Proceedings of the 8th International Conference on Applied Informatics
Eger, Hungary, January 27–30, 2010. Vol. 1. pp. 255–264.

Wavelet Analysis of Qos Based Network Traffic

Zoltán Gál[a], György Terdik[b]

[a]Service Center for Informatics, University of Debrecen
e-mail: ZGal@unideb.hu

[b]Department of Informatics, University of Debrecen
e-mail: Terdik@delfin.unideb.hu

Abstract

Integration of different network services into a converged infrastructure is one of the most intensive challenges in the converged network application development area. Huge differences exist regarding several characteristics of data, voice and video content traffics. While data traffic requires error free services, real time applications like interactive video transfer and VoIP distress of the high delay and jitter values.

DiffServ becomes more and more popular Quality of Service (QoS) mechanism not only in WAN but LAN environment as well. Coloring the IP packets of different traffic streams based on QoS traffic classes generate increased diversification of the statistical characteristics detected at the measuring points of the production network. Self similarity, long range dependence and fractal characteristics of these packet flows are strongly influenced by the QoS parameters in congested network environment.

Several models are proposed for the qualitative and quantitative evaluation of physical phenomenon supervened on different OSI layers at the intermediate nodes. Most of these claims relatively long traces to evaluate the scale independence and fractal characteristics. The wavelets based Multi Resolution Analysis (MRA) proposes fast pyramidal algorithm requiring $\sim O(n)$ computation steps for determining the self similarity measure of the traces with n samples, consisting serious interest in the low delay aspect of the burst detection.

In this paper several UDP and TCP traffics are considered and statistically analyzed based on MRA method. Fast detection algorithm of real time traffic burstiness is presented for QoS based packet switched WAN environment with congestion.

Keywords: IP, TCP, UDP, VoIP, Wavelet, MRA, QoS, LRD, Self-similarity.

1. Introduction

The parameters of IP data traffic networks are strongly influenced by the type of applied communication mechanisms. In spite of most popular character of the Internet no global guarantees exits for communication services. Serious efforts are set forth development of QoS methods worldwide involving analyzing of bandwidth, delay variation, blocking rate of services, packet lose rate, reliability and other parameters [9]. The QoS mechanisms for IP data networks are based on two approaches: i.) High channel bandwidth conception represented by the IETF (Internet Engineering Task Force) assuming enough and large capacity of cheap channel capacity for all network applications working on those links; ii.) Managed bandwidth conception, having origin from the BISDN services of the ATM Forum and assuming finite channel capacities with optimum set of resource parameters. The first method suppose higher evolution rhythm of the communication technology than the demand of users, sustaining low price of bit transmission, and on the other side the switching delay time of switches/routers can be decreased continuously. Such an over dimensioned worldwide network could resolve lot of current problems, but the economical aspects of network resource wastage prevents this solution. The managed bandwidth conception allows increase of network resources for applications running on limited channel capacities. Several technologies offer such solutions. First data link layer technology with QoS features was Frame Relay, allowing traffic prioritization with the bit DE (Discard Eligible) of FR frame and FECN/BECN (Forward Explicit Congestion Notification / Backward Explicit Congestion Notification) mechanisms. Explicit QoS mechanisms were introduced at the ATM (Asynchronous Transfer Mode) technology by the set-up phase of connection oriented (CO) L2 link. The MPLS (Multi Protocol Label Switching) technology uses L2.5 label switching CO mechanism to provide service guarantees approximate to the ATM.

The introduction and dispersion of QoS mechanism were generated by the VoIP (Voice over IP) services running on the Internet [5]. The converged network services like data, video and voice on the same common infrastructure has strong influence in QoS demand composition. The majority of current Internet applications are accessed from Ethernet end nodes. This aspect implies cost reduction if the intermediate nodes of LAN/MAN networks have similar L2 technology because the inexistence of frame type modification during the transmission simplifies the amount of switching task at the core and distribution level [10]. The different types of network applications put up different transfer guarantees (see Fig. 1). The network resource demand of applications is varying in time even the applications are differently tolerant to the transmission delay and the delay variation. Several applications can support data loss while others cannot.

Two QoS mechanisms are proposed for popular network applications today. Intserv is based on RSVP (Resource Reservation Protocol) and assures quantitative metric and latency time for the guaranteed services. The connection allowance control process decides of the successful of the new connection request. This control mechanism suppose necessary RSVP intelligence at the intermediate nodes. Based on this strict assumption, Intserv mechanism is only useful in autonomous system environment. The second, Diffserv QoS mechanism utilizes special fields of the PDUs (Protocol Data Unit) and the QoS priority is transmitted as color code inside of each PDU for a given OSI layer. This classification of each PDU implies different treatment possibility inside of the routers and switches. There are distinguished eleven categories of traffic: voice, interactive video, streaming video, signaling, IP routing, network management, critical data, transactional data, bulk data, best effort, and scavenger. The tag field of L2 trunk mechanism (IEEE 802.1Q), and the QoS field (formerly TOS – Type of Service) of IP packet header constitute the most successful fundamental control channel through the packet switched communication network. This mechanism is proposed for ISPs and becomes more and more popular in the converged network environment.

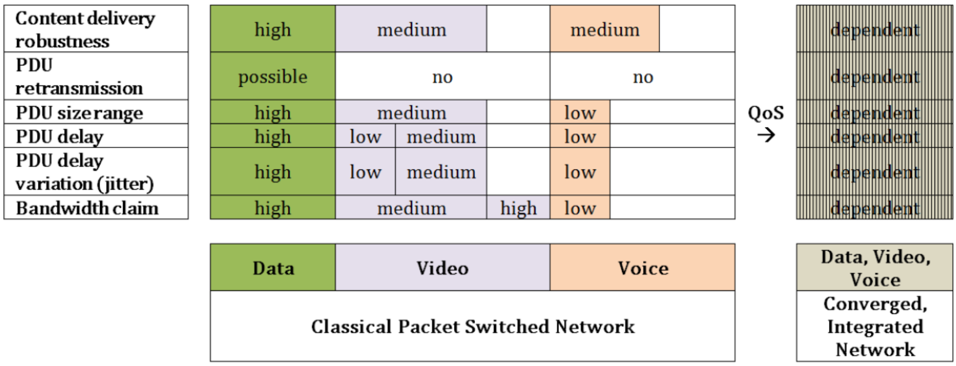

Figure 1: Characteristics of classical and converged packet switched networks

2. Wavelet analysis fundamentals

Several models are proposed for the qualitative and quantitative evaluation of physical phenomenon supervened on different OSI layers at the intermediate nodes. Most of these claims relatively long traces to evaluate the scale independence and fractal characteristics. Let $\mathrm{X_t}$, $\mathrm{t} \in \mathbb{Z}$ be a discrete time stochastic process. The continuous time series $\mathrm{Y}(\mathrm{t})$, $\mathrm{t} \in \mathbb{R}$ with stationary increment is called a cumulative (background) process of $\mathrm{X_t}$ and $\mathrm{X_t}$-t is the increment process of Y(t), if the following relation holds [6]:

$$X_t = Y(t) - Y(t-1), \ \forall t \in \mathbb{Z} \tag{2.1}$$

In the case of network traffic modeling, $\mathrm{X_t}$ is stationary in the sense that the behavior or structure is invariant to the time offset. In a different sense, the time interval $[0, \mathrm{t}]$ has an absolute frame reference feature. The real value, continuous process $\mathrm{Y}(\mathrm{t})$, $\mathrm{t} \in \mathbb{R}$ is $\mathrm{H}-\mathrm{ss}$ (H-self similar with Hurst parameter), if the following holds:

$$Y(at) =_{\mathrm{d}} a^{H} Y(t), \ \forall a > 0, \ t \geq 0, \text{ and } (0 < H < 1). \tag{2.2}$$

This relation means that processes $\mathrm{Y}(\mathrm{t})$, $\mathrm{t} \in \mathbb{R}$ and $\mathrm{Y}(\mathrm{at})$, $\mathrm{t} \in \mathbb{R}$ are the same in the sense of finite dimensional distribution. In traffic modeling $\mathrm{Y}(\mathrm{t})$ represents the quantity of data transferred in time interval $[0, \mathrm{t}]$. The time series generated form $\mathrm{X_t}$ by disjoint blocks of length m is named $\mathrm{X}^{(\mathrm{m})}$, aggregated process of grade m, if holds:

$$X_i^{(m)} = \frac{1}{m} \sum_{t=m(i-1)+1}^{mi} X_t, \ \forall \ \mathrm{t} \in \mathbb{Z}. \tag{2.3}$$

The second order self similarity implies the existence of correlated and asymptotically correlated structure during time aggregation. If process $\mathrm{Y}(\mathrm{t})$, $\mathrm{t} \in \mathbb{R}$ is H-sssi, then it is H-ss with stationary increment. If $\mathrm{Y}(\mathrm{t})$ is H-sssi with finite variance, then $0 < H < 1$ [6]. Process $\mathrm{X_t}$, $\mathrm{t} \in \mathbb{Z}$ is H-sss (H-Self similar and stationar) is called SRD (Short Range Dependent) if $<H \leq 1/2$ and is LRD (Long Range Dependent) if $1/2 < H < 1$. The autocorrelation function $\mathrm{r_X}(\mathrm{k})$, $\mathrm{k} \geq 1$ of second order self similar process $\mathrm{X_t}$, $\mathrm{t} \in \mathbb{Z}$ with Hurst parameter $0 < H < 1$, $H \neq 1/2$ can be estimated by following relation:

$$r_X(k) \sim \ \mathrm{H}(2\mathrm{H}-1) k^{2H-2}, \qquad if \ \mathrm{k} \to \infty. \tag{2.4}$$

The wavelets based Multi Resolution Analysis (MRA) proposes fast pyramidal algorithm requiring $\sim\mathrm{O}(\mathrm{n})$ computation steps for determining the self similarity measure of the traces with n samples, consisting serious interest in the low delay aspect of the burst detection. Decomposition

$$\{\mathrm{T_Y}(\mathrm{a}, \mathrm{t}) = \langle \mathrm{X}, \psi_{\mathrm{a,t}} \rangle, \ \mathrm{a} \in \mathbb{R}^+, \mathrm{t} \in \mathbb{R}\} \tag{2.5}$$

is called CWT (Continuous Wavelet Transform), where the coefficients are given by the inner products of analyzing function set and signal $\mathrm{Y}(\mathrm{t})$, $\mathrm{t} \in \mathbb{R}$. The set of functions $\psi_{\mathrm{a,t}}$ is generated from the ψ_0 reference pattern, called mother wavelet, in the following manner:

$$\left\{\psi_{\mathrm{a,t}}(\mathrm{u}) \equiv \frac{1}{\sqrt{\mathrm{a}}} \psi_0 \left(\frac{\mathrm{u}-\mathrm{t}}{\mathrm{a}}\right), \ \mathrm{a} \in \mathbb{R}^+, \mathrm{t} \in \mathbb{R}\right\}. \tag{2.6}$$

Any function element $\psi_{\mathrm{a,t}}(\mathrm{u})$ can be constructed from ψ_0 mother wavelet by operator $\mathcal{D}_\mathrm{a}$-dilatation (scaling) and then $\mathcal{T}_\mathrm{t}$-time shifting:

$$\begin{aligned}(\mathcal{T}_\tau \psi_0) &\equiv \psi_0(\mathrm{t}-\tau)\\ (\mathcal{D}_\mathrm{a} \psi_0) &\equiv 1/\sqrt{\mathrm{a}}\psi_0(\mathrm{t}/\mathrm{a})\end{aligned} \tag{2.7}$$

The ψ_0 mother wavelet (e.g. DaubechiesN, CoifletN, HaarN, SymmletN, etc.) spread in a narrow space in time and frequency domains and has most of its energy within a limited frequency (scale) band [1]. It satisfies the admissibility condition:

$$\int \mathrm{u}^\mathrm{k} \psi_0(\mathrm{u}) \, \mathrm{du} = 0, \mathrm{k} = 0, \ 1, \ 2, \ \ldots, \ \mathrm{N}-1, \tag{2.8}$$

that is it should be a bandpass or oscillating function, hence the name “wavelet”. Quantity $|T_Y(a,t)|^2$ is called scalogram and gives the energy level of signal Y controlled by dilatation parameter a around time moment t.

The MRA (Multi Resolution Analysis) proves that there exists a special sampling of the time-frequency plane, which contains all the information of signal Y as a subset of $T_Y(a,t)$, $a \in \mathbb{R}^+, t \in \mathbb{R}$. The construction method of this discrete subset is called DWT (Discrete Wavelet Transform). The basic idea of MRA is that signal Y is analyzed sequentially with rougher and rougher approximations. In this way, higher and higher frequency components are being left out from consideration. Detailed coefficients $d_Y(j,k)$ of the wavelet transformation are samples of set $T_Y(a,t)$ that place these elements in dyadic grid points:

$$d_Y(j,k) = T_Y\left(2^j, 2^j t\right), \tag{2.9}$$

where the base 2 logarithm of scale $\log_2\left(a = 2^j\right) = j$, is called octave_j. The ψ_0, mother wavelet utilized for analyzing second order processes decays at least exponentially in the time space, ensuring in this way second order statistical characteristics of the wavelet decomposition.

Wavelet coefficients $d_Y(j,k)$ of process $Y(t)$, $t \in \mathbb{R}$, with $H - ss$ feature demonstrate self similarity. For second order processes, this aspect implies following relation:

$$E\left[d_Y(j,k)^2\right] = 2^{j(2H+1)} E\left[d_Y(0,k)^2\right]. \tag{2.10}$$

If process $Y(t)$, $t \in \mathbb{R}$ is H-sssi, then the process of fixed scale index wavelet coefficients $\{d_Y(j,k),\ k \in \mathbb{Z}\}$ is a stationary process. In this way for $\forall\ k \in \mathbb{Z}$ relation (2.10) simplifies:

$$E\left[d_Y(j,k)^2\right] = 2^{j(2H+1)} C(H,\psi_0)\, \sigma^2, \tag{2.11}$$

where $C(H,\psi_0) = \int |t|^{2H} \left(\int \psi_0(u)\psi_0(u-t)du\right) dt$, és $\sigma^2 = E\left[Y(2.1)^2\right]$. This relation of the wavelet coefficients does not assume the Gauss distribution, and can be utilized in the analysis of scale and frequency dependence of LRD processes. In the wavelet domain, we only operate with $d_Y(j,\cdot)$ belonging to octave_j and we manipulate SRD processes during the transformation. Based on (2.8) the wavelet coefficients of mother wavelet ψ_0 all average at zero. The stationary character makes averaging possible to process $d_Y(j,\cdot)$, decreasing the variance in this way. The average statistics have low variance because of the SRD effect.

The random variable μ_j obtained from the $d_Y(j,k)$, wavelet coefficients of second order stationary process Y, is called energy function of process $d_Y(j,\cdot)$, which is the estimation of the standard deviation, too:

$$\text{Energy}_j = \mu_j \triangleq \frac{1}{n_j} \int_{k=1}^{n_j} |d_Y(j,k)|^2, \tag{2.12}$$

where n_j, is the number of existing coefficients of octave_j ($n_j = \lfloor 2^{-j} n \rfloor$) and n is the number of elements of signal Y. The analysis of second order dependence of signal Y is reduced in this way to the analysis of dependence of the energy μ_j of octave$_j$. The scale exponent can be obtained from the bias of the $\log_2(\mu_j)$ curve represented in function of octave$_j$.

$$y_j = \log_2(\mu_j) = \log_2\left(\text{Energy}_j\right) \approx (2H-1)j + c \tag{2.13}$$

The graph of quantity y_j in function of octave$_j$ inside the confidence interval is called second order log-scale diagram (2-LD). A linear segment or segments can be utilized for estimating the Hurst parameter or parameters. If more than one linear segment can be detected, then the process is monofractal, otherwise it is multifractal. Based on the weighted-least-square (WLS) method, the estimated Hurst parameter inside of confidence interval $[j_1, j_2]$ of the octaves is given by:

$$\widehat{H}(j_1, j_2) = \frac{1}{2}\left[\frac{\sum_{j=j_1}^{j_2} S_j j y_j - \sum_{j=j_1}^{j_2} S_j j \sum_{j=j_1}^{j_2} S_j y_j}{\sum_{j=j_1}^{j_2} S_j \sum_{j=j_1}^{j_2} S_j j^2 - \left(\sum_{j=j_1}^{j_2} S_j j\right)^2} + 1\right], \tag{2.14}$$

where $S_j = \frac{n \ln^2 2}{2^{j+1}}$ is weight. The scale independent time domain for the time series with sampling interval T can be calculated from interval $[j_1, j_2]$ of the octaves with the following formula:

$$[j_1, j_2] \ \rightarrow \left[2^{j_1}T,\ 2^{j_2}T\right]. \tag{2.15}$$

3. Analysis of network streams

There were analyzed three traffic streams, the historical Bellcore trace form 1989 and other two captured streams at the University of Debrecen in 2008 and 2007, respectively. These traces were selected in such way to have comparison possibility of Best-Effort and QoS data traffics with and without congestion (see Table 1.). VoIP trunk traffic in congestionless environment was analyzed secondly. The aggregated voice traffic of IP/PBX gateway was captured for a population of 1500 IP phones. The voice trunk link was 100 Mbps Ethernet and the capturing task was effectuated with $\tau = 10$ μsec accuracy in University of Debrecen environment on a working day for a one hour time interval. We did not use special QoS mechanism because the trunk link load was less than 1%. The aggregated voice streams were transmitted with DSCP = 0. In the third case the TCP traffic was transferred with Best-effort (DSCP = 0) mechanism in the QoS domain. The packet of the UDP based video conference fits the Ethernet frame and in this way the priority of the IP packets transmitting UDP video segments could be increased to upper layer by setting DSCP= 56 for these packets.

The independently analysis of a single trace of network transmission (frame interarrival time, or frame size) does not describe fully the physical phenomenon of PDU transmission process. At least two components of this physical process should be analyzed for satisfactory evaluation. This analysis conception become more and more popular just in the last years of network modeling [8]. Based on the transformation ON/(ON+OFF) we generated three complex time series Z_k, having real and imaginary parts the channel usage intensity (M_k) and the channel load $(\tan\varphi_k)$, respectively [2]:

$$Z_k = M_k + i \cdot \tan\varphi_k \ , k = 1, 2, \ldots, \ \lfloor t/T \rfloor \ , \ \ i^2 = -1, \tag{3.1}$$

where T is the sampling period of the traffic.

Data trace	BC-pAug89	VoIP-Trunk	VBR-video-congested
Capture location	Bellcore Morristown Research and Engineering Laboratory, USA	University of Debrecen, Hungary	University of Debrecen, Hungary
Capture date	29-Aug-1989 11:25-11:30	26-Mar-2008 10:00-11:00	2-Apr-2007 20:01-20:02
No. of captured frames	115,888	739,829	6,186
Traffic types	IP in LAN/ WAN network (100% data, 0% HTTP)	VoIP trunk (100% IP voice, 0% data)	TCP + UDP (TCP: data, UDP: video)
L2 channel	IEEE 802.3, 10 Mbps	IEEE 802.3, 100 Mbps	IEEE 802.3, 1 Mbps
QoS type	"Best-effort", no congestion	"Best-effort", no congestion	Congested channel (QoS) UDP (DSCP=56), TCP (DSCP=0)
Sampling period (T)	100 ms	100 ms	100 ms
Accuracy (τ)	10 μs	10 μs	10 μs
Channel intensity mean ($\widetilde{M}$)	3.86	20.54	9.51
Channel intensity deviation	1.67	9.21	2.01
Channel load mean ($\widetilde{\tan(\varphi)}$)	14.44 %	0.0025%	53.94%
Channel load deviation	4.43 %	0.0013%	5.51%

Table 1: Analyzed traces

We analyzed the scale dependence of the real and imaginary LRD parts of $Z(t)$ complex series with wavelet method, and we estimated the Hurst parameters. Because of the strong correlation between the channel intensity and channel load, the scalograms are similar, but the Hurst parameters are different. The time series component transmitting higher energy shows less LRD (see Fig. 5.). The majority of $Z(t)$complex time series are situated in a compact region of the $(M, \ \tan\varphi\ )$ plane, and the histogram of the channel intensity and channel load shows no Gauss distribution for all three traces.

The aggregated traffic of several voice sessions running on the same Ethernet voice trunk channel without congestion are LRD processes and have estimated Hurst parameters of channel load and channel intensity: $\widehat{H_{\tan\varphi}} = 0.64$, $\widehat{H_M} = 0.62$. This phenomenon origins from the constant size of voice frames generated by the codecs implying approximately linear relation between channel load and channel intensity parameters. These values are appropriate but different because there were running more than one voice codec (G.711, G.729, etc) simultaneously between the IP phones and the voice gateway.

The presence of Diffserv QoS mechanism in service provider network radically influences the traffic behavior of classical network applications. This changes can be detected in the third traffic trace and can measured with the complex time series generated by the ON/(ON + OFF) method. The Hurst parameters of the channel intensity and channel load for different QoS conditions are uncorrelated. Mixed Best-Effort TCP traffic with Diffserv based QoS controlled UDP traffic is LRD at the channel intensity time series, and the stationary character of the channel load is cancelled out for both transport layer protocols. The VBR streaming video traffic controlled by the Diffserv QoS mechanism produces more smooth traffic in congested environment. The congested Best-effort based TCP traffic is able to use only a part of the remaining network resources. If the video stream is not treated by QoS and is congested, the video service becomes inoperable.

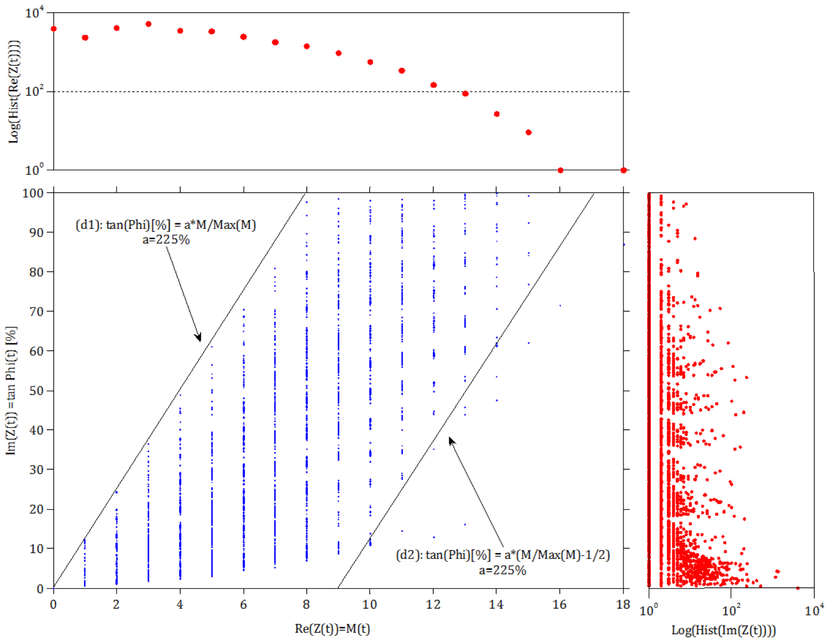

Figure 2: Histograms of the real (M-channel intensity) and imaginary ($\tan\varphi$ -channel load) parts, Bellcore trace

The explanation of this phenomenon is that remarkable part of Bellcore LAN/WAN traffic was generated by TCP-based SMTP and FTP protocols in short packets on lightly loaded 10 Mbps Ethernet channel with Best-effort (without QoS) method, and an important part of the channel was used by control signaling.

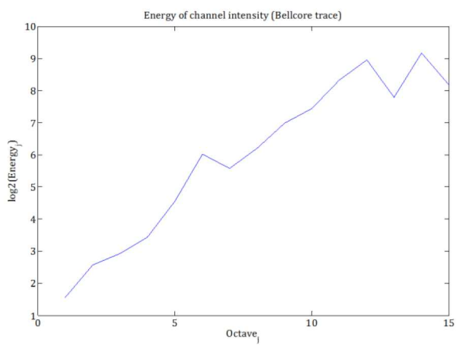

Figure 3: Energy function (M), Bellcore trace

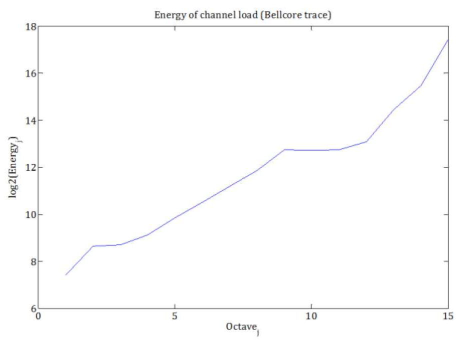

Figure 4: Energy function ($\tan\varphi$), Bellcore trace

The energy functions M and $\tan\varphi$ of a given trace determined by wavelet method are different and the linear interval location differs, as well (see Fig. 3. and Fig. 4.). Similar types of traffic (e.q. voice trunk) traversing channel without congestion implies similar but different energy curves.

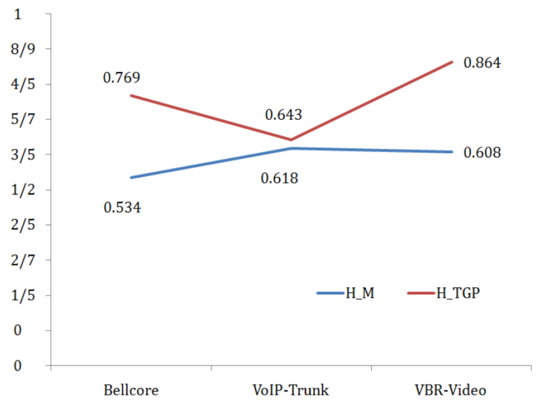

Figure 5: Hurst parameters of complex time series Z

In case of video and data congestion the QoS mechanisms implies higher Hurst parameters of the video and the distance between $\widehat{\mathrm{H}_{\tan\varphi}}$ and $\widehat{\mathrm{H}_{\mathrm{M}}}$ becomes relatively high (0.25).

4. Conclusions and remarks

Both components generated with the ON/(ON + OFF) method are necessary for network traffic analysis because both time series contain information for describing the physical process of content transmission. The ON/(ON + OFF) method is useful not only for Best-effort processes, but for other QoS processes, as well. The $ON/(ON+OFF)$ method based on wavelet transformation is able to estimate Hurst parameter using relatively short traces. This aspect is advantageous in practice, where detection of fractal characteristics of the actual traffic should be estimated asymptotically in real time.

Acknowledgement. This work is supported by TÁMOP 4.2.1./B-09/1/KONV-2010-0007/IK/IT project. The project is implemented through the New Hungary Development Plan co-financed by the European Social Fund, and the European Regional Development Fund.

References

[1] I. Daubechies: Ten Lectures on Wavelets (S.I.A.M., Philadelphia, 1992).

[2] Zoltán Gál: VoIP LAN/MAN traffic analysis for NGN QoS management, *Infocommunications Journal*, Volume LXIV, ISSN: 0866-5583, pp. 22–29, 2009.

[3] ITU-T Study Group 2, *Teletraffic Engineering Handbook*, 2006.

[4] Xiaolong Jin and Geyong Min: QoS Analysis of Queuing Systems with Self-Similar Traffic and Heavy-Tailed Packet Sizes, *Proceedings IEEE Communications Society ICC 2008.*

[5] Luis F. Ortiz: Solving QoS in VoIP: A Formula for Explosive Growth? – Special Report, *Brooktrout Technology, July 2001.*

[6] Kihong Park, Walter Willinger: *Self-Similar Network Traffic and Performance Evaluation*, Copyright, John Wiley & Sons, Inc., 2000.

[7] Karim Mohammed Rezaul, Vic Grout: An Overview of Long-range Dependent Network Traffic Engineering and Analysis: Characteristics, Simulation, Modelling and Control, *Proceedings of the 2nd international conference on Performance evaluation methodologies and tool*, Nantes, France, 2007.

[8] Gy. Terdik, T. Gyires: Internet Traffic Modeling with Lévy Flights. *IEEE/ACM Transactions on Networking*, Vol. 17, No. 1, pp. 120–129, February 2009.

[9] X. Xiao, L. M. Ni: Internet QoS: A Big Picture. *IEEE Network Magazine*, March 1999.

[10] White Paper: Advanced QoS. *Allied Telesis, http://www.alliedtelesis.com*