systems theory were used. Optimal parameters for embedding delay and embedding dimension for the time series were obtained respectively by time delayed mutual information and Cao method based on false nearest neighbours. These parameters were changing with time and scale of measurement. An average time delay of the time series was about 4 to 5 elements, an average embedding dimension of the time series was ranging from 5 to 9 elements (Table 1). From reconstructed phase space, attractors were extracted. For visualization purpose, a dimension of the attractors was reduced by Principal Component Analysis (PCA) method. The weight of each dimension showed no dominant value at different time scales after PCA transformation (Fig. 1). In most cases the majority of information

was held in four state spaces variables of dynamic system, which created the attractor. Reconstructed attractors were drawn in two-dimensional coordinates system (Fig. 2). Further investigation showed that the attractor can exhibit different geometric shapes at different time scales. Also with increase of time scale it was harder to find attractors partly due to increasing non-stationarity in frames traffic.

Adres

Arkadiusz BIERNACKI: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16, 44-101 Gliwice, Polska, abiernacki@star.iinf.polsl.gliwice.pl .