Long-Range Dependence in WiMAX traffic. A Preliminary Analysis

Cristina Stolojescu
"Politehnica" University, Electronics and Telecommunications Faculty,
Communications Department, Timisoara, Romania
Institut Telecom; Telecom Bretagne, UMR CNRS 3192 Lab-STICC,
Université européenne de Bretagne, Brest, France
c.stolojescu@etc.upt.ro

Abstract— **Last years, long-range dependence (LRD) has become a key concept in analyzing the traffic in a telecommunication network. The aim of this paper is to analyze the downlink traffic in a WiMAX network in terms of LRD. For the estimation of Hurst parameter we adopted the Rescaled Adjusted Range Method (R/S). We have obtained estimated values for each downlink trace from the data base which correspond at the considered network. Some particularities of the network can be established analyzing these values. Rules for the optimization of the network's exploitation can be derived analyzing these particularities.**

Keywords- long-range dependence, Hurst parameter, WiMAX traffic.

I. INTRODUCTION

Several empirical studies of traffic have been made recently for various communication networks. These studies have proved that the actual Internet traffic is self-similar (fractal) or long-range dependent, [1-6].

Long-Range Dependence (LRD) is a statistical measure for the speed of decaying of the autocorrelation of a time series. A random process has a Short-Range Dependence (SRD) if its autocorrelation decays fast.

More intuitively the LRD measures the length of the memory of a random process.

The predominant way to quantify the LRD of a random process is the estimation of its Hurst parameter, H. The Hurst parameter can not be calculated, it can be only estimated [3]. Some estimators of H are proposed in [4, 10].

A time series exhibits LRD if $0.5<H<1$. The LRD is stronger if the parameter H has a higher value.

The structure of this paper is the following. Section II offers a short introduction to the topic of LRD in communications networks. In Section III we will define the Hurst parameter and we will refer to a method for estimating it, the Rescaled Adjusted Range (R/S) method. The aim of section IV is to estimate the LRD of the WiMAX traffic. Section V is dedicated to some concluding remarks.

II. LONG-RANGE DEPENDENCE

Processes with LRD are often confused with self-similar processes (SSP). Some SSP may exhibit LRD, but not all random processes having LRD are SSP. There are various definitions of LRD in the literature. One of them is the following. A stationary process X_t is LRD if the series with the general term given by the k^{th} sample of its autocorrelation function (ACF), $\rho(k)$, diverges:

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty . \tag{1}$$

The ACF of the process X_t measures the similarity between X_t and a shifted version of itself, X_{t+k}:

$$\rho(k) = \frac{E[(X_t - \mu)(X_{t+k} - \mu)]}{\sigma^2} , \tag{2}$$

where μ represents the mean value of X_t, σ represents its standard deviation and E denotes the statistical expectation operator.

The condition for a stationary process X_t to exhibit LRD (1) has an asymptotically behavior, [8]:

$$\rho(k) \sim c_p k^{-\alpha} , \tag{3}$$

where c_p is a constant and $\alpha \in [0,1]$. The Hurst parameter of the process X_t can be defined using the variable α:

$$H = 1 - \alpha / 2 . \tag{4}$$

When the autocorrelation $\rho(k)$ decays slowly then equation (1) is verified and the corresponding stationary process X_t exhibits LRD.

Tacking into consideration the Wiener-Hincin theorem, the LRD of a stationary process X_t can be also defined in terms of its power spectral density [8]. A stationary process X_t exhibits LRD if its power spectral density verifies the following relation:

$$f(\lambda) \sim c_f \, |\lambda|^{-\beta} , \tag{5}$$

where c_f is a positive constant and $\beta \in [0,1]$.

978-1-4244-8458-4/10/$26.00 ©2010 IEEE

In this case the relation between the Hurst parameter H and β is the following:

$$H = (1+\beta)/2 . \qquad (6)$$

Previous work, [1-6] proved the utility of H for the analysis of the Internet traffic. In 1993 was identified the presence of LRD in data sets captured on Ethernet Local Area Network (LAN) traffic, [8]. In the case of Ethernet LAN traffic, LRD is manifested in the absence of a natural length of a `burst'; at every time scale ranging from a few milliseconds to minutes and hours, bursts consist of bursty sub-periods separated by less bursty sub-periods. So, a cause of LRD is the hidden periodicities which are present in the time series analyzed. It is also shown that the value of the Hurst parameter typically depends on the utilization level of the Ethernet and can be used to measure `burstiness' of LAN traffic.

In the literature, four possible origins for LRD in networks are commonly cited. These are as follows:

(1) LRD is inherent directly in the source of data.

(2) LRD is a result of the aggregation of heavy-tailed data streams.

(3) LRD is a result of feedback mechanisms in the TCP protocol.

(4) LRD arises from network topology.

These causes are explained in detail below. It is important to emphasize that these explanations are not contradictory. Each might make a contribution to the packet traffic behavior of the network.

(1) The evidence that LRD arises directly in the source of data comes mainly from studies of video traffic. Variable-bit-rate (VBR) video traffic by its nature exhibits LRD at source. The LRD, in this case, arises from an encoding mechanism whereby video is encoded as a series of differences between frames with occasional full updates. Measurements of Internet traffic show that real time video traffic is likely to be only a very small percentage of Internet traffic. It seems unlikely, therefore, that VBR video traffic could be the main component of LRD observed in aggregate traces. Of course, if VBR video traffic contains LRD at source then it is likely that other applications might have traffic distributions with unexpected statistical effects.

(2) The proposal that LRD in Internet traffic arises from the aggregation of heavy-tailed data streams is similar to the previously mentioned mechanism but has a slightly less direct causal mechanism. A causal connection between heavy-tailed sources and LRD was long suspected and finally was proven. There are good reasons to believe that source traffic on the Internet is heavy-tailed file sizes and sizes of accessed web documents have been shown to have heavy tails.

(3) Another potential cause of LRD is the feedback mechanisms in the Transmission Control Protocol (TCP). Markov chains were used to model TCP timeout and congestion window behavior and it was proved that these can cause local LRD (that is, LRD up to a certain time scale).

(4) Finally, there remains the distinct possibility that LRD is an emergent property of the networks themselves. Measurements made have shown that, even when packet inter-departure times are independent, arrival times at the destination show LRD. This obviously indicates that round trip times in networks are LRD processes.

Determining the origin of LRD in Internet networks remains an important research area and it is uncertain which (if any) of these four causes is really at the heart of the problem. The possibility remains that it is a mixture of some or all of them.

The aim of this paper is to apply the LRD analysis in the case of a relatively new network technology. The WiMAX technology was applied recently. We have not found in the literature studies about the LRD of the traffic of WiMAX networks. Possessing a data base containing traces for a WiMAX network we considered useful to estimate their LRD. We have chosen the downlink traffic to take into account all the sources of LRD. A simplified analysis could be done for the uplink traffic, when some sources of LRD can be eliminated. Indeed video traffic does not exist in uplink. The feedback mechanisms in the TCP protocol does not manifest in uplink. The present preliminary LRD analysis of the WiMAX traffic will take into account all the four sources of LRD, already denoted by (1) - (4). From a network designer point of view the most interesting LRD is that denoted by (4).

III. THE ESTIMATION OF HURST PARAMETER

The degree of LRD increases as $H\uparrow 1$, [7]. A value of H equal to 0.5 or smaller indicates the lack of LRD or the presence of SRD, [2].

There are various statistical techniques to estimate H. We can classify the estimators of H into two general categories, [4]:

1. the estimators operating in the time domain, which can be computed using the following methods:
   - absolute value method,
   - variance method,
   - rescaled adjusted range (R/S) method,

and,

2. the estimators operating in the frequency or wavelet domain, which can be computed using the following methods:
   - periodogram,
   - Whittle estimator, [10],
   - proposed by Abry-Veitch, [2].

In the following we will use the R/S method to estimate the Hurst parameter.

A. R/S method

The R/S analysis was proposed by Hurst in 1951. Given a sample of observations X_k, with k = 1,...,N, we must subdivide the whole sample into K non-overlapping blocks and calculate the rescaled adjusted range $R(t_i,n)/S(t_i,n)$ for each of the new "starting points" $t_1 = 1$; $t_2 = N/K+1$; $t_3 = 2N/K+1;\ldots$ and for the variable n which satisfy ($(t_i - 1) + n \le N$). R represents a range series, while S is the standard deviation series. Representing $\log(R(t_i,n)/S(t_i,n))$ versus $\log(n)$ is obtained the R/S plot. It is a straight line with the slope equal with the Hurst exponent, [9].

IV. RESULTS

This section presents the evaluation of H with the aid of the R/S method in the case of WiMAX traffic. Its goal is to highlight the particularities of WiMAX traffic from a LRD perspective.

A. Data base

In the following we will use historical data obtained by monitoring the traffic from 67 Base Stations (BS) composing a WiMAX network. The period of collection is of eight weeks, from March 17th till May 11th, 2008. Each BS has its own data set which is composed of numerical values representing the total number of packets from the downlink channel. We have selected the downlink connection tacking into account the considerations made at the end of section II.

The value of the interval between two consecutive recorded data is of 15 minutes. It can be easily deduced that for a given BS we have the following number of samples: 96 samples/day, 672 samples/week and a total number of 5376 samples.

B. Evaluation of H

We used for our experiments SELFIS (SELF-similarity analysIS), a java-based software tool for self-similarity and long-range dependence analysis, developed by T. Karagiannis and M. Faloutsos [3].

As a first experiment we calculated the value of R/S estimator for all the 67 time series, corresponding to all BSs. The results are presented in Table I.

TABLE I. H VALUES FOR THE TIME SERIES CORRESPONDING TO ALL 67 BASE STATIONS.

BS	**H**	BS	**H**
1	0,693	35	0,678
2	0,628	36	0,729
3	0,658	37	0,667
4	0,682	38	0,626
5	0,691	39	0,719
6	0,676	40	0,697
7	0,606	41	0,698
8	0,665	42	0,756
9	0,665	43	0,622
10	0,643	44	0,66
11	0,657	45	0,681
12	0,689	46	0,641
13	0,656	47	0,56
14	0,692	48	0,608
15	0,645	49	0,618
16	0,641	50	0,704
17	0,706	51	0,667
18	0,641	52	0,629
19	0,618	53	0,703
20	0,657	54	0,628
21	0,657	55	0,654
22	0,6	56	0,636
23	0,723	57	0,603
24	0,706	58	0,591
25	0,717	59	0,669
26	0,679	60	0,648
27	0,74	61	0,664
28	0,656	62	0,661
29	0,665	63	0,581
30	0,619	64	0,631
31	0,637	65	0,657
32	0,719	66	0,727
33	0,653	67	0,628

We can observe that the values of H are between 0.57 and 0.756, so Hurst parameter belongs to the theoretical interval that proves the presence of LRD ($H \in [0.5,1]$). The mean value of H equals 0.66.

Next, we have chosen a certain BS (for example BS 61) and we have split the time series into weeks. We have obtained eight new time series and for each of these series we have calculated the new values of H. In this case the minimum value equals 0.472, while the maximum value is equal to 0.649. So we can conclude that all the values of H are smaller than the mean value of the entire series (composed by the eight weeks). This observation can be explained by the presence of a hidden periodicity into the entire series. The corresponding period is higher or equal with the duration of a week because the LRDs corresponding to each week are smaller.

Finally, we will split again the series, this time computing H for series containing daily values. The values for the first week, for BS 61 are presented in Table II.

TABLE II. H VALUES FOR THE 7 DAYS COMPOSING A WEEK

Day	R/S
1	0,385
2	0,523
3	0,429
4	0,325
5	0,396
6	0,473
7	0,891

Analyzing the results, we observe that the values of H are smaller that the mean value per week. There is an anomaly in the case of the 7^{th} day explained by the fact that in this series we have detected the presence of thirty values of 0, which produced an error of estimation. This comportment can be attributed to the presence of a hidden periodicity into the series with the duration of one week. The corresponding period must have a value higher or equal with the duration of a day. We have identified this periodicity in the following experiment. We have computed the power spectral densities (PSD) for the time series corresponding to every week of the entire series for BS 61 (see table III) in order to verify the existence of periodicities. The concentration of the PSD around a specific harmonic indicates a periodicity with a period corresponding to the frequency of that harmonic. All the PSDs corresponding to every week of the entire series for BS61 are concentrated around the eighth harmonic. The amplitudes of the eighth harmonic of those series are presented on the second column of table III. So, the dominant period across all traces is the 24 hours one. Hence the reduction of the weekly LRD is explained by the daily periodicity.

TABLE III. PSD CORRESPONDING TO EVERY WEEK FOR A BS ARBITRARILY SELECTED.

Week	PSD ($*10^{14}$)
1	1,2
4	4,2
5	12,2
2	3,5
8	20
3	2,5
6	5
7	2,4

Another interesting observation relating Table II is that the majority of the values from the second column are inferior to 0.5, indicating that the LRD was completely eliminated.

V. CONCLUSIONS

The goal of this paper is to realize a long-range dependence analysis for WiMAX traffic. We have shown that WiMAX traffic exhibit LRD by estimating Hurst parameter using R/S method. We proved that LRD depends on the duration of the time series. The entire series exhibits stronger LRD than each weekly series. In the same way we demonstrated that the value of Hurst parameter for each of the seven daily series is smaller than the H value corresponding to the given weekly series. This comportment can be explained by the presence of some hidden periodicities. We have highlighted a periodicity with the period equal with the duration of a day. The presence of another periodicity with period equal with the duration of a week can be also supposed. These periodicities are a consequence of social reasons and not of technical reasons. The network is more exploited during the day and less exploited during the night; the network is less exploited during the weekends. We have also proved that the daily series associated to BS 61 does not manifest LRD. The LRD sources (1)-(4) are not present in the case of BS 61. In consequence the BS 61 has a good localization in the topology of the WiMAX network under study. This procedure can be applied to all the BSs which compose the considered network indicating those with bad localization. This way, the network topology can be corrected. Other improvement strategies for the considered WiMAX network could be imagined after the LRD analysis of the uplink traffic. The analysis reported here is only preliminary. It will be completed in the future by the addition of a part dedicated to the uplink traffic.

ACKNOWLEDGMENT

The research reported in this paper was developed in the framework of a grant of the Romanian Research Council (CNCSIS) with the title "Using Wavelets Theory for Decision Making" no. 349/13.01.09. The authors thank Alcatel Lucent, Timisoara, for providing the WiMAX traffic data base and for helpful discussions around WiMAX network.

REFERENCES

[1] Kihong Park and Walter Willinger. "Self-Similar Network Traffic: An Overview.", Technical Report, 1999.

[2] P. Abry and D. Veitch. "Wavelet analysis of long range dependent traffic", IEEE Trans. Information Theory, 1998, volume 44(1), pages "2-15".

[3] Thomas Karagiannis, Michalis Faloutsos, "SELFIS: A Tool For Self-Similarity and Long-Range Dependence Analysis", 1st Workshop on Fractals and Self-Similarity in Data Mining: Issues and Approaches (in KDD), Edmonton, Canada, July 23, 2002.

[4] Thomas Karagiannis, Mart Molle, Michalis Faloutsos, "Long-range dependence: Ten years of Internet traffic modeling", IEEE Internet Computing. Special Issue - Measuring the Internet, September 2004.

[5] Ulanovs P. and Petersons E., " Modeling methods of self-similar traffic for Network performance evaluation", Scientifc Proceedings of RTU. Series 7, Telecommunications and Electronics, 2002.

[6] S. Ostring, and H. Sirisena. The Influence of Long-range Dependence on Traffic Prediction. IEEE International Conference on Communications, ICC 2001, vol. 4, pp "1000-1005".

[7] G. Rutka, " Neural Network Models for Internet Traffic Prediction", Proceedings of Electronics and Electrical Engineering, Lithuania, Kaunas, 2006, volume 4(68), pp. 55-58.

[8] Richard G. Clegg, "The Statistics of Dynamic Networks", Technical Report, University of York, Department of Mathematics, 2004.

[9] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of ethernet traffic", Proceedings of ACM SIGCOMM 93, 1993, pp 183-193.

[10] M. S. Taqqu, and V. Teverovsky, "On Estimating the Intensity of Long-Range Dependence in Finite and Infinite Variance Time Series. A Practical Guide to Heavy Tails", Statistical Techniques and Applications, R. J.Alder, R. E. Feldman and M.S. Taqqu, Ed., Birkhauser, Boston, 1998, pp. 177-217.