Nonlinear Analysis of Network Traffic Measurements

P. Akritas[1], P.G. Akishin[2], I. Antoniou[1,3], A.Yu. Bonushkina[2], I. Drossinos[4], V.V. Ivanov[1,2], Yu.L. Kalinovsky[2], V.V. Korenkov[2] and P.V. Zrelov[2]

[1]*International Solvay Institutes for Physics and Chemistry, CP-231, ULB, Bd. du Triomphe, 1050, Brussels, BELGIUM*

[2]*Laboratory of Information Technologies, Joint Institute for Nuclear Research, 141980, Dubna, RUSSIA*

[3]*Department of Mathematics, Aristoteles University of Thessaloniki, 54006 Thessaloniki, GREECE*

[4]*European Commission, Joint Research Center, I-21020 Ispra (Va), ITALY*

Abstract

We applied a *nonlinear* time series analysis approach to the traffic measurements obtained at the input of a medium size local area network. In order to reconstruct the underlying dynamical system, we estimated the correlation length and the embedding dimension of the traffic series. The estimated embedding dimension, based on the Grassberger-Procaccia algorithm, is high. In order to extract the regular part from traffic data and to decrease the system's dimension, we filtered out a high-frequency, *"noisy"* part, applying the wavelet filtering. Using the Principal Components Analysis (PCA), we estimated the number of feature components in the traffic series. The reliable values of the correlation length and the embedding dimension provided the application of a layered neural network for identification and reconstruction of the dynamical system. We have found that the trained neural network reproduces the statistical features of real measurements and confirms the PCA result on the dimension of the traffic series.

Introduction

A major challenge for the emerging high-speed integrated-services communication networks is to develop models that can realistically capture the behavior of the network traffic. The performance of the networks depends on the traffic assessment. Complexity is a key issue in the network geometry and the information traffic. Evidence of traffic complexity appears in many forms, such as the long-range correlations and self-similarities found in the statistical analysis of traffic measurements [1, 2].

We applied a nonlinear time series analysis to the traffic measurements obtained at the input of the Dubna University local area network, including approximately 250 interconnected computers. The performance of the data acquisition system was realized on the basis of standard IBM PC: see Fig. 1.

The recorded traffic data correspond approximately to 20 hours of measurements. The part of this series corresponding to 1 hour of measurements is presented in Fig. 2.

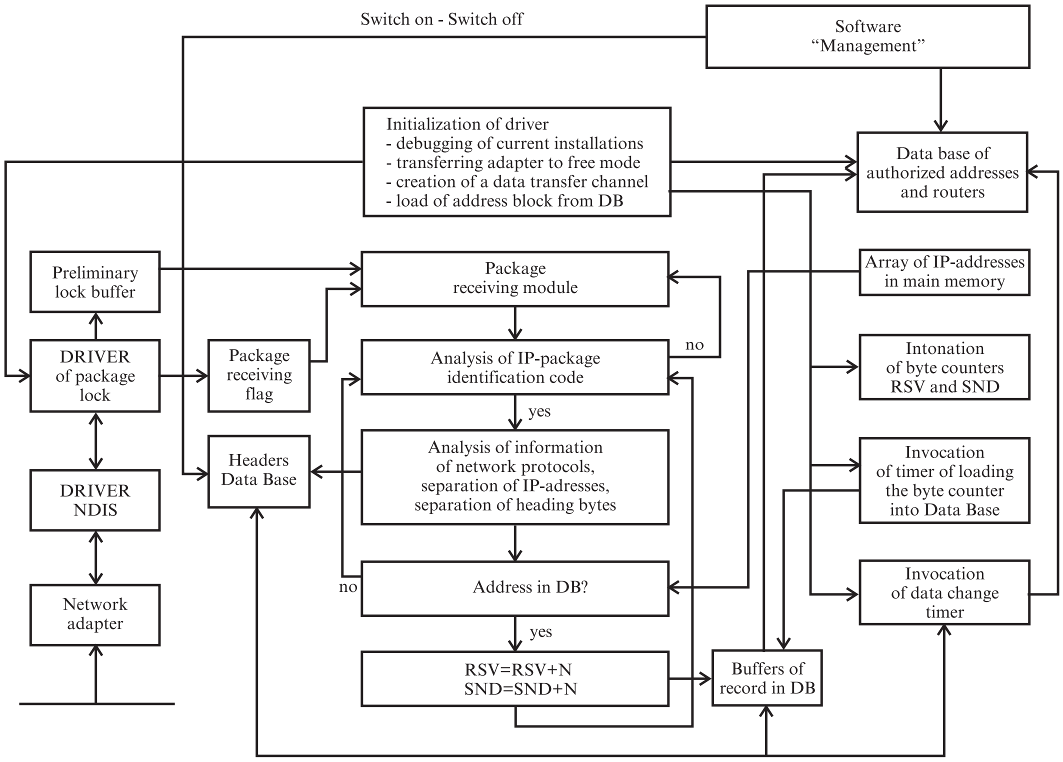

Figure 1: Scheme of a data acquisition system

Estimation of dynamical system parameters

In nonlinear time series analysis we observe the signal $\{x_i\}$ as a one-dimensional projection of a dynamical system operating in a space of vectors $\vec{y}_i$ of larger dimension [3]:

$$\vec{y}_i = (x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau}). \tag{1}$$

Here m is the dimension of the dynamical system, and τ is a "delay time", or the correlation length of series $\{x_i\}$.

In order to choose the independent components from the traffic data, we computed the correlation length τ. It corresponds to the point, where the linear auto-correlation function

$$C(\tau) = \frac{\sum\limits_{i=1}^{N} (x_{i+\tau} - \bar{x})(x_i - \bar{x})}{\sum\limits_{i=1}^{N} (x_i - \bar{x})^2} \tag{2}$$

first crosses the confidence tube corresponding the Gaussian white noise.

The dependence of the correlation length versus the aggregation bin size is plotted in Fig. 3. We see that for bin sizes from 0.1 sec up to 10 sec, the correlation length is in acceptable region: $\tau \sim 10$ sec. The points separated by the time interval τ can be considered as linear independent.

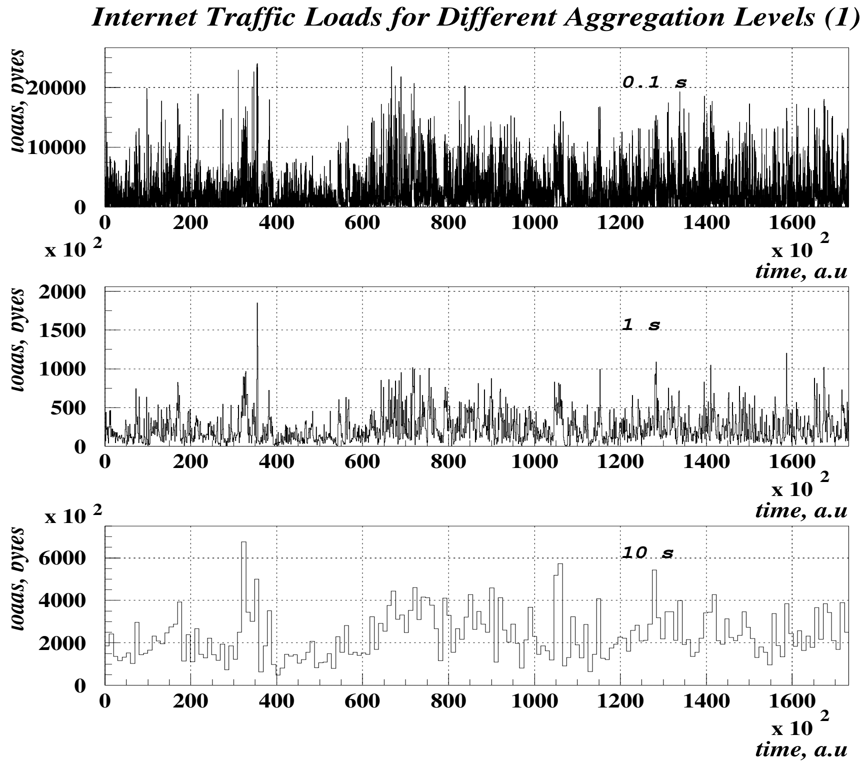

Figure 2: Traffic measurements aggregated with different bin sizes: 0.1 sec, 1 sec, 10 sec

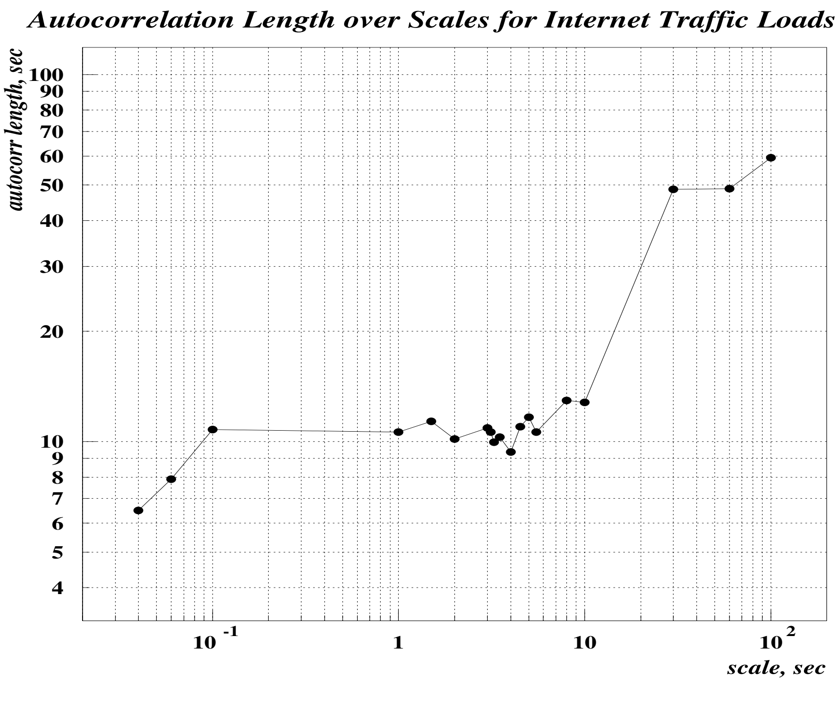

Figure 3: The dependence of the correlation length versus the aggregation bin

The embedding dimension counting for aggregated time series has been performed with the Grassberger-Procaccia algorithm, based on the correlation integral:

$$C_n^m(r) = \frac{2}{N(N-1)} \sum_{i \neq j} \Theta(r - |\mathbf{y_i} - \mathbf{y_j}|), \tag{3}$$

with the distance between two points given by

$$|\mathbf{y}_i - \mathbf{y}_j| = \max\left\{|x_i - x_j|, ..., |x_{i+(m-1)\tau} - x_{j+(m-1)\tau}|\right\}.$$

Here $\Theta = 1$ if its argument is non-negative and 0 otherwise.

The value $C_n^m(r)$ is the empirical probability that a randomly chosen pair of points $(\vec{y}_i, \vec{y}_j)$ will be separated by a distance less or equal to r. The integral $C_n^m(r)$ is computed for r ranging from 0 to the largest possible value $|\vec{y}_i - \vec{y}_j|$ and for m increasing from 1 up to the largest possible value. The slope of the logarithm $\log C_n^m(r)$ versus $\log r$ gives the lowest estimate of the embedding dimension: Fig. 4.

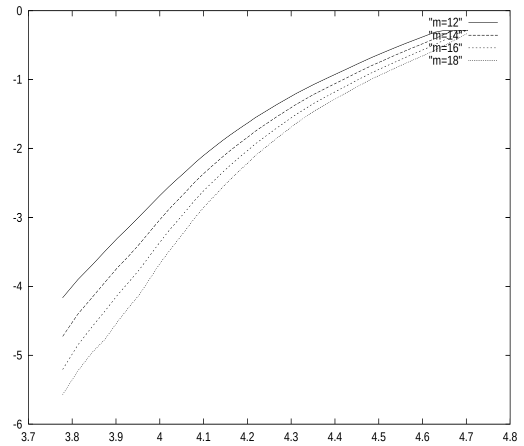

Figure 4: Correlation integrals $C_2^m(r)$ for traffic measurements aggregated with 1 sec bin: $\tau = 10$ sec and m =12, 14, 16, 18

For different parts of the time series analyzed, no saturation of the slope with respect to increasing m was found. According to the Takens theorem [4] this may imply a very high dimension of the studied time series.

Wavelet filtering and Principal Components Analysis

As usual we may consider the traffic measurements as a sum of a regular process and a stochastic part, related to the high frequency "noise". Elimination of the *noisy* part may simplify the analyzed data and reduce the dimension of the dynamical process. In order to achieve this, we used a discrete wavelet transform based on Daubechies wavelets [5]. In Figure 5 we present part of the original traffic series, the corresponding filtered signal and the noisy component.

After filtering out the noisy component from traffic measurements, we estimated the Grassberger-Procaccia dimension for the remaining regular part. We observed that for all curves the dimension about $16 \div 18$ seems to be close to saturation.

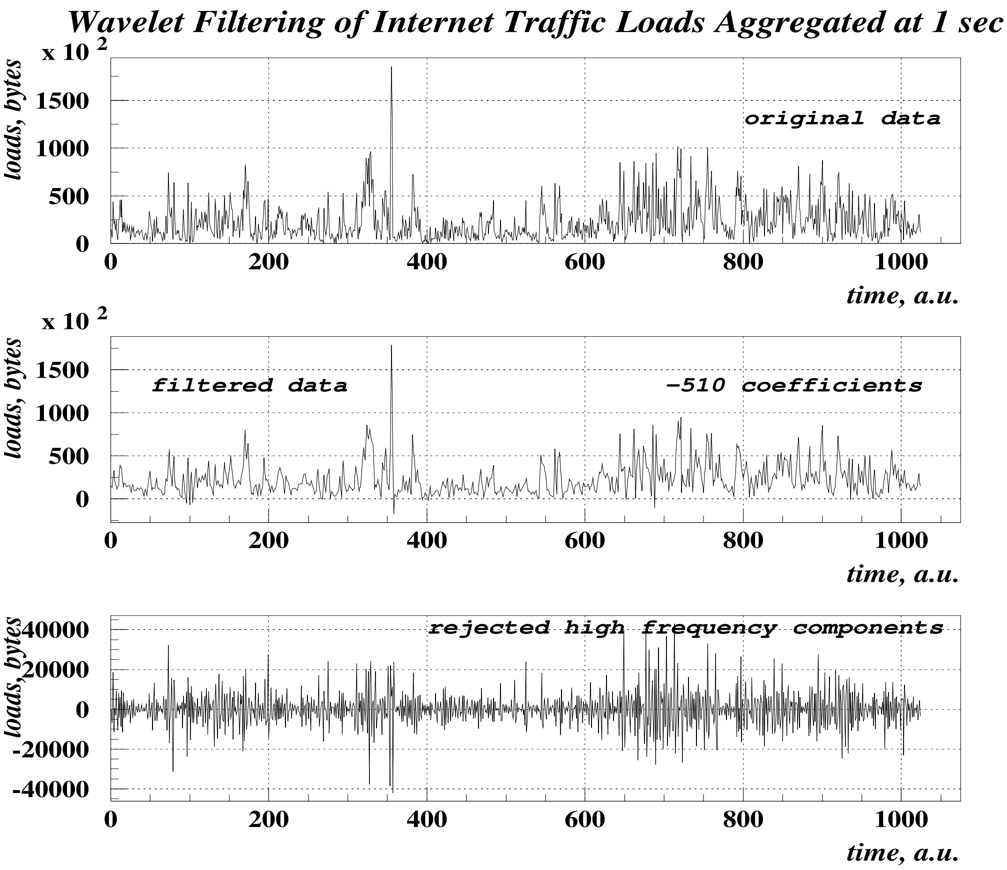

Figure 5: Traffic measurements: a) original traffic series, (b) filtered signal, (c) noisy component

The additional method which has been used for clarification of the dimensionality problem is the Principal Components Analysis [6]. We applied this method to the original traffic data and to the filtered data. Figure 6 shows the PCA results. We see that the intrinsic information of the analyzed system is accumulated in the few first components.

Reconstruction of dynamical system

In order to reconstruct the dynamical system corresponding to the traffic measurements, we used an artificial neural network (ANN). In our study we applied a layered neural network with the feed-forward architecture from the JETNET3 package [7]. For the ANN training we used a data set corresponding to approximately 34 minutes of measurements and aggregated with time bin 1 *sec*. These data were preliminary cleaned applying wavelet filtering for elimination of "noisy" component. The following parameters were used for the input vector formation: $\tau = 10\,sec$ and $d_E = 15 \div 20$. Figure 7 presents part of the traffic data (traffic.dat) and the result of the ANN approximation (train.dat) after 1000 training epochs. Figure 8 shows the statistical distributions of *Errors* (the difference between an actual value and the ANN

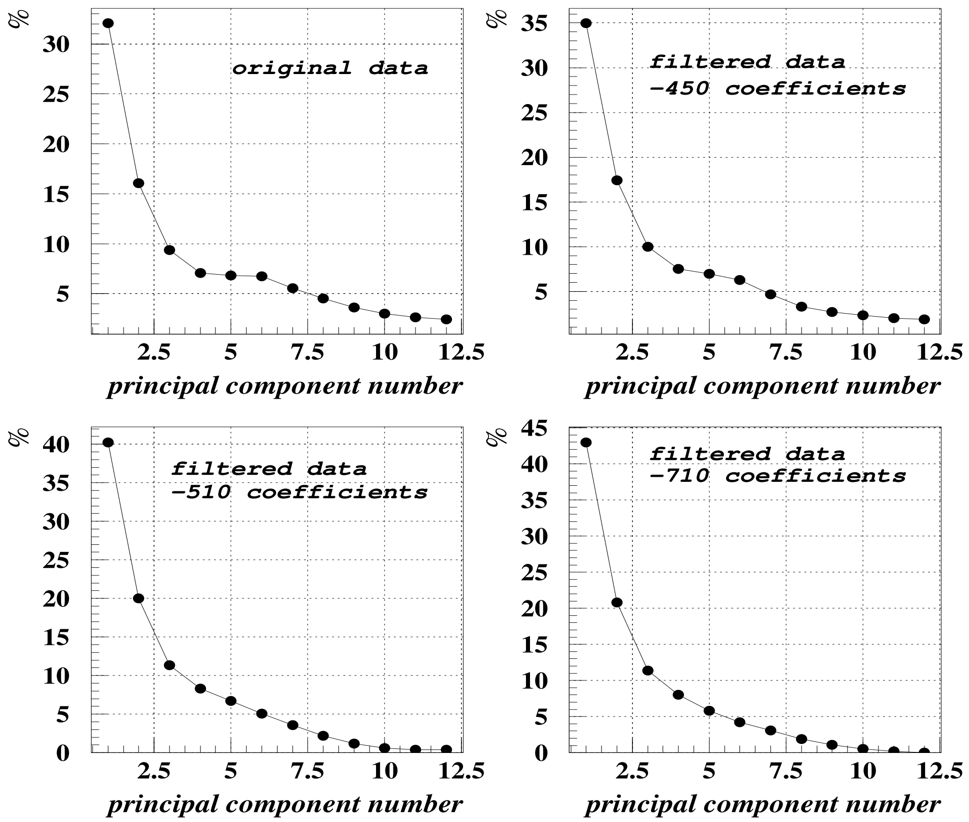

Figure 6: Principal components: (a) for original data, (b)-(d) for time series filtered at different thresholding level

approximation) and *Relative Errors* (the ratio error/actual value) after 1000 training epochs. We can see that, despite the highly chaotic character of time series, the ANN approximate this data quite well.

It is obvious that the ANN after training can be used as a model for simulation of traffic measurements. It can be seen as a nonparametric model, because this approach does not require any assumptions concerning possible parameters of the analyzed time series.

Figure 9 demonstrates the distributions of sizes of traffic packages (normalized to the interval [-1,1]) for the original traffic measurements (top figure) and for time series generated by the ANN (bottom figure). We see that the ANN model reproduces the statistical distribution of real data quite well.

Conclusion

In our work we systematically applied the nonlinear time series analysis to the traffic measurements obtained at the input of the medium size local area network. Using the estimated values for the time lag and the embedding dimension, we successfully ap-

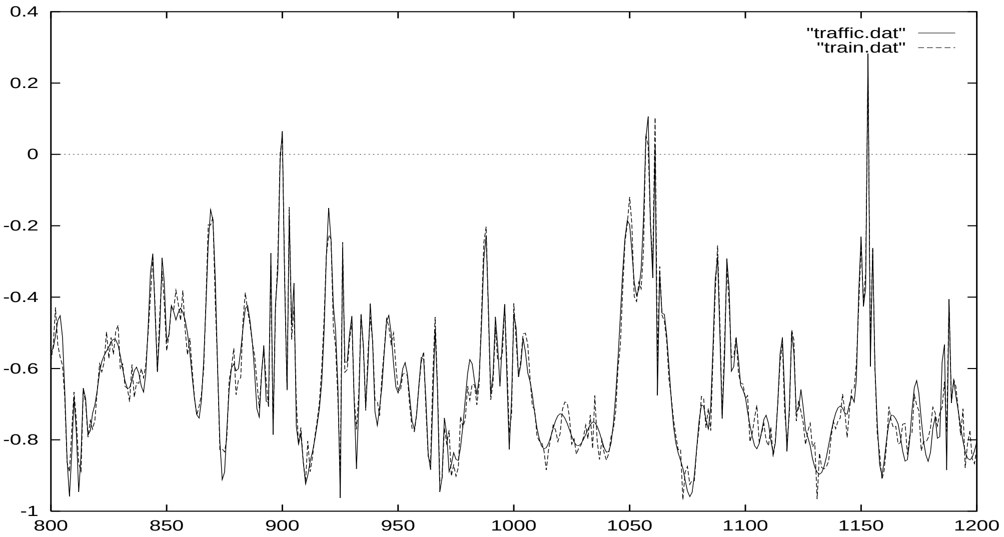

Figure 7: The result of the ANN approximation of the traffic series after 1000 training epochs

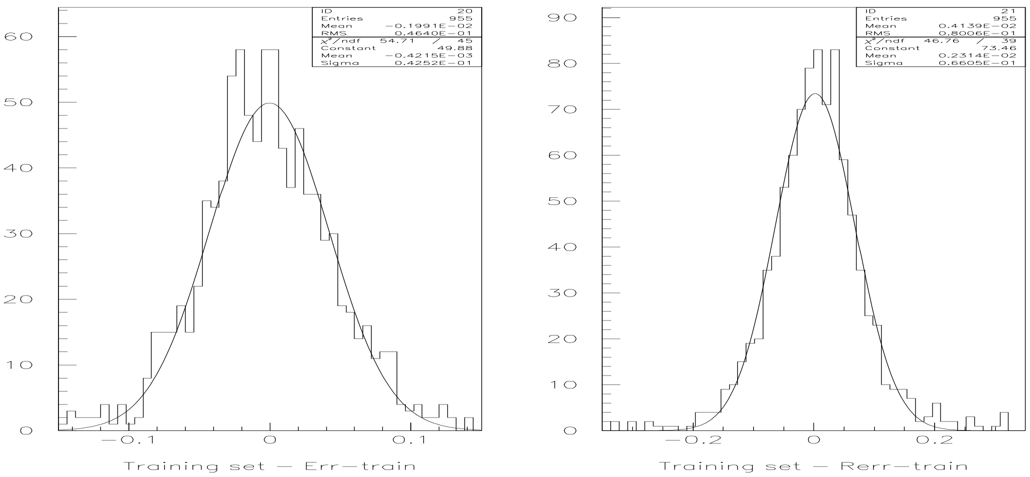

Figure 8: Distribution of *Errors* (left) and *Relative Errors* (right)

plied a layered feed-forward neural network for the identification and reconstruction of the dynamical system underlying the traffic measurements. We found that the trained neural network reproduces the statistical features of real traffic data.

Acknowledgements

We thank Yu. Kryukov for the preparation of traffic data.

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

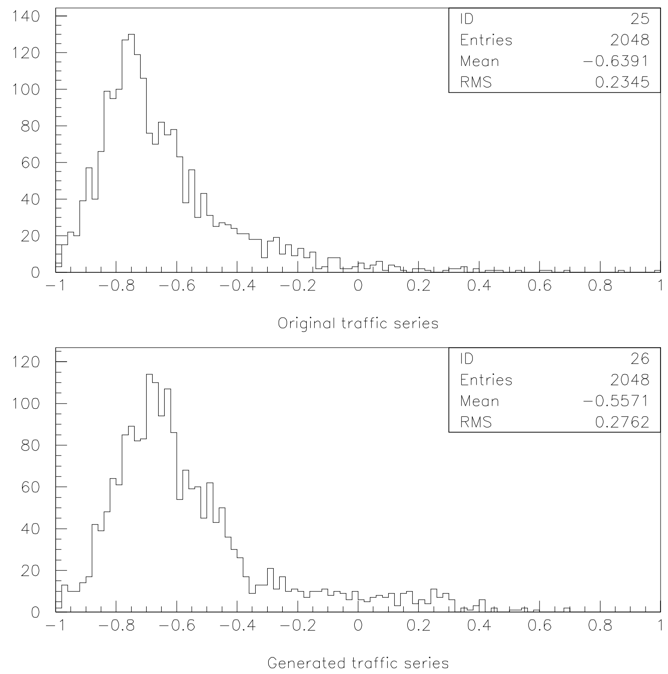

Figure 9: The distribution of sizes of the traffic packages (normalized to the interval [-1,1]) for : a) the original traffic measurements, and b) the generated by the trained ANN

References

[1] W Leland, M.Taqqu, W. Willinger, and D.Wilson: *On the Self-Similar Nature of Ethernet Traffic (Extended Version)*, IEEE/ACM Transactions on Networking, 2(1), pp. 1-15, February 1994.

[2] M.T. Lucas, D.E. Wrege, B.J. Dempsey, and A.C. Weaver: *Statistical Characterization of Wide-Area Self-Similar Network Traffic*, University of Virginia Technical Report CS97-04, October 9, 1996.

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

[4] F. Takens: *Detecting strange attractors in turbulence* in "Dynamical Systems and Turbulence", edited by D. Rand and L.S. Young, Lecture Notes in Mathematics **898** (Springer-Verlag, Berlin, 1981), 366.

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

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

[7] C. Peterson and Th. Rongvaldsson: *JETNET-3.0 – A Versatile Artificial Neural Network Package*, LU Tp 93-29, 1993.