The Analysis of *R/S* Estimation Algorithm with Applications to WiMAX Network Traffic

Artem A. Lenskiy and Soonuk Seol

School of Electrical, Electronics & Communication Engineering
Korea University of Technology and Education
1600 Chungjeol-ro, Byeongcheon, Cheonan, Chungnam, Republic of Korea 330-708
lensky@koreatech.ac.kr, suseol@koreatech.ac.kr

Abstract

One of the most popular scale exponent estimation algorithms is the Rescaled Range (R/S) analysis. The algorithm estimates the ratio of a range and standard deviation statistics in a window with increasing length. Firstly, through simulations with generated noise we analyze estimation precision of the algorithm and find that the algorithm overestimates and underestimates at lower and higher bounds of the estimation range correspondingly. Secondly, we analytically analyze the reasons for estimation errors. Based on spectral model of self-similar processes we find that the root of incorrect estimation is in the bias that persists in variance calculated for self-similar processes of limited length. Thirdly, we apply R/S algorithm to estimate scale exponents in WiMAX traffic, finally we compare the estimated exponent with the exponent obtained based on standard deviation statistics.

Keywords: *Scale exponent, long range dependency, fractal dimension, network traffic, R/S statistics, detrended fluctuation analysis*

1. Introduction

The network traffic is one of the numerous examples of processes with self-similar properties. It was discovered in the earlier 90s that previously widely used Poisson models do not take into account correlation in network traffic. Novel fractal models found to be more suitable to consider long range dependencies (LRD) of traffic loads. It turned out that LRD of the network effects its efficiency. It has been shown that TCP can induce fractal behavior in network traffic [1]. The analysis of network traffic helps improving network performance and design as it affects the total per-packet delay. The analysis of processes with self-similar properties consists in estimation of a process's scale-exponent. Therefore, the estimation precision is of a great importance in network traffic analysis. One of the most popular estimation algorithms is Rescaled Range analysis that relays on *R/S* statistic. The algorithm calculates range and standard deviation statistics for a window with an increasing length. Incorrect estimation may lead in misinterpretation of network traffic behavior. Ghaderi [2] proposed a network traffic predictor based on assumption that so called Hurst exponent a characteristic of self-similarity, rarely exceeds 0.85. However, our analysis shows that Rescaled Range analysis is significantly overestimates at lower values of H ($H \approx 0$) and underestimates H for values $0.5 < H$. Thus such a conclusion is incorrect provided that *R/S* estimation algorithm is used.

2. The Spectral Model of Self-similar Processes

There are two approaches to define statistically self-similar (SS) processes. In this paper we focus on the model defined in frequency domain [3]. The other model of self-similar processes is defined through autocorrelation function [4]. Both approaches are related to one another. The definition of the SS process by means of power spectral density (PSD) is given by:

Definition 1.

A statistically self-similar random process $x(t)$ is said to be $1/f$ process if its power spectral density is presented in the following form:

$$S_x(f) = \frac{\sigma}{f^{\gamma}} \tag{1}$$

The parameter γ is called the spectral exponent and σ is an arbitrary constant. Spectral exponent γ plays the key role in the behavior of self-similar processes.

Notice that this definition says nothing about the frequency bounds, meaning that f can take any values from zero to infinity which leads to undefined power spectral density at infinity and zero frequency for some γ values. The self-similarity directly follows from the definition (1) as:

$$S_x(af) = a^{-\gamma} S_x(f) \tag{2}$$

The spectral exponent is only one among a number of scale exponents that are widely used in characterization of self-similar processes. The most widely used method for scale exponent estimation is the R/S algorithm. R/S algorithm estimates a parameter H that is linearly linked to spectral exponent γ. In the following section we review and analyze the R/S algorithm and underline over- and underestimation problems associated with the algorithm.

Algorithm 1. *R/S* Analysis

1) Divide x into $M(L) = N/L$ adjacent sub-windows $x = \{X^{(0)}, X^{(1)}, \dots, X^{(M(L)-1)}\}$ with L samples in each sub-window, so $X^{(i)}(L) = \{x_{i \cdot L+1} \dots x_{(i+1) \cdot L}\}$

2) Compute the accumulated departures $Y^{(i)}(L)$ for each sub-window $X^{(i)}(L)$

$$Y^{(i)}(L) = \sum_{k=1}^{L} \left[X_k^{(i)}(L) - \langle X^{(i)}(L) \rangle \right], \tag{3}$$

where $\langle X^{(i)}(L) \rangle$ is the mean value of ith sub-window of size L.

3) Calculate the range of deviation within each sub-window with accumulated departures $Y^{(i)}(L)$ via

$$R^{(i)}(L) = \max\left(Y^{(i)}(L)\right) - \min\left(Y^{(i)}(L)\right) \tag{4}$$

4) Find the sample standard deviation defined for each sub-window $X^{(i)}(L)$

$$S^{(i)}(L) = \sqrt{\frac{1}{L} \sum_{k=1}^{L} \left[X_k^{(i)}(L) - \langle X^{(i)}(L) \rangle \right]^2} \tag{5}$$

5) Compute *R/S* statistics for each $L = \{L_1 \dots L_K\}$ as follows

$$\langle R(L)/S(L) \rangle = \frac{1}{N} \sum_{i=0}^{N-1} \frac{R^{(i)}(L)}{S^{(i)}(L)} \tag{6}$$

6) Finally the Hurst exponent H is estimated by solving a least mean square problem:

$$E^2(H, b) = \sum_{j=1}^{K} \left[\log\left(\left\langle \frac{R(L_j)}{S(L_j)} \right\rangle \right) - H \cdot \log(L) - b \right]^2 \tag{7}$$

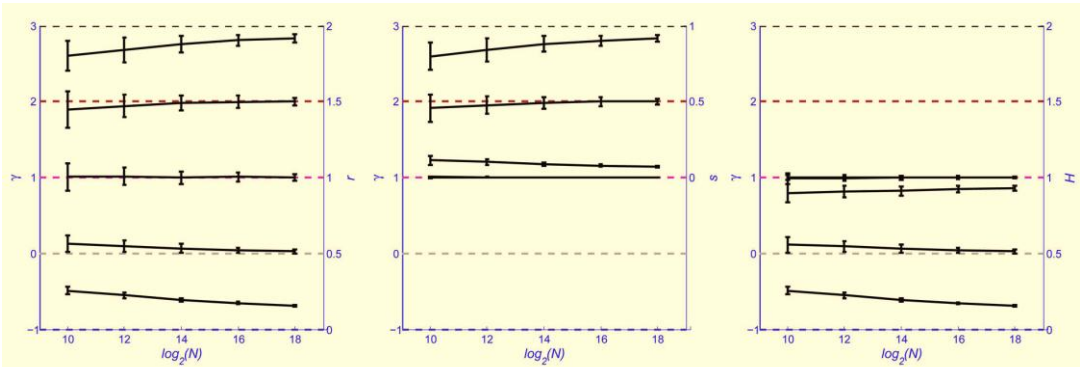
3. R/S Analysis and its Estimation Problems

The *R/S* analysis was originally created by the British hydrologist Harold Edwin Hurst [5, 6] while he was studying the problem of water storage on the Nile river. Later, it was popularized by Benoit Mandelbrot especially in the area of long term dependency analysis of the stock market. The *R/S* analysis estimates the parameter *H* which is a scale exponent and it is related to spectral exponent γ as

$$H = \frac{\gamma+1}{2}, \text{ for } -1 < \gamma < 1 \tag{8}$$

Given the time series $x = \{x_1, x_2, \dots, x_N\}$ of length *N*, the *R/S* statistics are calculated as outlined in algorithm 1.

Form the algorithm it can be observed that *H* is a slope of a line that is a best fit for the relationship $\log L_j \sim \log\left(\langle \frac{R^{(i)}(L_j)}{S^{(i)}(L_j)} \rangle\right)$. Assuming that $R(L_j)$ and $S(L_j)$ do not significantly fluctuate around their means $\langle R(L_j) \rangle$ and $\langle S(L_j) \rangle$, and taking into account that $0 < \langle S(L_j) \rangle < \langle R(L_j) \rangle$, we can separately analyze contributions of numerator and denumerator in estimation of *H*. Thus, we estimate scale exponents *r* and *s* using only the average of range deviations $R(L)$ over cumulative departures (figure 1 a)) and using the average of standard deviations $S(L_j)$ (figure 1 b)) correspondingly (*S* statistics).



(a) (b) (c)

Figure 1. The left ordinate corresponds to value of γ and the right ordinate represents the value of the scale exponent respectively. a) Shows the estimations of *r* based on range analysis of integrated time-series, b) shows the estimations of s based on standard deviation, c) shows the estimations of *H* with the R/S analysis.

By estimating the slope of the line $log(\langle R(L) \rangle)$ against $log\ (L)$, and $log(\langle S(L) \rangle)$ against $log\ (L)$ and then subtracting one from the other gives us an approximation of the slope of the line fit for $log\left(\langle \frac{R(L)}{S(L)} \rangle\right)$ against $log\ (L)$. Such approximation allows us to analyze the behavior of *R/S* statistics and understand how range deviation and standard deviation contribute to the estimation process. To check the precision of each approach we conducted three groups of simulations. We generated five sets of processes for each $\gamma = \{-1,0,1,2,3\}$. This choice of γ covers the range of γs for the most often appearing self-similar processes occurring in nature. For instance $\gamma = 0$ and $\gamma = 2$ correspond to white and Brownian processes. Network processes are usually processes with $\gamma = 2$ such processes are uncorrelated, however with

occurring congestion the value of γ diverges [1]. The process with $\gamma = 1$ is called pink noise and its current value is approximately equally correlated with values from recent and the very distant past. Heart rate variability is a good example of processes with $\gamma = 1$.

To analyze how the length of processes effects estimation precision, we generated 100 realizations with a different number of samples $N = \{2^{12}, 2^{14}, 2^{16}, 2^{18}\}$ for five mentioned above values of γ. We change the window size in accordance with the length of the process in the following manner$L = \{2^{\log N-6} \sim 2^{\log N-2}\}$. For each pair of γ and N we calculated the mean and standard deviation. These calculated values are shown in figure 1(a).

The relationship between the estimated scale exponent r and the spectral exponent γ is given by:

$$r = \frac{\gamma+1}{2}, \tag{9}$$

Note that $\langle R(L) \rangle$ statistics is calculated for cumulated sum. As it can be seen from figure 1 (a) among all values of γ the only accurate estimation is for processes with $\gamma = 1$, ($r = 1$). For processes with $\gamma < 1$ estimations are overestimated and for the processes with $1 < \gamma$ are underestimated. The precision gets better with increasing windows' sizes and time-series length. Nevertheless, there is a strong bias for estimated values with $\gamma = 1, (r = 0)$ that almost does not depend on *N*.

The relationship between estimated scale exponent *s* and the spectral exponent γ is

$$s = \frac{\gamma-1}{2}, \tag{10}$$

From figure 1 (b) it is easy to notice that estimation of scale exponents by observing the growth of standard deviation is only suitable for processes with $\gamma = \{1,2,3\}$. Exponents close to $\gamma = 1$ are overestimated and exponents around $\gamma = 3$ are underestimated. The estimation improves as the numbers of samples in sub-windows and in time-series increase.

The estimated scale exponents for generated processes with $\gamma = \{-1,0,1\}$ are zero, due to the stationary nature of these processes. For such processes standard deviation does not depend on the observation time. The standard deviation increases for processes with $1 \leq \gamma$.

Figure 1 (c) shows the estimation accuracy of *H*. The quality of estimation slowly increases with the increasing number of samples in sub-windows as well as in processes. The relation between *H* and γ is given in eq. (3). The estimation behavior in figure 1 (c) coincides with the following equation,

$$H = r - s \tag{11}$$

that supports the assumption we made earlier. Moreover, from this fact we conclude that the range of deviations and standard deviation behaves similarly. Thus, by subtracting corresponding values of *s* from values of *r*, we obtain good approximation of *H*. Therefore, we can conclude that the division by standard deviation in the *R/S* algorithm leads to narrowing of the estimation range to $\gamma \in (-1,1)$and promotes significant underestimation for processes with $\gamma = 1$. The estimations for processes with $1 \leq \gamma$ overlap with the constant $\gamma = 1$. Thus the *R/S* analysis combines the estimation drawbacks of estimation algorithm for exponents *r* and *s*. The exponent *H=0* is overestimated in figure 1 (c) in the same fashion as r overestimated for a process with true $r = 0$ in figure 1 (a) and *H=1* is underestimated similarly to the underestimated *s* in figure 1 (b).

4. The Sources of Estimation Errors

As we have seen in section 3, by measuring the increase in the standard deviation, the scale exponent *s* is estimated. From the simulation results we have also observed that scale exponent *r* of cumulated departures has similar behavior and the difference between these two exponents is a good approximation of *H*. Therefore we further for simplicity base our analysis on the behavior of standard deviation.

The variance of self-similar processes increases with the increasing observation interval *L* following the power law (12).

$$\sigma^2 = \int_{f_l}^{f_h} f^{-\gamma} df = \begin{cases} \frac{f^{1-\gamma}}{1-\gamma}\Big|_{f_l}^{f_h} & \gamma \neq 1 \\ \ln(f)\Big|_{f_l}^{f_h} & \gamma = 1 \end{cases} \tag{12}$$

Thus the reason for wider estimation range of *r* exponents using range deviation is in the prior summation step. Recalling that integration procedure is equivalent to multiplying the process in frequency domain by $1/jf$, so it leads to the following spectral density after integration:

$$S(f) = \frac{1}{f^{\gamma+2}} \tag{13}$$

Thus, the summation step in *R/S* algorithm increases the spectral exponent by 2 thus *r* is increased by 1 according to (9). The summation step allows estimating *r* exponents for processes with $-1 < \gamma$.

To study the over- and underestimation problems let us denote $f_h = T^{-1}$, where T is sampling period and $f_l = \frac{1}{L}$. Then the variance changes as follows:

$$\sigma^2(L) = \frac{1}{\xi}[L^{\xi} - T^{\xi}], \tag{14}$$

where $\xi = \gamma - 1$.

For $T \ll L$ the relationship (14) converges to

$$L^{\xi} \propto \sigma^2(L), \tag{15}$$

on other hand when *L* is comparable with the sampling period *T*, the power law rolls off and the rate of the deviation from the power law depends on ξ.

The variance exponents are estimated as a ratio between the logarithm of the observation intervals and the logarithm of the variance calculated for a process with sampling period *T*:

$$\hat{\xi} = \lim_{L\to\infty} \hat{\xi}(L) = \lim_{L\to\infty} \left[\frac{\ln(L^{\xi} - T^{\xi})}{\ln L} - \frac{\ln \xi}{\ln L}\right] \tag{16}$$

Although when *L* tends to infinity, the limit converges to a correct value of the exponent, the rate of convergence varies for various exponents. The difference in the convergence rate results in incorrect estimation on the limited range of scales.

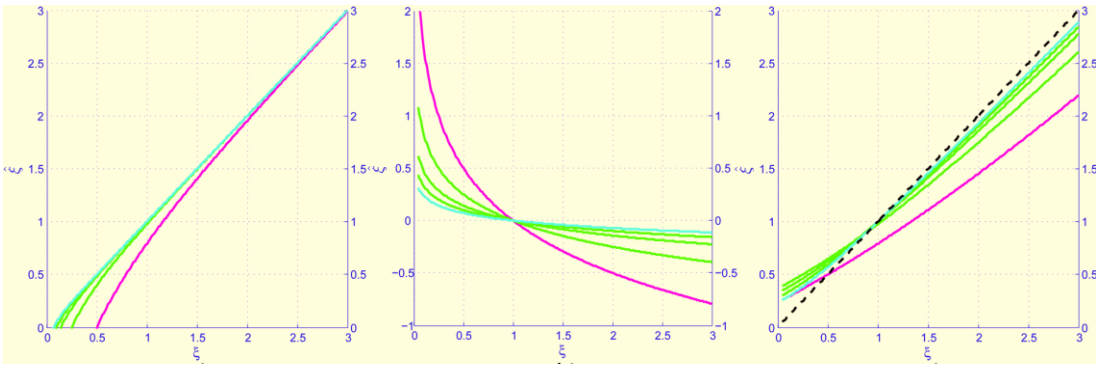
Let us first analyze the cause of the overestimation problem. To do so, we take a closer look at the right side of expression (16), that consists of two summands. The first summand is

$$\hat{\xi}_1(L) = \frac{\ln(L^{\xi} - T^{\xi})}{\ln L} \tag{17}$$

For the self-similar processes with infinitely fine details i.e. high frequency limit $f_h = T^{-1} = \infty$, the summand (17) reaches a constant $\hat{\xi}_1(L) = \xi$. However, when $T \neq 0$ and $L = T$, the numerator tends to $-\infty$. Thus, in estimation procedure for the case when observation interval is comparable with the sampling period, the variance behavior rolls off.

Figure 2 (a) shows the plots of the expression (17) for $\xi \in [0.05; 3]$ while $L = \{2^2, 2^4, 2^7, 2^{10}, 2^{14}\}$. As it can be seen for longer L the underestimation effect is less adverse and it is especially strong at lower values of ξ. Thus, the process discretization leads to underestimation, however we saw from the simulation results (Figure 1c) that estimated values are overestimated. The cause of overestimation is the second summand in (16), i.e.:

$$\hat{\xi}_2(L) = -\frac{\ln \xi}{\ln L} \tag{18}$$



(a) (b) (c)

Figure 2. The plots show the relation between the true ξ and the estimated $\hat{\xi}$ for $L = \{2^2\}$ magenta, $L = \{2^4, 2^7, 2^{10}, 2^{14}\}$ green, and $L = \{2^{14}\}$ cyan curves. a) $\hat{\xi}_1(L)$, b) $\hat{\xi}_2(L)$and c) $\hat{\xi}(L) = \hat{\xi}_1(L) + \hat{\xi}_2(L)$, the black line correspond to ideal match of the estimated $\hat{\xi}$ and the true ξ.

When $L \to \infty$, $\hat{\xi}_2(L)$ tends to zero, and summand (18) does not influence the estimation. On the other hand for finite observation interval, the summand $\hat{\xi}_2(L)$ significantly contributes to the total variance, especially for greater ξ. Figure 2(b) shows the plots of (17) for $\xi \in [0.05; 3]$ and $L = \{2^2, 2^4, 2^7, 2^{10}, 2^{14}\}$. The longer the observation interval the closer $\hat{\xi}_2(L)$ to zero at all values of ξ. Nevertheless the expression (17) diverge at $\xi = 0$.

The estimated exponent is a sum of (17) and (18) and is shown on figure 2(c). It is seen now that when the true values of ξ are close to zero, the estimated exponents are always overestimated. The overestimation effect is less for longer L. From figure 2(c) it is also seen that for longer L, variance exponents $1 < \xi$ are always underestimated. The underestimation gets worse for greater values of ξ.

Thus, comparing the effects of (17) and (18) on the estimation we conclude that even though the estimation is effected by the finite high frequency limit, the main cause of the overestimation and underestimation is in the nature of variance measurement of self-similar processes, specifically in $-\ln \xi$ summand. The estimation error is minimal for$\xi = 1$, i.e. $\gamma = 2$ that corresponds to brown noise. When $\xi = 1$, the estimation is only effected by the finite high frequency limit, that makes the estimation underestimated (fig. 2(a)). Generally, for longer observation intervals, scale exponents ξ are overestimated for $\xi < 1$ and underestimated for $1 < \xi$. In terms of *Hurst* parameter *H*, the underestimation occurs for $H < 0.5$ and overestimation occur for $0.5 < H$.

5. Experiments

The estimation algorithm is also analyzed with real network traffic data. The data have 613 samples of average daily downlink traffic throughput in bit per second (figure 4 (a)). The network traffic data are collected from a real router in mobile WiMAX networks in Seoul, Korea, from October 2010 to June 2012. In the target router, there have been about 4,100 active subscribers in average for that period. Note that the subscribers had been growing in general but sometimes were shrunken due to rearrangement of the base stations among routers. In general, 100~300 base stations are connected to one router and the router is directly connected to the backbone network.

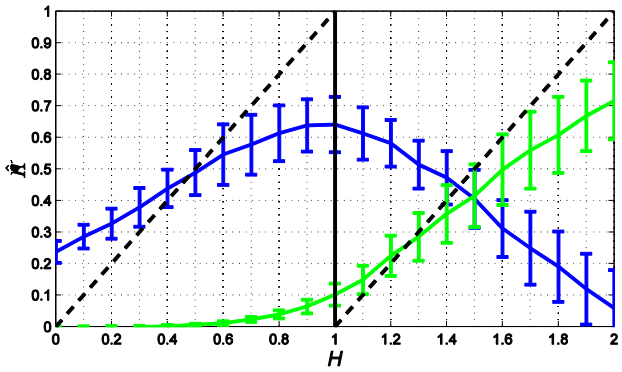
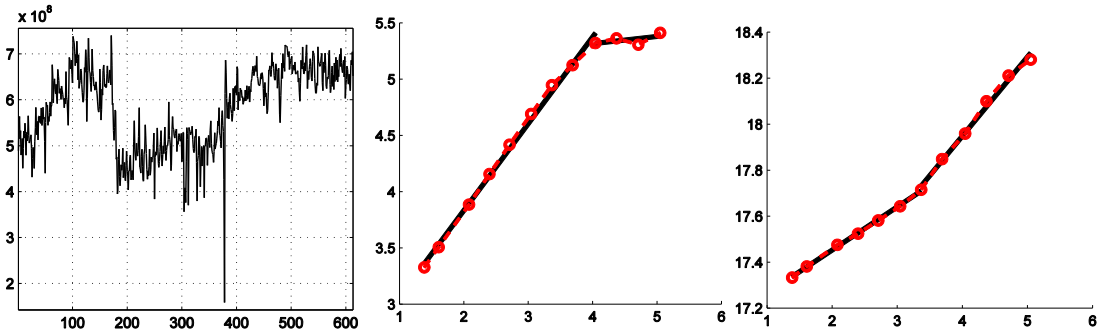


Figure 3. The abscissa and ordinate are true and estimated scale exponents respectively. Green and blue lines are value of estimated exponents using *R/S* statistics and standard deviation correspondingly. The black dashed lines correspond to perfect estimations using R/S and S statistics.

Prior to estimating exponents for real data we generated groups of 100 random sequences with $H \in [0,2]$ and with the step 0.1. The length of each sequence is equal to 613, i.e. the length of available WiMAX downlink traffic throughput. The average and standard deviation among estimated exponents were calculated in each group. To estimate exponents we used *R/S* and *S* statistics with $L = [1.4^{(4\ldots15)}]$. The estimation is even worse on such small sequences (fig. 3) than on longer sequences (fig. 1(b),(c)). Figure 3 also shows the behavior of estimated exponents $1 < H$ using *R/S* statistics. The processes with scale exponents $1 < H$ are better estimated with *S* statistic. However, the results are still unsatisfactory.

To check WiMAX network traffic on self-similarity, we plot log(*R/S*) versus log(*L*) . We chose the same range of $L = [1.4^{(4\ldots15)}]$ as in simulations above. Analyzing figure 4(a) it is easy to see that WiMAX traffic is self-similar with a crossover occurring at a point $L = 57$ that is approximately equal to the number of days in two month. The slope of the first segment is 0.77, whereas the slope of the second segment is 0.06. Both values diverge from 0.5, and therefore we conclude that WiMAX traffic is not perfectly Brownian. Moreover the first estimated exponent is above 0.5 and thus it is underestimated. The slope of the second segment is almost zero. Such small estimated *H*, according to figure 3 (blue line), only could correspond to$2 < H$, thus *R/S* algorithm is not capable of estimating the exponents for this process. A better choice for estimating the exponent for WiMAX traffic is by using *S* statistics. The crossover this time occurred at a point $L = 29$ that is approximately equal to the number of days in a month. This phenomenon has to be studied. The slope of the first segment is 0.19, whereas the slope of the second segment is 0.34. According to simulations (figure 3, green line) these value are more trustful. The corresponding

values of γ eq. (10) are 1.38 and 1.68. From these values we can derive stronger negatively correlated behavior of the traffic within one month time and weaker negatively correlated behavior with time range over a month.



(a) WiMAX daily network traffic (b) Relationship log(*R/S*) vs log(*L*) (c) Relationship log(*S*) vs log(*L*)

Figure 4. WiMAX Network Traffic Analysis

6. Conclusion

The *R/S* algorithm is suitable only for process with Hurst parameter around 0.5. The estimation error tends to zero when lengths of sub-windows tends to infinity. We found that the estimation quality depends on the lengths of sub-windows. Especially for processes with lower values of γ, longer sub-windows should be used to minimize the overestimation effect. However, this is often impossible, as in our example of WiMAX data, when only records for about one and a half year are available. Therefore *R/S* finds a limited area of applications and other estimation algorithm should be applied. We applied *S* statistics to estimate scale exponent for WiMAX traffic data. From the prominent crossover and the estimated values $\gamma < 2$ we concluded negative correlation in WiMAX network traffic. The phenomenon of the crossover point has to be further studied with more reliable estimation algorithms [7].

Acknowledgements

This paper was supported by the Education and Research Promotion Program of KOREATECH.

This paper was supported by the New Professor Research Program of KOREATECH.

References

[1] N. Wisitpongphan and J. M. Peha, "Effect of TCP on self-similarity of network traffic", in 12th IEEE international conference on computer communications and networks, **(2003)**.

[2] M. Ghaderi, "On the relevance of self-similarity in network traffic prediction", School of Computer Science, University of Waterloo, **(2003)**.

[3] M. S. Keshner, "1/f noise," Proceedings of the IEEE, vol. 70, **(1982)**, pp. 212-218.

[4] B. B. Mandelbrot and V. Ness, "Fractional Brownian motions, fractional noises and applications," SIAM Review, vol. 10, **(1968)**, pp. 422-437.

[5] E. E. Peters, "Chaos and order in the capital markets", New York: John Wiley, **(1991)**.

[6] H. E. Hurst, "Long-term storage capacity of reservoirs", Trans. Am. Soc. Civ. Eng., vol. 116, **(1951)**, pp. 770-808.

[7] W. Bucaoto and A. Lenskiy, "Relationship of the Heartbeat and Interbreath Interval Dynamics in Young and Elderly People", presented at the The 43rd ISCIE International Symposium on Stochastic Systems: Theory and Its Applications, Shiga, Japan, **(2011)**.