http://dx.doi.org/10.5755/j01.eee.19.3.3702 ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 19, NO. 3, 2013

Real-time Estimation of Traffic Self-similarity Parameter in Simulink with Wavelet Transform

E. Grab[1], S. Sarkovskis[1]
[1]*Department of Transport Electronics and Telematics, Riga Technical University*
Lomonosova street 1/V, LV-1019, Riga, Latvia
elans.grabs@rtu.lv

***Abstract*—The article describes P/M/1/K queuing model and Hurst parameter estimation unit with Wavelet Transform for real-time estimation of the traffic self-similarity parameter in Simulink. It consists of simulation results, which show the possibility to estimate Hurst parameter in real-time by means of Wavelet transform. It is offered to iteratively calculate the averaged Hurst parameter estimation, which improves the accuracy to the accuracy of the estimation over entire traffic series and in some cases even higher, as it has been shown in simulation results analysis. The estimation deviations have been analyzed as well as variance of such deviations. The results also give directions for further research to improve accuracy of traffic parameters estimation.**

***Index Terms*—Wavelet transform, Hurst parameter, real-time estimator, ethernet networks.**

I. INTRODUCTION

It is widely common for network traffic to be correlated in long-time scale, i.e. the traffic is self-similar (fractal). This has been confirmed long time ago by many researchers, for example data sets described in [1] show these scale dependent properties for actual Ethernet traffic from Local Area Network. There are also other examples mentioned in [2], such as video-traffic with variable bit-rate, Wide Area Networks, etc.

Such long-term correlation significantly impacts queue length, and as of such – waiting delays for data packets. It can be calculated, that for standard Markov traffic with utilization $\rho = 0.8$ and buffer overflow probability $P_{Loss} = 10^{-6}$ it's necessary to supply memory capacity of $K = 55$ data packets. Compared to this, the self-similar traffic with same utilization level and buffer overflow probability would require memory capacity of $K = 5 \cdot 10^8$ packets estimated by formula from [3]–[8], for Hurst parameter of $H = 0.9$ (very high self-similarity) in order to provide efficient service.

In order to efficiently manage data flows, it is necessary to estimate parameters of these flows to allocate appropriate amount of resources.

Manuscript received March 19, 2012; accepted May 14, 2012.
This work has been supported by the European Social Fund within the project „Support for the implementation of doctoral studies at Riga Technical University“.

II. P/M/1/K MODEL WITH ESTIMATION

This work is based on the P/M/1/K model from [3], which has been extended by adding self-similarity parameter estimation tool based on wavelet transform. The algorithm of this tool is proposed in [2] and it has been adopted for real-time use in Simulink environment.

For self-similar traffic generation most commonly used distribution is Pareto distribution [1], [8]. Its distribution function, in general has 3 parameters, however for network traffic generation the most commonly used form of Pareto distribution function [8] is

$$f(x) = \frac{\alpha}{\beta}\left(\frac{\beta}{x}\right)^{\alpha+1}, \tag{1}$$

where distribution parameters are calculated as follows for traffic with entity intensity λ and self-similarity H:

$$\alpha = 3 - 2H, \tag{2}$$

$$\beta = \frac{\alpha - 1}{\alpha\lambda}. \tag{3}$$

The diagram of P/M/1/K Simulink model with Wavelet-estimator tool is presented in Fig. 1. The model consists of one “Source” (Pareto distribution), which generates entities with intensity of $\lambda = 50$, one “FIFO Queue” of length K and one “Server” unit with service intensity $\mu = 62.5$ (Exponential distribution). There are also additional blocks, such as entity sinks to count arrived and lost entities, “Buffer” to accumulate entities for Hurst parameter estimation and output switch which directs data flow to sinks depending on Queue length limit K. There is also estimation block to calculate data packet loss probability.

Note, that all parameters of queuing model can be adjusted from “Model Parameters” block. The described above blocks are pure P/M/1/K model, which has been extended by wavelet-estimation block in bottom left corner of Fig. 1. The “Estimator” block is Embedded MATLAB Function block, which implements Hurst parameter estimation algorithm described in details in [2]. Model parameter T is an observation period, which can be divided in discrete sub periods Δt in which number of sent packets is being registered. There are recommendations available on choosing these values in [4], [5].

Authors of [2] have shown that the choice of specific wavelet for Hurst parameter estimation is based only on its number of vanishing moments *N*. The possibility to choose *N* value allows searching for trends and excluding their negative effects in data processing [2].

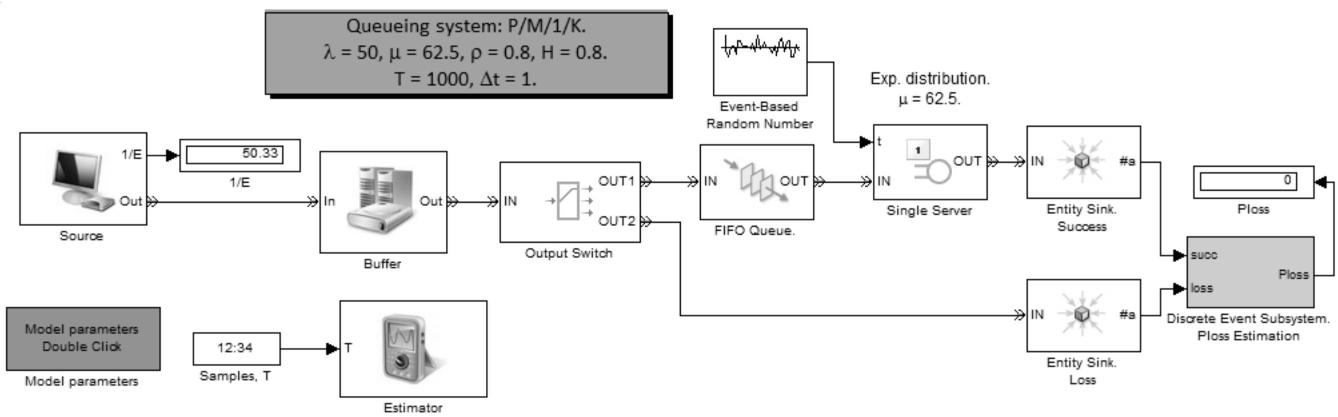

Fig. 1. The diagram of P/M/1/K Simulink model with Hurst parameter wavelet-estimator for traffic intensity of λ = 50, self-similarity parameter H = 0.8 and service intensity of μ = 62.5.

Theoretically, the higher is *N* value, the more accurate estimator can be created, however *N* value also increases computation costs, which affects its performance. The model described in this article is based on Daubechies-*N* wavelets, where *N* is determined by the number of wavelet vanishing points. This allows to easily increment *N* value by choosing the corresponding wavelet from Daubechies wavelet basis.

Note, that the algorithm from [2] can be improved to make more precise evaluations, as it has been done in [6]. The model being described in this article implements those improvements.

The algorithm proposed in [2], [6] calculates momentary Hurst parameter values over *T* sized packed groups by applying wavelet transform. According to [7], one should calculate average Hurst parameter over time instead of momentary values. This addition to the model has been made and wavelet-estimator outputs both these values in time. The averaged Hurst parameter is iteratively calculated by

$$\hat{H}_{avg}[n] = \frac{(n-1)\cdot \hat{H}_{avg}[n-1] + \hat{H}[n]}{n}, \qquad (4)$$

where $\hat{H}_{avg}[n]$ is averaged Hurst parameter estimate; $\hat{H}_{avg}[n-1]$ is previous averaged Hurst parameter estimate; $\hat{H}[n]$ is current momentary Hurst parameter estimate and *n* is the number of current estimate stored in counter. Note, that due to self-similar traffic nature, the estimation of self-similarity parameter can't be realized in situations when there is no data about entity count (a "pause"). In such cases instead of calculation according to (4) the previous averaged value remains the same

$$\hat{H}_{avg}[n] = \hat{H}_{avg}[n-1]. \qquad (5)$$

There is also another MATLAB program for Hurst parameter estimate calculation over entire entity count series (full traffic). This method is supposed to provide the highest estimation accuracy and will be used to compare the iteratively averaged value at the end of the entity count series.

III. Simulation Results

During the modelling process the Hurst parameter evaluates are saved to MATLAB workspace, which allows taking all benefits from Simulink integration into MATLAB.

These data arrays have been plotted as graphs – the set Hurst parameter value (source), momentary Hurst parameter value and averaged over time Hurst parameter value. The example for Wavelet transformation window length T = 1000 for self-similarity parameter H values of 0.6, 0.7, 0.8 and 0.9 are shown in Fig. 2–Fig. 5, respectively. The simulation time for all scenarios in this article is $T_{sim} = 50 \cdot T$, where T is the observation period in normalized time units. The length of one discrete sub period Δt = 1 and the number of entity count measurements in one window length is $M = T/\Delta t = 1000$.

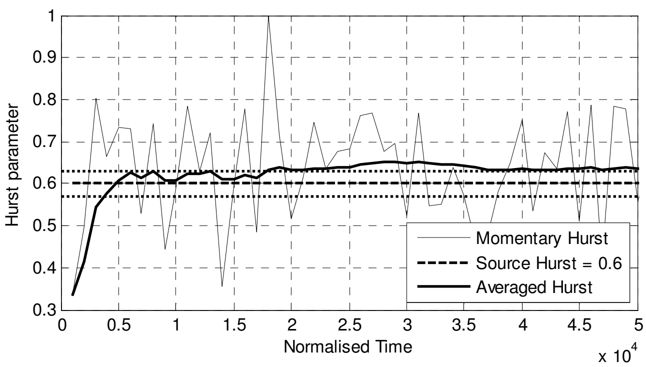

Fig. 2. Graphs of self-similarity parameter estimates for source generator set value of H = 0.6 and window length T = 1000.

It is possible to define a 5% deviation interval (or any other) to determine how fast the averaged value converges to actual Hurst parameter value set for generation of entities according to (2.1) and (2.2). These 5% deviation intervals are also shown in Fig. 2–Fig. 5.

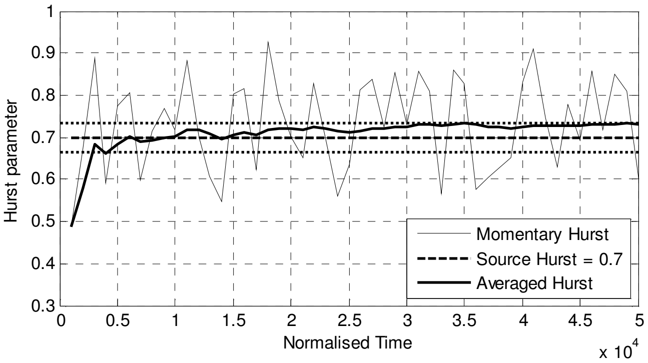

Fig. 3. Graphs of self-similarity parameter estimates for source generator set value of $H = 0.7$ and window length $T = 1000$.

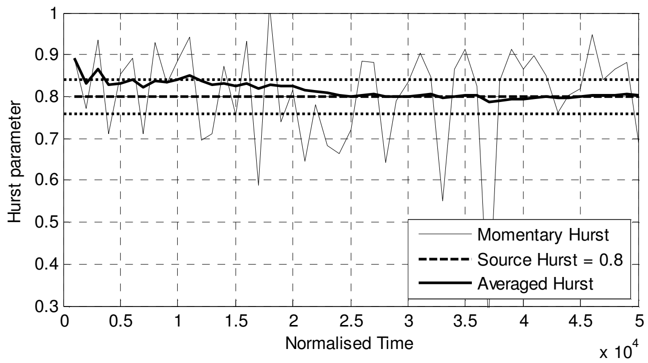

Fig. 4. Graphs of self-similarity parameter estimates for source generator set value of $H = 0.8$ and window length $T = 1000$.

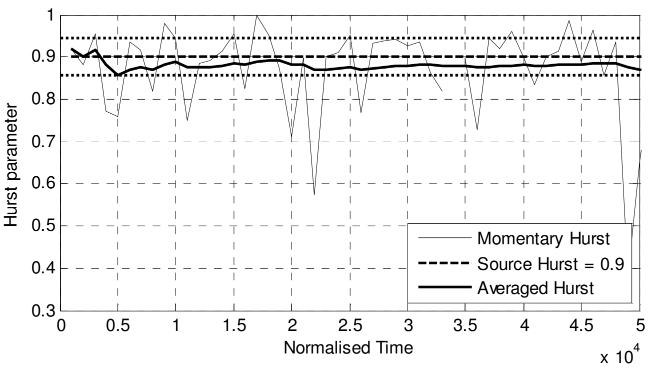

Fig. 5. Graphs of self-similarity parameter estimates for source generator set value of $H = 0.9$ and window length $T = 1000$.

The estimation error can be seen in Fig. 6 for all specified Hurst parameter values. The relative estimation error is calculated as percentage according to the following expression

$$\delta_{\hat{H}} = \frac{\left|\hat{H} - H\right|}{H} \cdot 100\%, \tag{5}$$

where H and $\hat{H}$ are actual source Hurst parameter value and its estimate, respectively. The 5% deviance interval has been marked as well for reference.

As it can be seen from Fig. 2–Fig. 5, deviation of momentary Hurst parameter evaluates from actual value is higher at lower parameter values. However, in all cases the averaged Hurst parameter value is converging to actual value after some interval, which doesn't seem to vary greatly for different Hurst parameter values. The exception can be seen in Fig. 2, where averaged Hurst parameter estimates leave the 5% deviation interval.

In Fig. 5 it can be noted, that roughly at the time interval t = (32500; 35000) the Hurst parameter can't be estimated. The estimation can be calculated incorrectly or even can't be calculated at all if there are no generated entities or the amount of such entities is insufficient to determine self-similarity parameter, i.e. the traffic intensity is very low. In such cases it's impossible to adequately define self-similarity parameter over fixed length window alone, and it could require information about the previous entities.

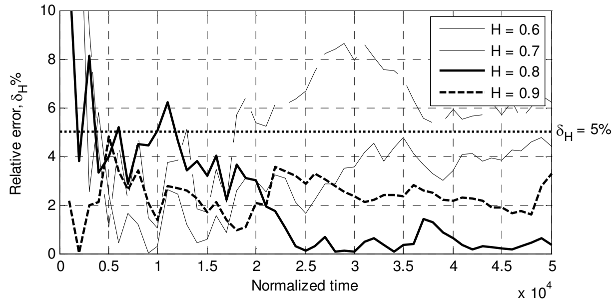

Fig. 6. Graphs of self-similarity parameter estimate relative error for different source generator set values H.

Thus, according to (4) and (5) expressions the averaged Hurst parameter value in such cases remains unchanged, which can be clearly seen from Fig. 5 at the corresponding time interval. Note, however, that in larger scale these pauses can contain information about process self-similarity, since entity inter arrival time has Pareto distribution with infinite variance.

The simulation results have been summarized in Table I. The table contains variance values of estimation deviances presented in Fig. 2–Fig. 5 for all specified source Hurst parameter values. The Table I also includes the last result for iteratively calculated by (4), (5) average Hurst parameter estimate and Hurst parameter single estimate over entire realization of generated traffic, calculated in post-processing mode by the same algorithm from [2], [6].

TABLE I. COMPARATIVE SIMULATION RESULTS FOR DIFFERENT SOURCE HURST PARAMETER VALUES.

Source Hurst parameter	Iteratively averaged Hurst parameter estimate		Full realization Hurst parameter estimate
	Average	Variance, σ^2	
$H = 0.6$	0.637	0.0030	0.641
$H = 0.7$	0.731	0.0016	0.729
$H = 0.8$	0.803	0.0004	0.799
$H = 0.9$	0.871	0.0001	0.863

As it can be seen from Table I data, the variance of estimated Hurst parameter values decreases along with Hurst parameter actual value increase. The estimates themselves however are at certain point contradictive. From the Table I data it can be clearly seen, that both for very high and low self-similarity traffic the short-term estimates give more precise result. On the contrary, for Hurst parameter values between these extreme values the long-term estimates achieve higher values.

It can also be seen from Table I, that for $H = 0.8$ both estimates give very accurate results, meaning both of these methods are almost equally effective for this specific self-similarity parameter value.

The further research is necessary to determine, whether it is possible to improve accuracy by implementing both of these methods in parallel computations.

There is also additional research required about choice of the number of entity count measurements in one window length $M = T / \Delta t$, especially for Hurst parameter values close to the mentioned above extreme values.

IV. CONCLUSIONS

The analysis of the modelling results shows that it is possible to use this model to estimate self-similarity parameter in real-time and make decisions on forecasted required memory allocation in traffic control systems. However, short-time evaluation can give inaccurate results, so it might be possible to improve accuracy by implementing long-time evaluation as well, which can be added in future work. The results from Table I suggest such possible solution.

Another possible solution for estimation accuracy improvement is to implement additional estimator with use of different transform – for example, Empirical Mode Decomposition (EMD). In particular, there is possibility to use Hilbert-Huang Transform (HHT) in conjunction with Wavelet Transform for traffic analysis as it has been shown in [9]. In future the estimator described in the article can be either extended, or modified accordingly.

Finally, the traffic parameters estimation system can be extended to estimate additional parameters, which would allow taking more accurate decisions on quality assurances for end users.

REFERENCES

[1] W. Leland, M. Taqqu, W. Willinger, D. Wilson, "On the Self-Similar Nature of Ethernet Traffic", *IEEE/ACM Transactions on Networking,* IEEE Press Piscataway, vol. 2, no. 1, pp. 203–213, 1994. [Online]. Available: http://dx.doi.org/10.1109/90.282603

[2] P. Abry, D. Veitch, "Wavelet-Analysis of Long-Range-Dependent Traffic", *IEEE Transactions on Information Theory,* IEEE Computer Society Washington, DC, USA, vol. 44, no. 1, pp. 2–15, 1998. [Online]. Available: http://dx.doi.org/10.1109/18.650984

[3] S. Sharkovsky, E. Grab, "Modeling Self-Similar Traffic in Networks", in *Proc. of the 52nd International Scientific Conference,* Latvia, Riga, 2011, pp. 19–23.

[4] M. Kulikovs, E. Petersons, S. Sharkovsky, "Integral measurement process of incoming traffic for measurement-based admission control", in *Proc. of the 2010 IEEE Region 8 International Conference on Computational Technologies in Electrical and Electronics Engineering*, IEEE Press Piscataway, pp. 183–186, 2010. [Online]. Available: http://dx.doi.org/10.1109/SIBIRCON.2010.5555296

[5] M. Kulikovs, S. Sharkovsky, E. Petersons, "Comparative studies of methods for accurate Hurst parameter estimation", *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering)*, no. 7, pp. 113–116, 2010.

[6] D. Veitch, P. Abry, M. Taqqu, "On the Automatic Selection of the Onset of Scaling", FRACTALS, vol. 11, no. 4, pp. 377–390, 2003.

[7] O. Sheluhin, S. Smolskiy, A. Osin, *Self-Similar Processes in Telecommunications*. John Wiley & Sons Ltd, 2007, p. 314.

[8] Ю. Лосев, К. Руккас, „Анализ моделей вероятности потери пактов в буфере маршрутизатора с учетом фрактальности трафика“, *Вестник Харьковского Национального Университета. Серия: Математическое моделирование. Информационные технологии. Автоматизированные системы управления,* Харьковский национальный университет им. В. Н. Каразина, no. 10, pp. 163–169, 2008.

[9] X. Cheng, K. Xie, D. Wang, "Network Traffic Anomaly Detection Based on Self-Similarity Using HHT and Wavelet Transform", in *Proc. of the 5th International Conference on Information Assurance and Security*, IEEE Press Piscataway, pp. 710–713, 2009. [Online]. Available: http://dx.doi.org/10.1109/IAS.2009.219