E11-2003-148

I. Antoniou[1,2], V. V. Ivanov[1,3], Valery V. Ivanov[3,4],
P. V. Zrelov[3]

PRINCIPAL COMPONENT ANALYSIS OF NETWORK TRAFFIC MEASUREMENTS

Reported at the International Seminar
«Advanced Computing and Analysis Technics in Physics Research»,
June 24–28, 2002, Moscow, Russia

---

[1]International Solvay Institutes for Physics and Chemistry, CP-231, ULB, Bd. du Triomphe, 1050, Brussels, Belgium
[2]Department of Mathematics, Aristoteles University of Thessaloniki, 54006 Thessaloniki, Greece
[3]Laboratory of Information Technologies, Joint Institute for Nuclear Research, 141980, Dubna, Russia
[4]University Scientific Center, Joint Institute for Nuclear Research, 141980, Dubna, Russia

Introduction

In [3] we applied a nonlinear analysis technigue [4] to the traffic measurements obtained at the input of the intermediate size Local Area Network (LAN). We demonstrated that this approach can be successfully used for deeper understanding of main features of the traffic data. At the same time, we found that due to a very complicated character of traffic series the traditional algorithms of nonlinear analysis do not give reliable estimations of the analyzed time series. For instance, the Grassberger-Procaccia algorithm gives a very high dimension for original traffic measurements. However, after filtering out a high frequency component, which can be considered as noise, we obtained a more realistic result for the embedding dimension of the underlying process. This result has been confirmed independently by the Principal Component Analysis (PCA) [3] in frames of the "Caterpillar"-Singular Spectrum Analysis (SSA) scheme [1, 2].

The PCA is a well-known technique in multivariate data analysis [5]-[9], which consists in applying a linear transformation to the original data space into a *feature space*, where the data set may be represented by a reduced number of "effective" features and yet retain most of the intrinsic information content of the data. The "Caterpillar"-SSA approach is a novel scheme very efficient for analysis of time series corresponding to any arbitrary signal [1, 2].

In our study we use the traffic measurements obtained at the input of Dubna University [10] LAN, which includes approximately 200-250 interconnected computers. In Section 1 we describe the data acquisition system of this LAN, realized on the basis of a standard PC. In Section 2 we present a basic concept of the "Caterpillar"-SSA scheme. In Section 3 we apply the "Caterpillar"-SSA technique to the traffic measurements and analyze the leading components responsible for the main part of the network traffic. In Section 4 we study residual components and propose a statistical method for their selection and elimination from a whole set of principal components.

1. Data acquisition system

The measurements of network traffic are realized at the external side of the input lock of LAN. The performance of the data acquisition system is based on realization of an open mode driver [11]: see Fig. 1.

In standard conditions the network adapter of a computer is in a mode of detecting a carrying signal (main harmonic 4 – 6 MHz). After appearing in the cable bits of the package preamble, the network adapter comes to a mode of 1 bit and 1 byte synchronization with the transmitter and starts receiving first bytes of the package heading. As soon as one succeeds in extracting the MAC-address of the shot receiver from the first bytes taken by the adapter, the network adapter compares it to its own. In the case of a negative result of the comparison, the network adapter ceases to record the shot's bytes into its internal buffer and cleans its contents and then waits until the

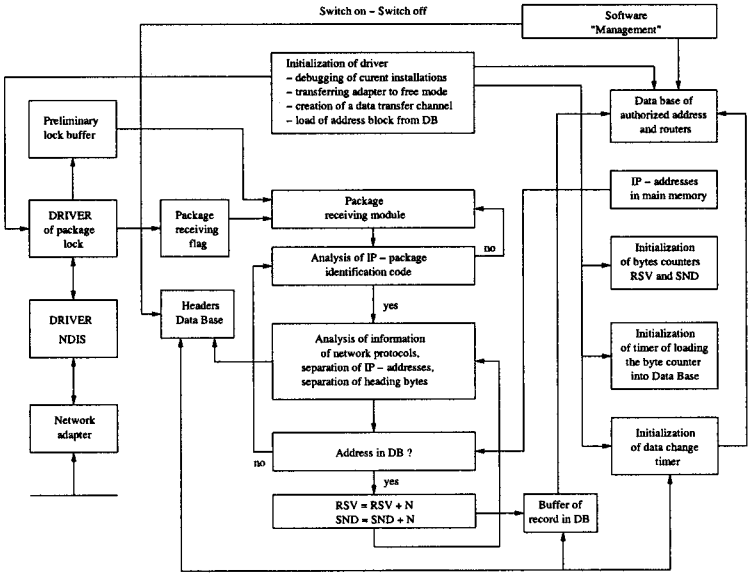

Figure 1: Scheme of a data acquisition system

next package appears.

In order to provide conditions for receiving and analysis of all the packages transmitted over the network, it is necessary to move the adapter devices to a free mode when all possible shots are recorded in the buffer. This operation is executed through the instructions of the NDIS driver.

The free mode driver records the accepted packages in the preliminary capture buffer and displays the flag of receiving the package. Then the receiving package module is activated and analysis of the margin of the package's type is carried out to extract TCP/IP packages from the whole stream.

After identification it is possible to separate and delete the data block as well as to record the headers to the SQL-server database. The recording is performed together with the time data with a frequency up to 10 kHz. Although the recording is performed with buffering, the mode of saving the packages' headers requires enormous server's resources, as in this case there is a permanent procedure of recording with small portions to the hard disk. That is why this mode is switched on if required at the management system's instruction.

The system also provides control over the external traffic of the local area network on the basis of controlling the records in the router table. Initial information on the legal IP addresses is saved in the database of the LAN computers from which data on legal addresses are loaded into the main memory array. The users which do not participate in forming the external traffic, are not taken into account when calculating the number of transferred and received bytes. In order to decrease the number of sessions of recording the information on the external traffic in the database, a timer

of load out of the buffer and a timer of changing a current date have been introduced into the system.

The recorded traffic data correspond approximately to 20 hour (1600000 records with a frequency up to 10 kHz, which corresponds to $1\,ms$ bin size) measurements. The part of this series corresponding approximately to 1 hour measurements and aggregated with different bin sizes is presented in Fig. 2. Two protocols are used in the

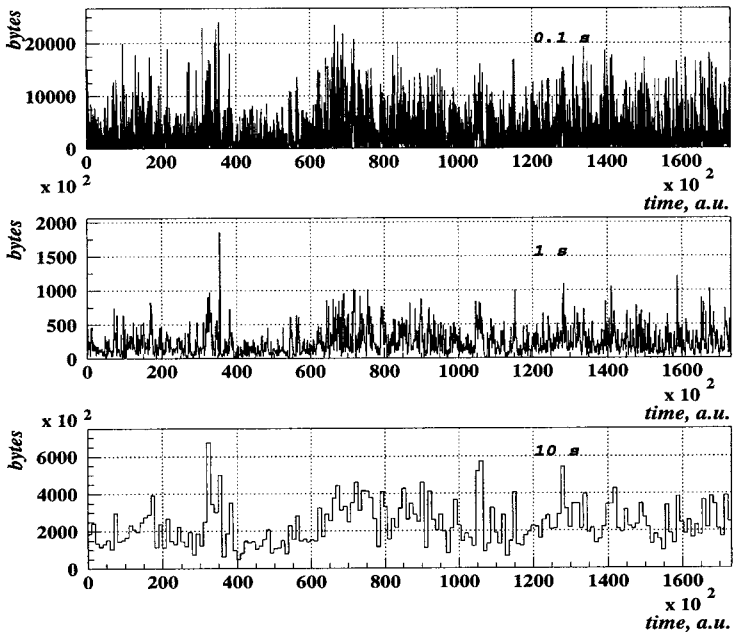

Figure 2: Traffic measurements aggregated with different bin sizes: $0.1\,s$, $1\,s$ and $10\,s$

"Dubna" LAN. The NetBEUI protocol is applied only for internal exchanges, and the TCP/IP – for external communications. The contribution of the NetBEUI traffic is estimated around 1-6 packages per second during daily working hours, which is negligibly small compared to the TCP/IP traffic. In this connection, we may neglect the influence of non-IP traffic on the TCP/IP traffic.

2. Basic concept of the "Caterpillar"-SSA technique

The "Caterpillar"-SSA approach [1, 2] is applied to the analysis of time series corresponding to any arbitrary signal $f(t)$, $t > 0$ determined in equidistant points. The basic "Caterpillar"-SSA scheme includes four main steps:

1. transformation of one-dimensional series into a multidimensional form,

2. singular value decomposition of the multidimensional series,

3. principal components analysis and selection of feature components,

4. reconstruction of one-dimensional series using the selected components.

The transformation of one-dimensional series

$$x_i = f(t_i) = f[(i-1)\Delta t], \quad i = 1, 2, \ldots, K \tag{1}$$

into a multidimensional series is realized by representing (1) in matrix form:

$$X = (x_{ij})_{i,j=1}^{k,L} = \begin{pmatrix} x_1 & x_2 & x_3 & \ldots & x_L \\ x_2 & x_3 & x_4 & \ldots & x_{L+1} \\ x_3 & x_4 & x_5 & \ldots & x_{L+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_k & x_{k+1} & x_{k+2} & \ldots & x_K \end{pmatrix}, \tag{2}$$

where $L < K$ is called the caterpillar or window length and $k = K - L + 1$.

Then the eigenvalues λ_i, $i = 1, 2, \ldots, L$ and eigenvectors $\vec{V}_i$, $i = 1, 2, \ldots, L$ of the covariance matrix $C = \frac{1}{k} X X^T$ are determined. The matrix of eigenvectors V is used for transition to principal components

$$Y = V^T X = (Y_1, Y_2, \ldots, Y_L), \tag{3}$$

where Y_i $(i = 1, 2, \ldots, L)$ are rows of k elements.

The equality

$$\sum_{i=1}^{L} \frac{\lambda_i}{L} = \sum_{i=1}^{L} \alpha_i = 1$$

permits to estimate the contribution α_i (in decreasing order) of the i-th principal component into the analyzed series. This contribution can be interpreted as a fraction of information related to a single component, and it helps, together with analytical and visual analysis of eigenvectors and principal components, to select feature components for reconstruction of one-dimensional series. The selection of specific components usually depends on the goal which we pursue and the informative content of particular components (see, for example, [12, 13, 14, 1, 2]).

3. PCA of traffic measurements: analysis of leading components

The caterpillar length (or window) C_L is chosen based on the analysis of the autocorrelation function for traffic measurements [3]. In this study we used different values of C_L, starting from the minimal value $C_L = 12$ up to $C_L = 20$.

Figure 3 shows part of the daily traffic measurements aggregated with the bin size $1\,s$ used in this study.

One of main results of the application of the "Caterpillar"-SSA technique to the analyzed series is presented in Fig. 4. It shows the contribution of the eigenvalues in percentages for $C_L = 12$ and 20. This information permits to estimate the number of principal components, which effectively contribute to the analyzed series.

Taking into account [19], it is reasonable to assume that the packet size distributions, corresponding to the leading components, may be described by the log-normal

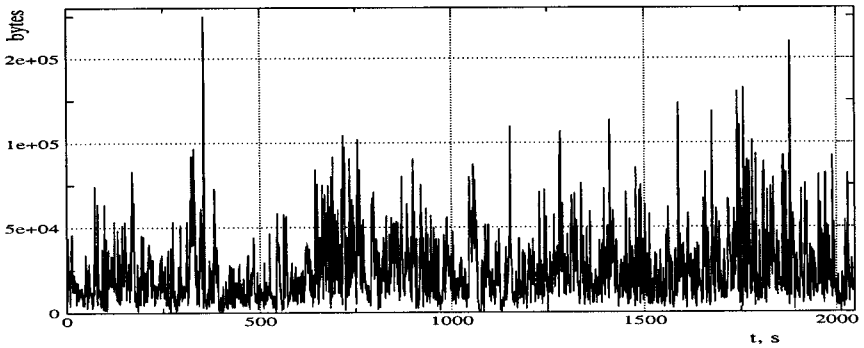

Figure 3: Traffic measurements aggregated with the bin size 1 s

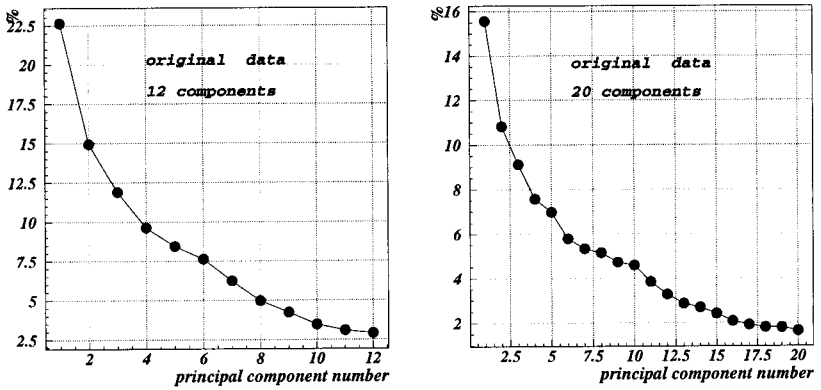

Figure 4: Contributions of eigenvalues in percentages for the original traffic data: $C_L = 12$ (left plot) and 20 (right plot)

distribution. In order to check whether these distributions follow the log-normal form, we fitted them by the log-normal function [20]:

$$f(x) = \frac{A}{\sqrt{2\pi}\sigma}\frac{1}{x}\exp\left[-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right], \tag{2}$$

where σ and μ are parameters and A is a normalizing factor. The fitting procedure was realized with the help of the MINUIT package [21] in frames of well-known PAW (Physical Analysis Workstation, see details in [22]).

We present in Table 1 the results of fitting the packet size distributions, corresponding to different number N of leading components (the results presented here are for $C_L = 20$), by function (2), and ν is the number of degrees of freedom for the χ^2-test.

Table 1: Results of fitting the packet size distributions, corresponding to the sum of N leading components, by the log-normal function (2)

N, leading comp.	σ	μ	ν	χ^2
1	0.273 ± 0.009	10.44 ± 0.01	47	87.49
2	0.304 ± 0.005	10.40 ± 0.01	44	66.82
3	0.349 ± 0.007	10.38 ± 0.01	47	53.10
4	0.377 ± 0.008	10.37 ± 0.01	47	63.52
5	0.420 ± 0.011	10.35 ± 0.01	47	68.50
6	0.432 ± 0.012	10.34 ± 0.01	46	59.12
7	0.426 ± 0.008	10.35 ± 0.01	47	49.03
8	0.444 ± 0.007	10.34 ± 0.01	47	34.39
9	0.463 ± 0.008	10.33 ± 0.01	43	38.94
10	0.482 ± 0.009	10.32 ± 0.01	47	37.76
11	0.489 ± 0.008	10.31 ± 0.01	47	55.64
12	0.500 ± 0.009	10.32 ± 0.01	47	59.00
13	0.506 ± 0.008	10.32 ± 0.01	43	51.97
15	0.518 ± 0.009	10.31 ± 0.01	46	55.16
17	0.516 ± 0.008	10.30 ± 0.01	47	78.59
19	0.513 ± 0.008	10.30 ± 0.01	44	101.6

Figure 5 shows the dependence of χ^2/ν versus N (for $C_L = 20$). Two lines parallel to the abscissa axes show the significance levels (or the probability that the observed chi-square will exceed the value χ^2 by chance even for a correct model: see, for instance, [18, 20]) $\alpha = 10\%$ (the top line, $\chi^2/\nu = 1.247$) and $\alpha = 89.5\%$ (the bottom line, $\chi^2/\nu = 0.732$) corresponding to the χ^2 test for $\nu = 47$.

This dependence demonstrates that the testing distribution does not pass the null-hypothesis (2), when only the first leading component is taken into account. Then, with the increase of N, the value of χ^2 is rapidly decreasing and for $N = 3$ one can see a quite good level of correspondence ($\alpha = 22\%$) of the distribution to the null-hypothesis (Fig. 6).

This result is of great interest because only 3 first components (of 20) already form the fundamental part of the information traffic. Figure 7 shows the seria reconstructed on the basis of the first, second and third leading component, correspondingly, after the subtraction the caterpillar average value. Figure 8 presents the dependence of the autocorrelation function

$$C(\tau) = \frac{\sum_{i=1}^{K}(x_{i+\tau} - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{K}(x_i - \bar{x})^2}, \qquad \bar{x} = \frac{1}{K}\sum_{i=1}^{K} x_i. \tag{3}$$

One can see from these figures that the autocorrelation function corresponding to the sum of 3 leading components is close to the autocorrelation function for the original data. Their summary contribution into the general dispersion is around 40 % (see Fig. 4 for $C_L = 20$).

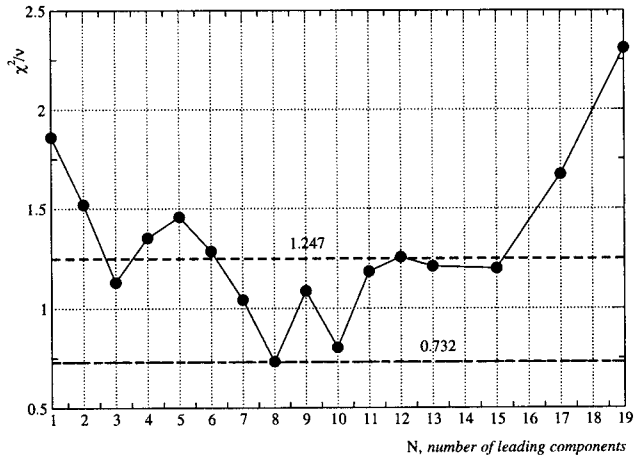

Figure 5: The dependence of χ^2/ν versus the number of leading components

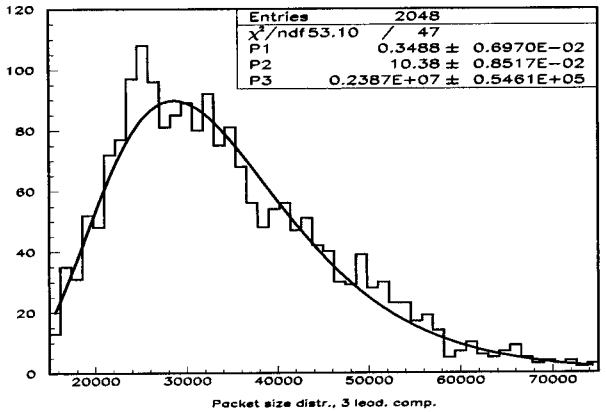

Figure 6: Fitting the distribution corresponding to 3 leading components by function (2)

This result has been confirmed for the shorter caterpillar length, $C_L = 12$. In this case only 2 leading components, their lump contribution approximately coincides with the contribution of the 3 leading components for $C_L = 20$ (see Fig. 4), reproduce the

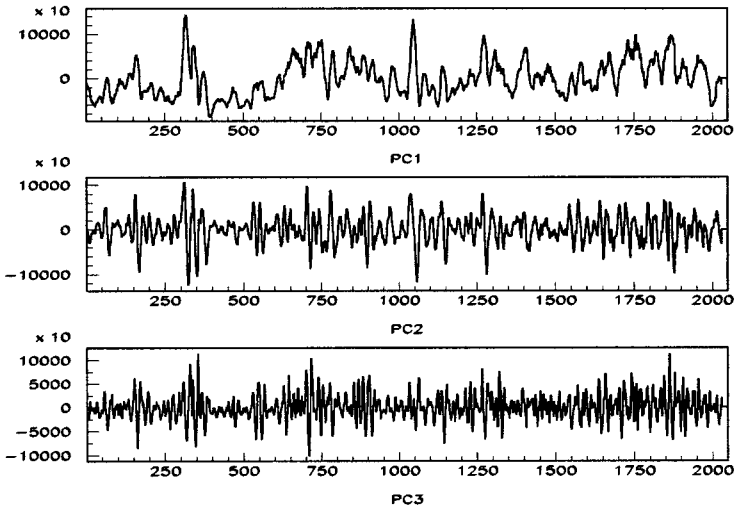

Figure 7: Time seria corresponding to 3 leading components (after subtraction of the caterpillar average value)

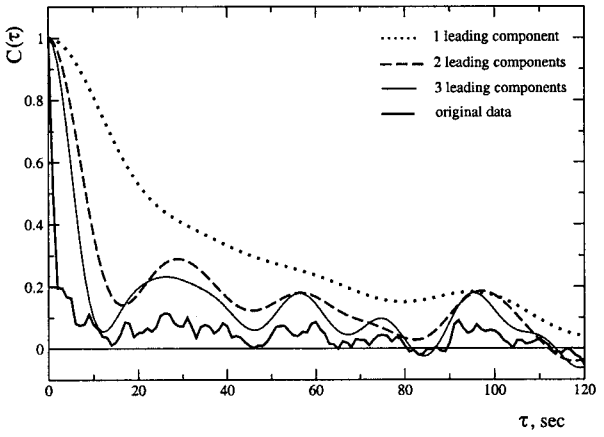

Figure 8: Autocorrelation functions $C(\tau)$ of reconstructed seria corresponding to different number of leading components

log-normal form of the traffic.

Further increase of N leads to unexpected increase of χ^2 (for $N = 4$ and 5) together with the decrease of the significance level below 10%. Then the value of χ^2/ν

rapidly decreases and reaches its record minimal value 0.732 for $N = 8$. The corresponding statistical distribution is presented in Fig. 9. It demonstrates both a very good level of correspondence of the reconstructed distribution to the null-hypothesis ($\alpha = 89.5\,\%$) and a reliable accuracy of approximation on all regions of the analyzed distribution. The summary contribution of 8 leading components into the general dispersion is around 66 %.

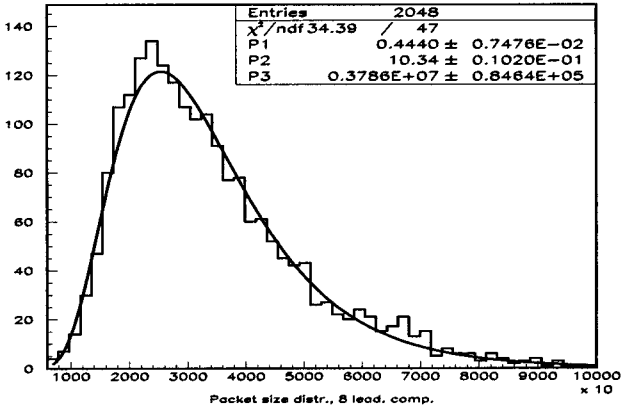

Figure 9: Fitting the distribution corresponding to 8 leading components by function (2)

Figure 10 shows the reconstructed series using the "Caterpillar"-SSA method (for $C_L = 20$) on the basis of 8 leading components. One can clearly see that it reproduces characteristic features of the original series presented in Fig. 3.

4. PCA of traffic measurements: analysis of residual components

In the region of large N there is a growth of χ^2 especially noticeable at $N \geq 15$: see Fig. 5. Such tendency may be caused by the influence of the residual components related to small irregular variations, which do not fit in the basic model of network traffic (2) and can be interpreted as stochastic noise.

Figure 11 shows the series reconstructed on the basis of the smallest residual component, namely, the component 20. One can clearly see that this series has significantly different character compared to the original traffic measurements. It looks like a non-stationary process symmetric against zero mean value.

Figure 12 shows the statistical distribution corresponding to the series presented in Fig 11. It quite well follows the Gaussian distribution that is confirmed by the χ^2-test (see Fig. 12). The autocorrelation function of the corresponding series shows that it behaves like noise.

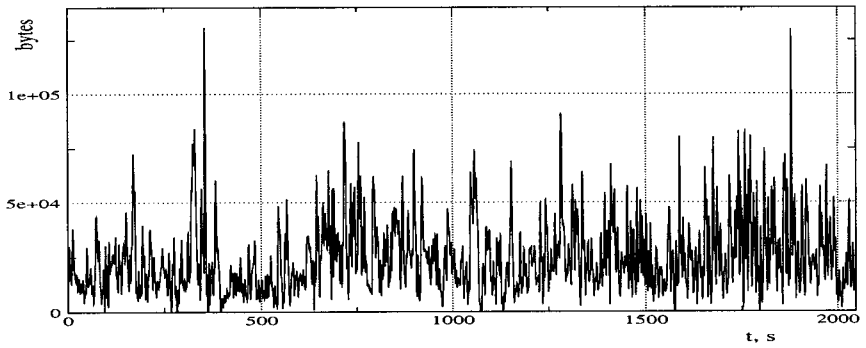

Figure 10: Traffic series measurements reconstructed by the caterpillar method (for $C_L = 20$) on the basis of 8 leading components

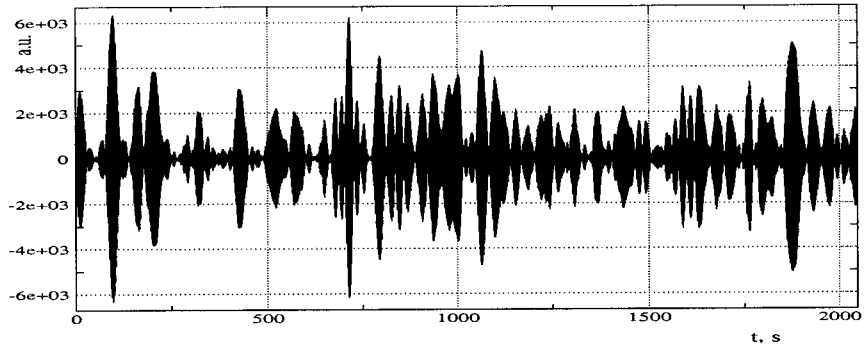

Figure 11: Traffic series reconstructed by the caterpillar method ($C_L = 20$) on the basis of the smallest component

However, when increasing the number of residual components, their summary distribution quickly starts lose the symmetric form together with growth of correlations between the series terms.

In order to estimate the amount of residual components that can be eliminated from the original time series without the influence on its fundamental part, we divide all principal components into two parts:

1. first part corresponding to the leading components and responsible for the log-normal form of the packet size distribution,

2. second part related to residual components, which is described by a symmetric statistical distribution and behaves like a stochastic noise.

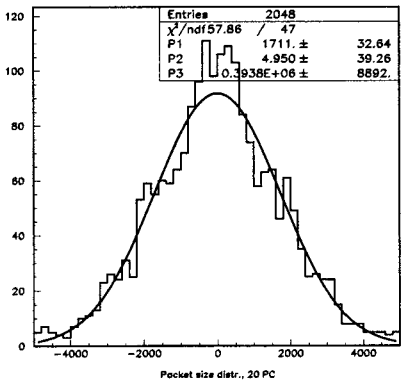

Figure 12: Statistical distribution of the time series presented in Fig. 11; the fitting curve corresponds to the Gaussian distribution

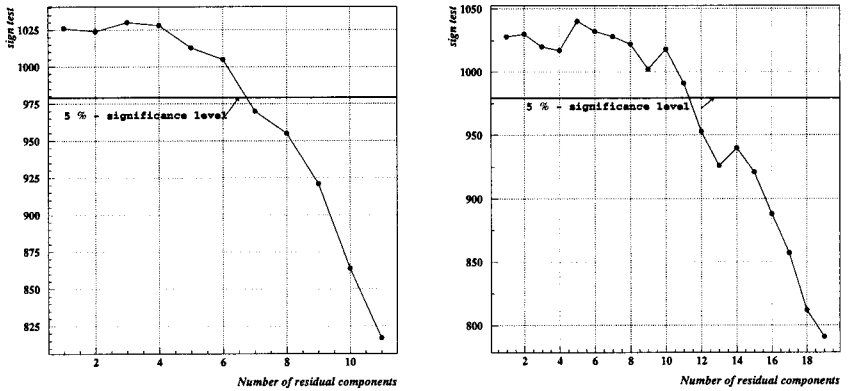

Figure 13: The values of the sign test μ versus the number of residual components for the caterpillar length $C_L = 12$ (left figure) and $C_L = 20$ (right figure)

As the criterion for selection of the second part we used the "moment" of the symmetry violation for the series corresponding to the residual components. The well-known sign test has been used for testing the symmetry against zero of residual distributions. The sign test has the following form:

$$\mu = \sum_{i=1}^{n} \Theta(X_i), \tag{4}$$

where $X_1, \ldots, X_n$ are observables, n is the sample size, and Θ is the Heaviside function:

$$\Theta(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

When the null-hypothesis is valid the, μ distribution is approximated (in case of large n) by:

$$P\{\mu \le m \mid n,p\} \approx \Phi\left(\frac{m-np+0.5}{\sqrt{np(1-p)}}\right),$$

where Φ is the distribution function of the normal distribution, $p = 0.5$ and $n = 2048$ (in our case).

Figure 13 shows the dependence of the μ value versus the number of residual components (for caterpillar lengths 12 and 20). It is clearly seen that the μ value exceeds the reliable confidential level, when the number of residual components is greater than 6 for $C_L = 12$ and 11 for $C_L = 20$.

In order to confirm the results obtained by the sign test, we applied a more powerful criterion based on the ω_n^2 statistics [25]. This criterion tests the symmetry against $x = 0$ of the distribution function $F(x)$ of observables $X_1, \ldots, X_n$, i.e. the null-hypothesis H_0: $F(x) = 1 - F(x)$. The corresponding ω_n^2 statistics has the following form:

$$\omega_n^2 = n \int_{-\infty}^{\infty} \left[F_n(x) + F_n(-x) - 1\right]^2 dF_n(x), \tag{5}$$

where $F_n(x)$ is the empirical distribution function. It is more convenient to calculate the values of the statistics (5), using the following formula

$$\omega_n^2 = \sum_{j=1}^{n} \left[F_n(-X_{(j)}) - \frac{n-j+1}{n}\right]^2,$$

where $X_{(1)} \le \ldots \le X_{(n)}$ is the variational series constructed on the basis of observables.

Figure 14 shows the dependences of ω_n^2 values versus the number of residual components for two cases of the caterpillar length: $C_L = 12$ and 20. These dependences have distinct characteristic features at $k = 4$ for $C_L = 12$, and $k = 7$ for $C_L = 20$ (one can see that the number of such components approximately equals to one third of the caterpillar length), after which, when k is increasing, there is a quick rise of ω_n^2. This rise means that the residual series loses its symmetric character, because in the second part the components responsible for the log-normality are involved.

One can see from Fig. 14 that the number of residual components $k = 6$ for $C_L = 12$ and $k = 11$ for $C_L = 20$ correspond to the 5% - significance level for the ω^2-criterion. This coincides with the result obtained for the sign test (Fig. 13). These estimates of the number of components, which do not noticeably influence the fundamental part of traffic, qualitatively coincides with the result obtained in Section 3 applying the χ^2-test (Fig. 5).

Conclusion

We applied the "Caterpillar"-SSA approach [1, 2], which is an extension of the Principal Component Analysis, to network traffic measurements, in order to understand the main features of the principal components of network traffic. Our analysis of the leading components has shown that only a few first components already form the fundamental

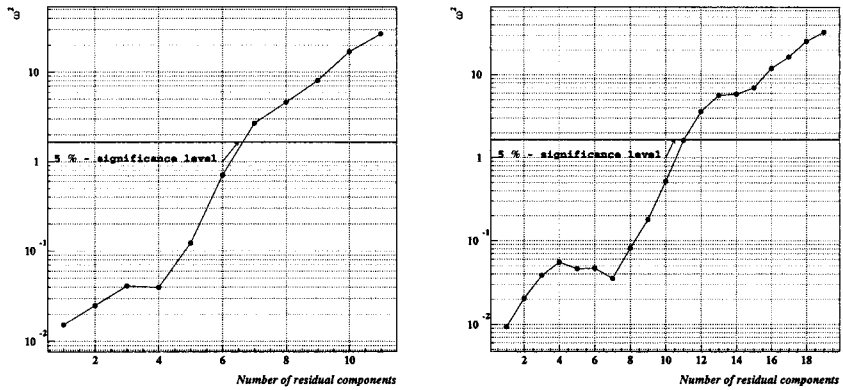

Figure 14: The dependences of ω_n^2 values versus the number of residual components for two cases of the caterpillar length: $C_L = 12$ (left figure) and $C_L = 20$ (right figure)

part of the information traffic and that the correspondence to the log-normal distribution remains valid up to intermediate values of N. In the region of large N there was found a noticeable growth of χ^2, that can be explained by the influence of residual components related to small irregular variations. Based on feature characteristics of residual components, we developed a statistical method that permits to estimate the number of components which do not play a noticeable role in the fundamental part of traffic and can be eliminated from the whole set of components.

Thus, the statistical analysis of traffic measurements based on the joint application of χ^2 and ω^2 tests gives the possibility to split the whole set of components into two classes. The first class includes the leading components responsible for the main contribution to the traffic, and the second class involves residual contributions that can be interpreted as noise. A more detailed analysis of the boundary region between these two groups may provide additional information on traffic components and, thus, simplify the understanding of traffic dynamics.

Acknowledgments

We are grateful to Prof. I. Prigogine and Prof. V. G. Kadyshevsky for encouragement and support.

This work has been partly supported by the European Commission in the frame of the Information Society Technologies (IST) program, the IMCOMP project (IST-2000-26016).

References

[1] D.L. Danilov and A.A. Zhigljavsky, Eds.: *Principal Components of Time Series: Caterpillar Method*, St. Petersburg University Press, 1997 (in Russian).

[2] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky: *Analysis of time series structure: SSA and related techniques*, Chapman & Hall/CRC, 2001.

[3] P. Akritas, P.G. Akishin, I. Antoniou, A.Yu. Bonushkina, I. Drossinos, V.V. Ivanov, Yu.L. Kalinovsky, V.V. Korenkov and P.V. Zrelov: *Nonlinear Analysis of Network Traffic*, "Chaos, Solitons & Fractals", Vol. **14(4)**(2002) pp.595-606.

[4] Henry D.I. Abarbanel: *Analysis of Observed Chaotic Data*, 1996 Springer-Verlag New York, Inc.

[5] R.W. Preizendorfer: *Principal Component Analysis in Meteorology and Oceanography.* New York: Elsevier, 1988.

[6] I.T. Jolliffe: *Principal Component Analysis*, New York, Springer-Verlag, 1986.

[7] J.E. Jackson: *A User's Guide to Principal Component Analysis.* John Wiley & Sons: New York 26-62(1992).

[8] K. Karhunen: *Uber lineare methoden in der Wahrscheinlichkeitsrechnung*, Annales Academiae Scientiarum Fennicae, Series A1: *Mathematica-Physica* **37**, 3-79 (Transl.: RAND corp., Santa Monica, CA, Rep. T-131, 1960).

[9] M. Loéve: *Probability Theory*, 3rd ed. New York: Van Nostrand, 1963.

[10] The State University "Dubna": http://www.uni-dubna.ru.

[11] P.V. Vasiliev, V.V. Ivanov, V.V. Korenkov, Yu.A. Kryukov and S.I. Kuptsov: *System for Acquisition, Analysis and Control of Network Traffic for the JINR Local Network Segment: the "Dubna" University Example*, JINR Communications, D11-2001-266, JINR, Dubna, RUSSIA, 2001.

[12] D.S. Broomhead and G.P. King: *Extracting Qualitative Dynamics from Experimental Data*, Physica D, **20**, 217-236 (1986).

[13] D.S. Broomhead and G.P. King: *Time-series Analysis*, Proc. Roy. Soc. London, **423**, 103-110 (1989).

[14] R. Vautard, P. Yiou and M. Ghil: *Singular Spectrum Analysis: A Toolkit for Short, Noisy Chaotic Signals*, Physica D, **58**, 95-126 (1992).

[15] "CATERPILLAR" Version 1.00. Copyright 1997 Caterpillar Group. Program for time series analysis.

[16] C.K. Chui: *An Introduction to Wavelets.* Academic Press: New York, 1-18(1992).

[17] I. Daubechies: *Wavelets*, Philadelphia: S.I.A.M., 1992.

[18] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery: *Numerical Recipes in C: The Art of Scientific Computing*, II-d Edition, Cambridge University Press 1988, 1992.

[19] I. Antoniou, V.V. Ivanov, Valery V. Ivanov, and P.V. Zrelov: *On the Log-Normal Distribution of Network Traffic*, Physica D 2901 (2002) 1-14.

[20] W.T. Eadie, D. Dryard, F.E. James, M. Roos and B. Sadoulet: *Statistical Methods in Experimental Physics*, North-Holland Pub.Comp., Amsterdam-London, 1971.

[21] F. James: *MINUIT – Function Minimization and Error Analysis*, Reference manual, version 94.1, CERN Program Library D506, 1998.

[22] R. Brun, O. Couet, C. Vandoni and P. Zanarini: *PAW - Physics Analysis Workstation*, CERN Program Library Q121, 1989.

[23] D.S. Broomhead and G.P. King: *Extracting qualitative dynamics from experimental data*, Physica **20D** (1986), 217.

[24] A.M. Albano, J. Muench, C. Schwartz, A.I. Mees, and P.E. Rapp: *Singular value decomposition and the Grassberger Procaccia algorithm*, Phys. Rev. **A38** (1988), 3017.

[25] G.V. Martinov: *Omega-squared criteria*, Moscow, "Nauka", 1978 (in Russian).

Received on July 28, 2003.

Антониоу Я. и др.
E11-2003-148

Анализ главных компонентов измерений
информационного трафика

К измерениям информационного трафика применен метод главных компонентов на основе подхода «Caterpillar»-SSA [1,2]. Этот подход оказался очень эффективным для понимания основных особенностей компонентов, формирующих информационный трафик. Статистический анализ показал, что уже несколько первых компонентов формируют основную часть информационного трафика. Остаточные компоненты играют роль небольших нерегулярных возмущений и могут интерпретироваться как стохастический шум. Используя характерные особенности остаточных компонентов, мы разработали статистический метод для их отбора и последующего исключения из общего числа главных компонентов.

Работа выполнена в Лаборатории информационных технологий ОИЯИ.

Препринт Объединенного института ядерных исследований. Дубна, 2003

Antoniou I. et al.
E11-2003-148

Principal Component Analysis
of Network Traffic Measurements

We applied the Principal Component Analysis, especially the «Caterpillar»-SSA approach [1,2], to the network traffic measurements. This approach proved to be very efficient for understanding the main features of term forming the network traffic. The statistical analysis of leading components has demonstrated that a few first components already form the main part of information traffic. The residual components play a role of small irregular variations which do not fit in the basic part of network traffic and can be interpreted as a stochastic noise. Based on the feature characteristics of residual components, we developed a statistical method for the selection and elimination of residuals from the whole set principal components.

The investigation has been performed at the Laboratory of Information Technologies, JINR.

Preprint of the Joint Institute for Nuclear Research. Dubna, 2003

Макет *Т. Е. Попеко*

Подписано в печать 06.08.2003.
Формат 60 × 90/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,18. Уч.-изд. л. 1,98. Тираж 320 экз. Заказ № 54044.

Издательский отдел Объединенного института ядерных исследований
141980, г. Дубна, Московская обл., ул. Жолио-Кюри, 6.
E-mail: publish@pds.jinr.ru
www.jinr.ru/publish/