ATM Traffic Prediction Methods Using Wavelet Analysis

[1]Solis M. Barreto, P.A., [2]Dantas, M.J.P., [3]Lemos, R. P.

[1] Department of Computer Science, [2]Department of Mathematics and Physics
Catholic University of Goiás (UCG)
Av. Universitária 1069, Setor Universitário 74605-010 Goiânia, Goiás, Brazil
[3]Department of Electrical Engineering-Federal University of Goiás (UFG)
Praça Universitária s/n Setor Universitário 74605-220 Goiânia, Goiás, Brazil

Abstract-This work introduces the wavelet transform as an important element for ATM traffic prediction. Two methods are proposed. The first method proposes data fittings by the fGn model of parameter *H,* adequate for long dependence stationary processes. The estimated *H* is accomplished by a method based on Wavelet analysis. A small order Wiener filter is projected to implement the prediction and the final proposal results in a filter coefficient correction method that improves the prediction quality. The second method proposes a combination of wavelet transforms to feed forward artificial neural networks. Wavelet transforms are used to preprocess the nonlinear time-series in order to provide a step-closer phase learning paradigm to the artificial neural network. The network uses a variable length time window on approximation coefficients over all scales. This approach improves the generalization ability as well as the accuracy of the artificial neural network for ATM traffic prediction. Both prediction methods are evaluated on traffic data files from *Bellcore*.

Key words: ATM Traffic, prediction, wavelet analysis.

1 Introduction

During the last years, many researchers have observed the self-similarity of Asynchronous Transfer Mode (ATM) traffic. For the prediction process is of fundamental importance the certain modeling of the statistical multiplexing process, which promotes the traffic switching with variable bandwidths on different applications. Recently, several works modeled ATM traffic as a fractal phenomenon, with self-similarity and long dependence characteristics [1][2][3].

The multiplexing process of nodes in ATM networks uses a finite buffer, the long-time dependence produces delays and loss of cells, compromising the network quality service parameter and producing congestion situations. Then, the traffic prediction task with some antecedence may provide the network controllers a multiplexing adjustment to avoid or minimize these effects.

2 ATM traffic models

ATM traffic has the self-similarity property that establishes the preservation of the statistical properties in different time scales. The fractal geometry and the fractal process offer facilities for the understanding, modeling and analysis of self-similarity processes [4][10]. A stochastic process $x(t)$ defined in the time interval $(-\infty,\infty)$ for any real number $a>0$, is statistically self similar with Hurst parameter H if follows the relation :

$$x(t) =_d a^{-H} x(at) \tag{1}$$

where $=d$ denotes equality in the statistical sense. In the case of the ATM traffic, the self-similarity modeling helps to extract the traffic burstiness in different time scales. Previous studies have shown that in the case of ATM traffic, the Hurst parameter is $H>0.5$ [1].

Self-similar processes have been used to model the traffic fractal behavior [3][4], due to the fact that the increment processes of these models are long dependence stationary. These properties are also present in the incremental traffic process [1][2]. These processes are called stochastic processes $1/f$ and constitute an important model class for different applications in signal processing, pointed recently at applications in network traffic [7][9]. Two models are highlighted in the literature: the fractional Brownian motion (fBm) and the fractional Gaussian noise (fGn). For fBm and fGn definitions, see Appendix A.

In the signal processing analysis there are several mathematical tools for signal decomposition. The use of wavelets is especially interesting for signals that have non-stationary variations and self-similarity characteristics [5]. The wavelet transform functions make it possible to frame a signal in regions of variable size that allows the analysis of the signal in different time scales, producing an automatic mapping of a signal in a time-frequency plane. For wavelet transform definitions, see Appendix B.

3 The traffic prediction problem with wavelet analysis application

3.1 fGn model with wavelet analysis

With this method, the ATM traffic increments process is modeled as a fractional Gaussian noise with Hurst parameter H. The fGn model adjustment to the traffic increments requires the H parameter estimate for the measured traffic. The idea to use the details (*Wavelet* coefficients) for the H parameter estimate is interpreted as a generalization method proposed by Allan in 1996 [11]. He stats that there is an association between self-similarity associate variance and long dependence, introducing an estimator that improves the variance estimate where the term:

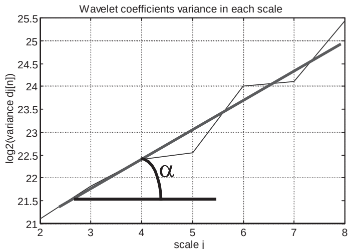
$$\frac{1}{N}\sum_{n}(X^{(T)}(n))^{2} \tag{2}$$

is substituted by "ALLAN's variance", defined as:

$$\frac{1}{N}\sum_{n}(X^{(T)}(n+1)-X^{(T)}(n))^{2} \tag{3}$$

In other words, there have been used dyadic bases for the wavelet representation, such that $X^{(T)}=2^{J}$, corresponds to the details, i.e. the *wavelet* coefficients. This improvement results from the fact that wavelets are capable of descorrelate long range dependence (LRD) processes, thus $\operatorname{var} d_{j,k} \approx 2^{j\alpha}$ where $\alpha = 2H\text{-}1$.

The H parameter estimate, using the method based on wavelet analysis [6] is obtained by adjusting a straight line to the points $\log_2$ (variance $d_{j,k}$) versus scale j, by the minimum squares method. The straight line inclination coefficient corresponds to an estimate of H, by Eq.4 as seen in figure 1.



$$\hat{H} = \frac{(\alpha + 1)}{2} \qquad \textbf{(4)}$$

Figure 1. *H* estimate for linear regression about the *wavelet* coefficients variances in the scales.

3.2 Artificial neural networks with wavelet analysis

Some work has been done for ATM traffic prediction using artificial neural networks (ANN)[14][15][16] ANNs are characterized by a non-parametric modeling where there is no need to understand the process statistical characteristics. During the learning process, the correlation between input and output variables is inferred, so there is no necessity to know the phenomenon statistical properties to be modeled. For signal processing some works include the use of neural networks as predictors. In [17], Liang and Page discuss about the learning capacity of an ANN and a multiresolution paradigm that decomposes the original signal in several versions using wavelet transforms looking for the improvement of the ANN generalization capacity.

Using the above work as a reference, the proposal of the predictor used in this work consists to input the ANN a set of training sets preprocessed using the wavelet transform. The training sets are formed with vectors of different sizes, unlike the traditional network learning which employs a single signal representation for the entire training process. The ANN will be trained with different scale versions of the original signal, trying to achieve a faster learning time and a less computational effort.

4 ATM traffic prediction methods

Given a time series with n samples, $\{x[i], i=1,2,\ldots n\}$, representing the number of cells produced by an ATM source in n discrete time intervals, the goal is to predict the series value in the time interval $n+k$, $k=1,2,\ldots,$ considering the series values until the time interval n. In order to compare the prediction quality, the normalized mean square error (NMSE) was adopted as a performance parameter. The NMSE is computed as:

$$NMSE = \frac{1}{\sigma^2 n} \sum_{k=1}^{n} [(x(k) - \hat{x}(k)]^2 \qquad \textbf{(5)}$$

where $x(k)$ is the observed value of the time series at time k, $x(k)$ is of predicted value of $\hat{x}(k)$ and σ^2 is the variance of the time series over the prediction duration. Thus, a NMSE=1 corresponds to predicting the estimated mean of the series.

4.1 fGn and Wiener predictor modeling

Before entering the filtering process, a decomposition of the signal using the wavelet transform leads to the H parameter estimate. From the fGn model autocorrelation functions $R_x[k]$ and the H parameter estimate, a Wiener filter of order M is projected. A recursive prediction process is executed to obtain the sequence next values $\{X[n]\}$, like described in Fig. 2.

Therefore the process is stationary, the Wiener filter outputs are optimal process estimates for the minimum square sense traffic increments [8]:

$$\hat{X}[n+k \mid n] = \sum_{m=1}^{M} h^{o}[m] X[n-m] \tag{6}$$

Optimal predictor coefficients are calculated from:

$$\mathbf{h}^{o} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xx} \tag{7}$$

where

$$\mathbf{r}_{xx} = [R_x(k), R_x(k+1), R_x(k+2), \ldots, R_x(k+M)]^T \tag{8}$$

and

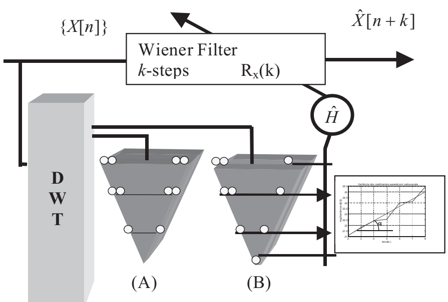
$$\mathbf{R}_{xx} = \begin{bmatrix} R_x(0) & R_x(1) & \ldots & R_x(M-1) \\ R_x(1) & R_x(0) & \cdots & R_x(M) \\ \vdots & \vdots & \ddots & \vdots \\ R_x(M-1) & R_x(M) & \ldots & R_x(0) \end{bmatrix} \tag{9}$$

Such that $R_x(.)$ is the autocorrelation function of the fGn model calculated from Eq.A-6 and Eq.A-7.

In order to compare the quality prediction the theoretical NMSE (e_T) is computed as:

$$e_T = 1 - \frac{\mathbf{h^o}^T \mathbf{r}_{xx}}{V_H} \tag{10}$$

and the normalized mean square error (e_S) is computed from Eq.5.



(A) Approximation coefficients
(B) Wavelet coefficients.

Figure 2. Traffic increments prediction algorithm with Wiener Filter projected by the fGn model and H parameter estimate using *wavelet* analysis.

The value of the H parameter estimate only depends on the variances from the decomposition wavelet scales. For a real time prediction, if exists a significant alteration of H, the Wiener filter coefficients are recalculated to continue the prediction process.

4.2 ANN predictor with phase learning.

The model consists of an ANN with a prior preprocessing of the signal using the Haar wavelet transform in the following manner: being X^s the original signal, $T_r(x^r)$ is the learning activity done in the x^r scale of the original signal X^r. The resulting x^r signal is composed only by the approximation coefficients. The details of the signal are discarded in each calculated scale. When r increases, the quantity of points of the original signal in the scale x^r diminishes in a proportion 2^r. Though, this can be interpreted as a visualization of the signal when fewer details and a lesser quantity of points. Interpreting the notation $T_j(x^j) \rightarrow T_i(x^i)$ as the training activity T_j with training set x^j preceding the training activity T_i with training set x^i, the training activities are given by the set $\{T_r(x^r),\ r=1,2,3,..n\}$, where $T_j(x^j) \rightarrow T_i(x^i)$ for $i=j-1$ e $n \geq j \geq 1$. Clearly appears that the set $T_0(x^0)$ denotes the training activity with the original signal. Figure 3 illustrates the model for the prediction task. The ANN weights are initialized just once during the first training activity (with the higher scale), afterwards they are just adjusted by the next learning activities with the lower scales.

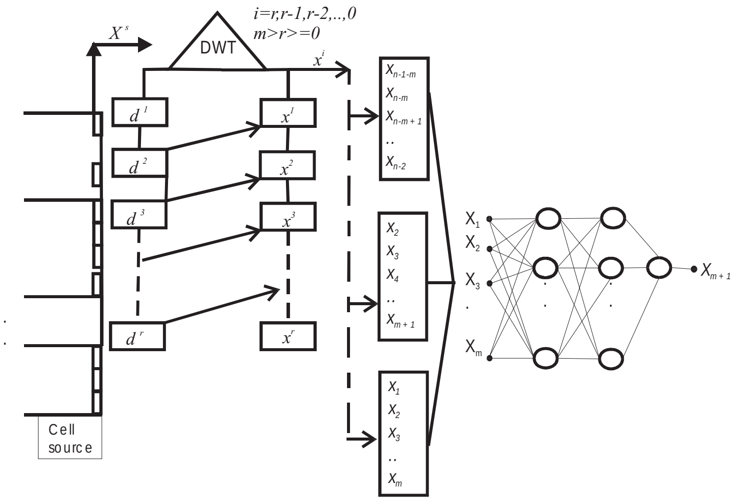


Figure 3. MLP network with DWT Haar signal preprocessing for ATM traffic prediction. The ANN input consists of m neurons, with two hidden layers and one output neuron. The vectors x^r, resulting of the approximation coefficients of Haar wavelet transform are the different input training sets.

5. Results

Due to unavailability of ATM network traffic data in operational networks, a traffic conversion procedure on *Bellcore* Ethernet traffic files was made, gathering the packages in regular time intervals and dividing the total byte number by the equivalent size of an ATM cell. Three resulting series were obtained: BC-OCTEXT, BC-OCTINT and BC-AUG. The series BC-OCTINT consists of 1,759 seconds, in a time scale of 1 second, corresponding to 29 minutes and 31 seconds of internal traffic. The series BC-OCTEXT consists of 122,797.83 seconds, corresponding to 1,000,000 of packets of external traffic in a minute scale, resulting in 2046 points corresponding to 34 hours and 10 minutes of traffic. The series BC-AUG consists of 3,142.82 seconds, corresponding to 1.000.0000 of Ethernet packets.

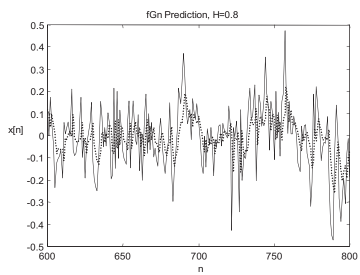
5.1 Results obtained with fGn and Wiener predictor modeling

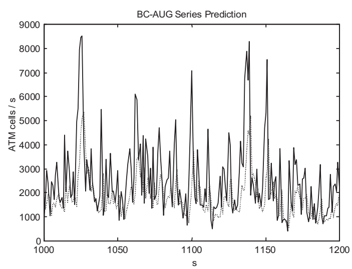
To separately evaluate the prediction quality and fGn adaptation to the traffic increments, the prediction of the time series BC-AUG (scale = 1s) was made, generated by the method *Random Midpoint Displacement* for $H = 0.8$. The parameter H was esteemed through wavelet analysis and the found value for the series BC-AUG resulted in 0.8107.

A comparison of the predicted curve and the real curve for fGn model for time series BC-AUG are illustrated in figure 4. The results of the experiments are summarized in the Tables 1(a) and 1(b). Predictors are evaluated with order $M = 6$ and $M = 100$ to 1 step and to 5-steps.

However the increments process of BC-AUG time series practically introduces the same H of fGn and it differs of a Gaussian process (see Fig. 5). This justifies that the simulated errors are greater than the theoretical ones. Soon the model does not adjust perfectly to the data, but a fractal model can be more difficult to treat. The signal prediction follows the original signal, but with a smaller amplitude. To improve the prediction, adjusting the amplitude and keeping the proportionality between filter coefficients whose sum equals to 1 induced the normalization of the optimal coefficients in the following form:

$$\mathbf{h}^{\mathbf{o}} = \frac{\mathbf{h}^{\mathbf{o}}}{\sum_{m} h^{o}[m]} \qquad \textbf{(11)}$$





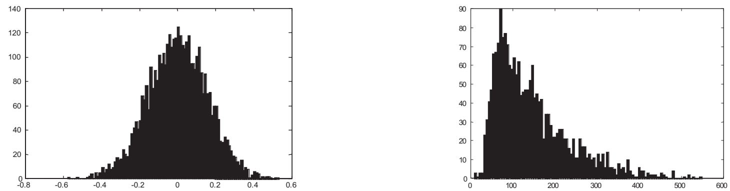
(a) (b)

Figure 4. (a) fGn Series (solid line) and prediction (dotted line), Wiener of order $M = 6$; (b) BC-AUG series (solid line) and prediction (dotted line), Wiener of order M=6.

Filter 1-step	fGn Series $H = 0.80$	BC-AUG Series $\hat{H} = 0.8107$	Filter 5-step	fGn Series $H = 0.80$	BC-AUG Series $\hat{H} = 0.8107$
M=6	$e_T = 0.7031$ $e_S = 0.7899$	$e_T = 0.6782$ $e_S = 0.8621$	M=6	$e_T = 0.8982$ $e_S = 0.9430$	$e_T = 0.8828$ $e_S = 1.5947$
M=100	$e_T = 0.6939$ $e_S = 0.7824$	$e_T = 0.6687$ $e_S = 0.6537$	M=100	$e_T = 0.8743$ $e_S = 0.9297$	$e_T = 0.8568$ $e_S = 0.9622$

(a) (b)

Table 1. Error comparison between fGn and BC-AUG prediction (a) Wiener of order $M = 6$ and $M = 100$ to 1 step; (b) Wiener $M = 6$ and $M = 100$ to 5 steps.

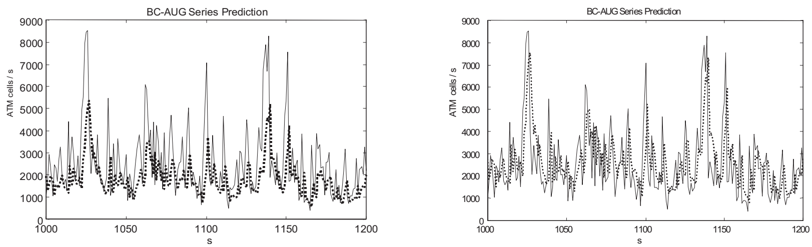


(a) (b)

Figure 5. (a) histogram of the fGn series (H=0.8) 4096 points and (b) of the BC-AUG series (s)($\hat{H}$ = 0.8107) 3142 points.

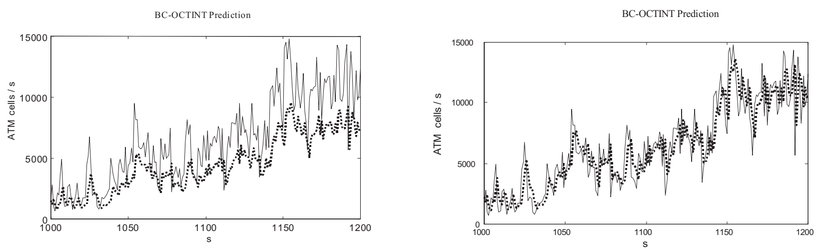
Figure 6 illustrates the improvement in the prediction with the adjustment in the great coefficients. The error obtained for BC-AUG prediction decreased from 0.8621(a) to 0.6736 in (b), well nearest to the theoretical error 0.6782.

In figure 7, for the BC-OCTINT series, scale of 1s ($\hat{H}$ = 0.8042), a theoretical error of 0.6934 was obtained, a simulated of 0.9140 before the adjustment and 0.3835 after (b). In a related work for the same time series [10], a LMS filter is implemented resulting in a error of 0.4769 with an order M = 20.



(a) (b)

Figure 6. Comparison between BC-AUG prediction: (a) without great coefficients adjustment, (b) with adjustment.



(a) (b)

Figure 7. Comparison between BC-OCTINT prediction: (a)without optimal coefficients adjustment, (b) with adjustment.

5.2 Results obtained with Artificial Neural Networks

In the case of the BC-OCTEXT time series, three sets were used: a training set, formed by elements 1 to 800, a validation set formed by elements from 801 to 1000 and a test set formed by elements 1001 to 2000. The prediction was made using a MLP network with m=10, two neurons in each hidden layer, resulting in a configuration of 10x2x2x1 and a learning rate of 0.01. In this case, the prediction quality was resulting NMSE=0.42506. Figure 8(a) shows the prediction result.

In the case of the BC-OCTINT time series, the training set was formed with 600 points (800 to 1499), the validation set from points 100 to 599 and the test set with points from 1500 to 1699. A MLP 20x10x10x1 was used with a learning rate of 0.01 and the prediction result quality was a NMSE value of 0.92008, which shows a low generalization capacity of the network, as seen in figure 8(b).

In the second experiment it appeared that the training sets could be reduced as well as the MLP complexity. Also, it was observed that the early stopping procedure adopted in the first experiment did not improve the prediction quality, so there was no need for defining a validation set. The training set for the time series BC-OCTEXT was formed with 64 points and the test set was formed with elements from point 1001 to 2000. The scale of decomposition using the predictor model described in figure 2 was $j=3$ resulting in 4 training sets x^3, x^2, x^1 and x^0, the latter corresponding to the original time series. The training sets are shown in figure 9. The MLP used was 4x2x2x1, resulting in a less computational effort for training. The prediction quality was NMSE=0.39951, as illustrated in figure 10.

In the case of the time series BC-OCTINT the training set was formed by the elements 1 to 256 and the test set with elements from point 1000 to 1699. It is important to note that in this second test set, there is an interval of abrupt variation that might increase the prediction difficulty. As with the time series BC-OCTEXT, the MLP used was a 4x2x2x1 and the prediction quality was NMSE=0.46195, showing a significant improvement. The prediction result is illustrated in figure 11.

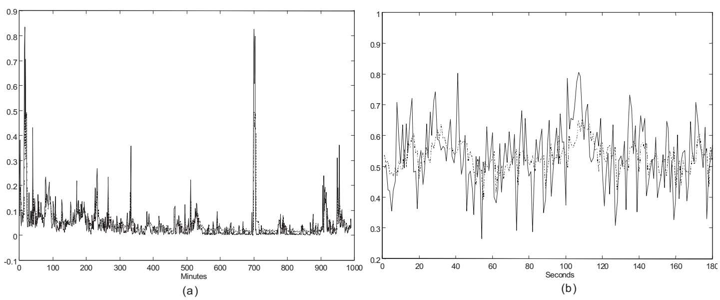


Figure 8. (a) BC-OCTEXT prediction with NMSE=0.42506. Real time series solid line, predicted time series dotted line. (b) BC-OCTINT prediction with NMSE=0.92008. Real time series solid line, predicted time series dotted line.

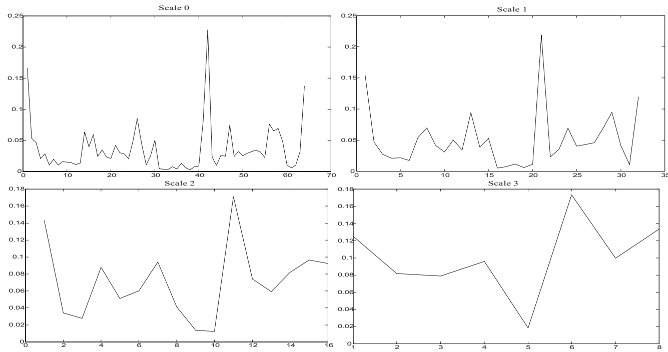


Figure 9. BC-OCTEXT: training sets, four scales x^o, x^1, x^2, x^3.

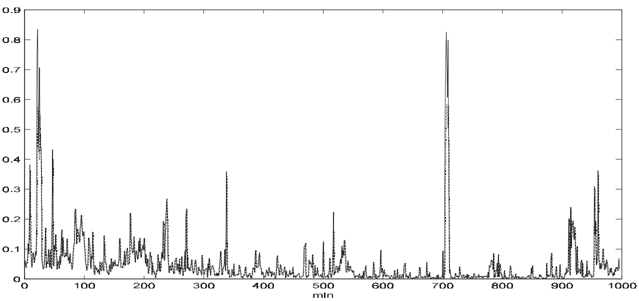


Figure 10. BC-OCTEXT MLP-wavelet transform prediction with NMSE=0.39951, real time series solid line, predicted time series dotted line.

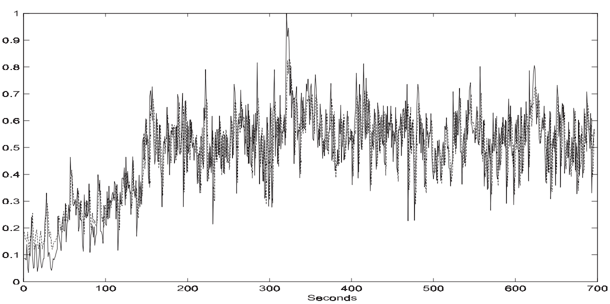


Figure 11. BC-OCTINT MLP-wavelet transform prediction with NMSE=0.46195, real time series solid line, predicted time series dotted line.

6. Conclusions

The fGn model adjustment of the traffic increments is not to perform and a juster model can lead to a more difficult treatment. The Wiener filter optimal coefficients normalization process results in a simulated error reduction, leading to a set of theoretical limits, validating the proposed method.

The conventional learning process in neural networks is sometimes inadequate for nonlinear, nonstationary signal prediction problems and often yields poor generalization performance. The combination of artificial neural networks and wavelet transforms for ATM traffic prediction permits a preprocessing of the signal that has a fractal behavior. The ATM traffic treated as a self-similar process, with same statistical information in different time scales induces the use of wavelet transforms to produce different training patterns with less or more details. The neural network learns gradually from simpler to more complex training sets introducing a phase learning paradigm. This process permits a neural network complexity reduction and the pursuit of predictors that have a better generalization capacity with a low computational cost that appears to perform well enough to be of practical value. The generalization of these results for other types of time series should depend on the compatibility of the wavelet transform and the process model that represents the time series.

Appendix A

Definition fBm: The discrete fractional Brownian motion (fBm) with parameter H, $0 < H < 1$ is a nonstationary stochastic process, $B_H[n]$, $n \in \mathbb{Z}$, Gaussian and H-sssi, such that $B_H[0] = 0$ with stationary increments and finite normal dimension distribution.

fBm has the following statistical properties:

Autocorrelation function:

$$R[n,k] = E\{B_H[n]B_H[n]\} = \frac{\sigma^2}{2}\left[|n|^{2H} + |k|^{2H} - |n-k|^{2H}\right] \qquad \textbf{(A-1)}$$

that reflects the nonstationarity of the process, variant with temporal displacement.

Variance:

$$Var\{B_H[n]\} = \sigma^2 |n|^{2H} \qquad \textbf{(A-2)}$$

Power density is given empirically by:

$$S_{BH}(\omega) = \sigma^2/|\omega|^{2H+1} \qquad \textbf{(A-3)}$$

Equation A-3 can be interpreted as a spectrum generalized of the self-similarity of the form:

$$B_H[n] = a^{-H} B_H[an-b], \quad a > 0 \text{ e } \forall b \qquad \textbf{(A-4)}$$

The fractional Brownian motion increment process is called fractional Gaussian noise (fGn) and it introduces a long term and slow decay dependence variance (high variability), when $½ < H < 1$.

Definition fGn: $X[n]$, $n \in \mathbb{Z}_+$ is a fractional Gaussian noise with parameter H and corresponds to the increments process of a fractional Brownian motion(fBm) with parameter H with the form:

$$X[n] = B_H[n] - B_H[n-m] \qquad \textbf{(A-5)}$$

Evaluating fGn statistical properties shows that the autocorrelation function is given by:

$$R_x(k) = \frac{V_H}{2}\left(|k-1|^{2H} + |k+1|^{2H} - 2|k|^{2H}\right) \quad \textbf{(A-6)}$$

where:

$$V_H = \Gamma(1-2H)\frac{\cos \pi H}{\pi H} \quad \textbf{(A-7)}$$

Appendix B

Wavelet analysis

In the wavelet analysis, the low signal frequencies are identified as the approximation coefficients and the high frequencies are identified as the details. This can be seen as the original signal passing in a filtering process, with a low-pass filter producing the approximation coefficients and a high-pass filter producing the details. Given a scale function ϕ^o and a basic wavelet ψ^o, the discrete wavelet transform (DWT) is defined as:

$$X(t) \rightarrow \{\{a_{j,k}, k \in Z\}, \{d_{jk}, j = 1,2,....,J, k \in Z\}\} \quad \textbf{(B-1)}$$

The coefficients are defined as the internal product between X and two sets of functions:

$$a_{jk} = < X, \phi^0{}_{j,k} > \qquad d_{jk} = < X, \psi^0{}_{j,k} > \quad \textbf{(B-2)}$$

$\psi^0_{j,k}$ (respectively $\phi^0_{j,k}$) representing translations and dilations of ψ^0 (respectively ϕ^0), called basic wavelet function. Mallat [18] introduced a pyramidal algorithm for computing wavelet transforms by using the wavelet coefficients as filter coefficients. For the composition the algorithm employs a lowpass filter *L* and a highpass filter *H*, as illustrated in figure B-1 for a Haar wavelet system with J levels.

The multiresolution analysis (MRA) resides in the computation of a wavelet system for the decomposition and reconstruction of a signal *x(t)* using an orthonormal base. There are two fundamental functions used to obtain a wavelet system: the scale function and the basic wavelet function, as seen in Eq. B-3 Eq. B-4, respectively, where Z are the integer set and a_k are the wavelet coefficients. Both functions are prototypes of a class of orthonormal functions.

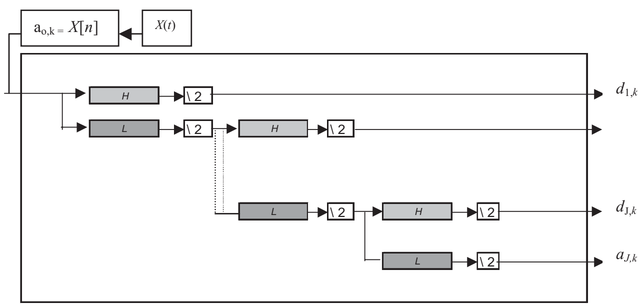


Figure B-1: Mallat's pyramidal algorithm (DWT HAAR , J levels)

$$\phi(t) = \sum_{k \in Z} a_k \phi(2t-k) \quad \textbf{(B-3)}$$

$$\psi(t) = \sum_{k \in Z} (-1)^k a_{k+1} \phi(2t-k) \quad \textbf{(B-4)}$$

7. Bibliographic References

[1] Leland, W., Taqqu,W. and Wilson, D., "On the self-similar nature of Ethernet traffic". *IEEE/ACM Transactions on Networking*, 2(1): 1-15, 1994.

[2] Beran, J., Sherman, R., Taqqu, M. and Wilinger, W., "Long range-dependence in variable-bit-rate video traffic". *IEEE Transactions on Communications*, 13(6): 400-410, 1995.

[3] Vidács, A., "Self-Similar Traffic Modeling Techniques in ATM Networks". *Master's thesis*, Technical University of Budapest, 1996.

[4] Patel, A. and Williamson, C., "Statistical multiplexing of self-similar traffic: theoretical and simulations results" 1997.

[5] Wornell, G., *Signal Processing with fractals: A wavelets Based Approach*, Prentice Hall, New Jersey, 1996.

[6] Abry, P. and Veitch, D., "Wavelets analysis of long-range dependent traffic". *IEEE Transactions on Information Theory*, : 2-15, 1998.

[7] Hirchoren, G. and Arantes, D.S., "Predictors for the discrete-time fractional Gaussian process". *IEEE Transactions on Signal Processing*, : 49-53, 1998.

[8] Haykin, S., *Modern Filters*. Macmillan Publishing Company, New York, 1989.

[9] Norros, I., "On the prediction of fractional Brownian motion". *Journal of Applied Probability*, 33: 400-410, 1996.

[10] Carvalho, M.M., "Prediction of Self-similar Traffic in Large Band Networks", Master´s *Degree Thesis,*, Unicamp, 1998.

[11] Allan, W., "Statistics of Atomics Frequency Standards", In Proc. IEEE Vol. 54, 221-239, 1996.

[12] Ryu B.K., Elwalid A.: "The importance of Long-Range Dependence of VBR Video Traffic in ATM Traffic Engineering Myths and Realities.", ACM SIGCOMM, San Francisco CA, 1996.

[13] Nguyen D., Widrow B.: "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights.", Proceedings of the International Joint Conference on Neural Networks, Volume 2. 357-363, 1993.

[14] Drossu R., Lakshman T.V., Obradovic Z., Raghanvendra C.: Single Multiple Frame Video Traffic Prediction Using Neural Networks Models. Computer Networks, Architecture and Applications. S.V. Raghavanan, B.N. Jain Editors. Chapmann and Hall) 146-158, 1995.

[15] Nordstrom E., Galmo O., Gustaffson M., Asplund L. :"Neural Networks for Admission Control in ATM Networks.", Proceedings of the First Swedish National Conference on Connectionism. Skivde, Sweden, 1992.

[16] Lobejko W. "Traffic Prediction by Neural Approach.", IEEE Transactions on Signal Processing, Vol 45 N.11, 571-575, 1996.

[17] Liang Y., Page E.:"Multiresolution Learning Paradigm and Signal Prediction.", IEEE Transactions on Signal Processing, Vol 45.,2858-2864, 1997.

[18] Mallat S.G., "Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$", Trans. Amer. Math. Soc. Vol 315, 69-87, 1989.