Suppression of Boundary Effect and Introduction of Scale Correlation for Wavelet based Traffic Prediction

Naoya Matsusue†, Hiroshi Hasegawa†, and Ken-ichi Sato†
† Dept. of Electrical Engineering and Computer Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-8603 Japan
E-mails: n_matusu@echo.nuee.nagoya-u.ac.jp, hasegawa@nuee.nagoya-u.ac.jp, sato@nuee.nagoya-u.ac.jp

***Abstract*— In this paper, we propose a Wavelet-based prediction method of the Internet traffic volume. By introducing maximal overlap formulation of the Haar wavelet, the proposed method is free from so-called boundary condition, which arises from processing delay of analysis filters of the Wavelet transform and refrains from full utilization of information on recent input signals. The proposed method is based on a vector autoregressive model so as to introduce inter-scale correlation of Wavelet coefficient serieses at different scales.**

I. Introduction

Recently traffic volume in the Internet is rapidly growing(For example, see [1]). Further traffic expansion will occur in the near future with the introduction of new broadband services including high definition TV (1k × 2k pixels) and super-high definition TV (2k × 4k pixels). This growth is much faster than that of the electrical processing and routing capability at network nodes, roughly doubles every eighteen-months, i.e. Moore's law. Thus photonic network technologies [2, 3] are being widely studied and dynamic control schemes of such networks [4–6] are developed for better network resource utilization and higher throughput. The effectiveness of the dynamic control heavily depends on accuracy of prediction of traffic volume since some operational delay is necessary.

Time series with long-memory characteristic [7, 8] can be found in several applications including analysis and prediction of network traffic volume. It is known that this characteristic makes it difficult to analyze stochastic properties of the time series and can be concealed into vector series given by the Wavelet transform [7]. Accuracy of analysis/prediction is enhanced if a prediction algorithm, for example adaptive signal processing technique, is applied to the vector series instead of the original time series [9, 10]. Then through the inverse Wavelet transform an estimate for the next time sample is derived. These Wavelet-based prediction methods suffer from so-called *boundary condition* [7] due to processing delay of analysis filters between inputs and outputs. That is, information of future input signals is necessary to derive a vector-valued coefficient at this moment, and this effect becomes severe as the analysis filter becomes longer. Conventional traffic prediction method [10] resolves this drawback by ignoring ambiguous current coefficients and predicting a coefficient that is two or more time samples ahead. Therefore, the method does not fully utilize information of recent inputs.

In this paper, we propose a traffic prediction method that is free from the boundary effect. Firstly, we introduce the Haar wavelet because of its shortest support of analysis filters. By combining maximal overlap formulation [7] with the Haar wavelet, it is shown that the vector-valued coefficient at this moment is derived without assuming information of the future input. In addition, after predicting coefficient at the next time sample, estimate of traffic volume is immediately derived as the sum of all components of the predicted vector-valued coefficient. Because of poor frequency response of the Haar wavelet, the components of vector series are strongly correlated each other. To fully utilize the correlation between components, instead of applying adaptive estimation technique [11] (ex. acoustic echo cancellation based on LMS, NLMS, etc.), we introduce a vector autoregressive model to predict the vector-valued coefficient. Finally we apply the proposed method to actual data of the Internet traffic volume and verify its applicability.

II. Preliminaries

Let the set of all real numbers and integers be $\mathbb{R}$ and $\mathbb{Z}$, respectively. For orthonormal analysis filters, highpass filter $(h[n])_{n\in\mathbb{Z}} \subset \mathbb{R}$ and lowpass filter $(g[n])_{n\in\mathbb{Z}} \subset \mathbb{R}$, respectively stand for wavelet function and scaling function, we have scaling coefficients and wavelet coefficients at ℓ th level ($\ell = 1, \ldots, L$) of an input signal $s : \mathbb{Z} \to \mathbb{R}$ as

$$\left.\begin{array}{rcl} c_\ell[n] & := & \displaystyle\sum_{k=-\infty}^{\infty} g[k]c_{\ell-1}[n-2^{\ell-1}k] \\ d_\ell[n] & := & \displaystyle\sum_{k=-\infty}^{\infty} h[k]c_{\ell-1}[n-2^{\ell-1}k] \end{array}\right\} \quad (n \in \mathbb{Z})$$

where $c_0 = s$ and L is maximum decomposition level. The above formulation is called *maximal overlap discrete wavelet transform* [7, Sec.5]. With this maximum overlap formulation, we can derive wavelet/scaling coefficients defined over all discrete time samples whereas usual decimated wavelet transform [7, 12] provides coefficients on a sparse grid $2^\ell\mathbb{Z}$ and is effective for several applications such as data compression.

Assume that power spectrum of the input s is given by $S_X(f)$. Then inter-level correlation between wavelet coefficients at ℓ and ℓ' th levels is [7]

$$\begin{array}{rl} & \mathrm{Cov}\{d_\ell[n], d_{\ell'}[n']\} \\ = & \displaystyle\int_{-1/2}^{1/2} e^{i2\pi f(2^{\ell'}(n'+1)-2^\ell(n+1))} H_\ell(f)H_{\ell'}(f)S_X(f)df \end{array} \quad (1)$$

where

$$\begin{array}{rcl} G_\ell(f) & := & \displaystyle\prod_{k=0}^{\ell-1} G(2^k f) \\ H_\ell(f) & := & H(2^{\ell-1}f)G_{\ell-1}(f) \quad \left(\dfrac{1}{2} \le f \le \dfrac{1}{2}\right) \end{array}$$

with the Fourier transform of the analysis filters H and G. Obviously perfect decorrelation is achieved when H is

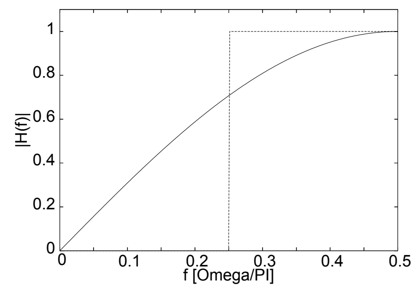


Fig. 1. Frequency responses of the Haar wavelet and the ideal filter

the ideal highpass filter. Indeed, for analysis of fractionally difference process by Daubechies wavelets, convergence to the ideal decorrelation, as the filter becomes longer, is shown in [13]. However the ideal filter requires infinite support and application of such filter is not practical.

On the other hand, wavelet transform with long analysis filters is severely affected by so-called boundary conditions [7, Sec.4.11 and Sec.5.11] that means we must assume some future inputs to derive current coefficient. Only the Haar wavelet, $(h[0], h[1]) = \frac{1}{2}(1, -1)$ and $(g[0], g[1]) = \frac{1}{2}(1, 1)$, is free from this effect. However, its frequency response is too poor to decorrelate the input sequence (See Fig.1).

III. Proposed Traffic Prediction based on the Maximal Overlap Haar Wavelet Transform and Vector Autoregressive Model

As we discussed in the previous section, the boundary condition can be resolved by use of the Haar wavelet. To tackle the problem of its poor decorrelation performance, we formulate the problem as a prediction based on vector autoregressive model and introduce the maximal overlap formulation to derive dense coefficients. Let a $L+1$ tuple vector of a scaling coefficient and wavelet coefficients, of the maximal overlap Haar wavelet at n th time sample, be

$$\begin{aligned} \boldsymbol{x}[n] &:= (c_L[n], d_1[n], \ldots, d_L[n])^T \\ &\quad - \mathrm{E}[(c_L, d_1, \ldots, d_L)^T] \quad (n \in \mathbb{Z}) \end{aligned} \tag{2}$$

where $\mathrm{E}[\cdot]$ stands for expectation. In practical applications, $\mathrm{E}[\cdot]$ would be approximated by ensemble average over some interval and we also employ this estimation scheme later. Let autocorrelation matrix be $R[m] := \mathrm{E}[\boldsymbol{x}[n]\boldsymbol{x}^T[n-m]]$. Assume that the vector series $(\boldsymbol{x}[n])_{n\in\mathbb{Z}}$ is generated by the following vector-valued linear process

$$\boldsymbol{x}[n] = \epsilon_n + \sum_{k=1}^{\infty} B_k \epsilon_{n-k}$$

where $(\epsilon_n)_{n\in\mathbb{Z}} \subset \mathbb{R}^{J+1}$ is a sequence of i.i.d. random vectors with zero mean and positive definite covariance matrix. Then the minimum mean square error linear predictor of $\boldsymbol{x}[n]$ based on $\boldsymbol{x}[n-1], \ldots, \boldsymbol{x}[n-M]$ is given by [14]

$$\widehat{\boldsymbol{x}}[n+1] = \sum_{m=1}^{M} A_m \boldsymbol{x}[n-m+1] \tag{3}$$

where $(A_m)_{m=1}^M \subset \mathbb{R}^{(L+1)\times(L+1)}$ satisfy the multivariate "Yule-Walker" equations

$$(A_1, A_2, \ldots, A_M) = G^T \Gamma^{-1} \tag{4}$$

with

$$G := \begin{bmatrix} R[1] \\ R[2] \\ \vdots \\ R[M] \end{bmatrix} \in \mathbb{R}^{M(L+1)\times(L+1)}$$

and

$$\Gamma := \begin{bmatrix} R[0] & R[-1] & \cdots & R[-M+1] \\ R[1] & R[0] & \cdots & R[-M+2] \\ \vdots & & \ddots & \vdots \\ R[M-1] & R[M-2] & \cdots & R[0] \end{bmatrix} \in \mathbb{R}^{M(L+1)\times M(L+1)}.$$

Summarizing the above discussion, we have the next prediction algorithm.

Algorithm 1: *For given traffic sequence $(s[n])_{n\in\mathbb{Z}}$, derive a vector-valued sequence $(\boldsymbol{x}[n])_{n\in\mathbb{Z}}$ in Eq.(2) through the maximal overlap Haar wavelet transform and estimation of expectation $\widehat{\mathrm{E}}[(c_L, d_1, \ldots, d_L)^T]$ by computing ensemble average over properly selected interval. After estimation of $(A_m)_{m=1}^M$ by Eq.(4), coefficients at the next time sample $\widehat{\boldsymbol{x}}[n+1]$ is predicted by Eq.(3). With a vector of coefficients*

$$\begin{aligned} &\left(\widehat{c_L}[n+1], \widehat{d_1}[n+1], \ldots, \widehat{d_L}[n+1]\right)^T \\ := \;& \widehat{\boldsymbol{x}}[n+1] + \widehat{\mathrm{E}}[(c_L, d_1, \ldots, d_L)^T], \end{aligned}$$

traffic at the next time sample is derived through the inverse Haar transform:

$$\widehat{s}[n+1] = \widehat{c}_L[n+1] + \sum_{k=1}^{L} \widehat{d_k}[n+1].$$

□

Remark 1:

1) With the predicted volume by the algorithm, dynamic network control scheme will setup/remove optical paths which are tunnels directly connect source and destination nodes with huge capacity (ex. 10-40Gbps). Evaluation of improvement in resource utilization is the next problem.
2) It would suffice to update $(A_m)_{m=1}^M$ and $\mathrm{E}[(c_L, d_1, \ldots, d_L)^T]$ for every time intervals, not every time samples. Moreover, we numerically verified that some delay for the update can be acceptable. Thus, the prediction and the update can be done in parallel and would be implemented on a pair of relatively slow processors. □

IV. Numerical Experiment

We employ a traffic data from *NLANR PMA*[15] (the detail is shown in Table I). The data is originally given as a list where each item stands for size of each data packet and its arrival time. We derived a step function of total size of packets arrived within ten seconds.

TABLE I

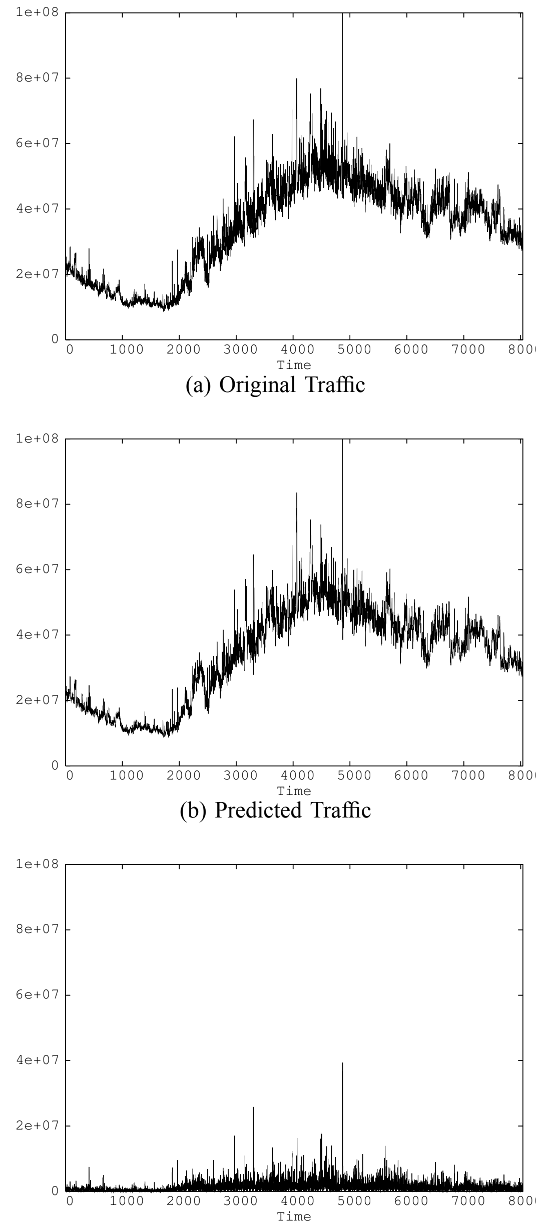
SPECIFICATION OF TRAFFIC EMPLOYED FOR NUMERICAL EXPERIMENT

Location	University of Leipzig Internet access link
Capacity	OC3 Packet-over-Sonet 155.53 Mbits/s
Date	1 – day (Nov 22 2002)
Accuracy	guaranteed to be less than 1 microsecond to UTC

TABLE II

EVALUATION OF RESIDUAL ERROR

NMSE	0.0038897
(Square root of NMSE)	(0.062367)
MRE	0.042424



(a) Original Traffic

(b) Predicted Traffic

(c) Prediction Error

Fig. 2. Result of Traffic Prediction

For the proposed method, the degree $M = 2$ is estimated through application of the *Akaike Information Criterion*. The level of wavelet decomposition is selected to $L = 3$. The coefficients (A_1, A_2) of vector autoregressive model and the expectation $\mathrm{E}[c_2, d_1, d_2, d_3]$ are periodically computed for every 600 samples. First 600 samples of the step function is used only for the estimation of parameters. The result of prediction for the rest are shown in Fig. 2. We can observe that the original traffic (Fig.2(a)) and the predicted traffic (Fig.2(b)) have similar shape. Following the evaluation scheme in [10], we introduce two measurements to evaluate the residual error in Fig.2(c):

$$e_{\text{NMSE}} := \frac{\sum_{n\in N} (s[n] - \widehat{s}[n])^2}{\sum_{n\in N} s^2[n]}$$

$$e_{\text{MRE}} := \frac{1}{|N|} \sum_{n\in N} \left| \frac{s[n] - \widehat{s}[n]}{s[n]} \right|$$

where $N \subset \mathbb{Z}$ is the set of indexes and $|N|$ stands for the number of the indexes. The result is shown in Table II. The square root of NMSE and MRE imply that estimation error at each time sample is approximately $\frac{1}{20}$ of magnitude.

ACKNOWLEDGMENT

This work is partially supported by NICT (National Institute of Information and Communications Technology).

REFERENCES

[1] Japan Internet Exchange, http://www.jpix.ad.jp/en/techncal/traffic.html

[2] K. Sato, *Advances in Transport Network Technologies: photonic networks, ATM, and SDH, Artech House*, Norwood, 1996.

[3] K. Sato, S. Okamoto, and H. Hadama, “Network Performance and Integrity Enhancement with Optical Path Layer Technologies,” IEEE J. Sel. Areas Commun., vol.12, no.1, pp.159-170, Jan. 1994.

[4] M. Kodialam and T. V. Lakshman, “Dynamic Routing of Locally Restorable Bandwidth Guaranteed Tunnels using Aggregated Link Usage Information,” *Proc. IEEE INFOCOM 2001*, pp.376–385, Apr. 2001.

[5] K. Shiomoto, E. Oki, W. Imajuku, S. Okamoto, and N. Yamanaka, “Distributed Virtual Network Topology Control Mechanism in GMPLS-based Multiregion Networks,” *IEEE J. Sel. Areas Commun.*, vol.21, no.8, pp.1254–1262, Oct. 2003.

[6] E. Oki, K. Shiomoto, D. Shimazaki, N. Yamanaka, W. Imajuku, and Y. Takigawa, “Dynamic Multilayer Routing Schemes in GMPLS-based IP+Optical Networks,” *IEEE Commun. Mag.*, pp.108–114, vol.43, no.1, Jan. 2005.

[7] D. B. Percival, A. T. Walden, *Wavelet Methods for Time Series Analysis,* Cambridge University Press, 2000.

[8] M. Krishna.P, V. M. Gadre, and U. B. Desai, *Multifractal based Network Traffic Modeling*, Kluwer 2003.

[9] L. Hong, G. Chen, and C. K. Chui “A Filter-Bank-Based Kalman Filtering Technique for Wavelet Estimation and Decomposition of Random Signals,” *IEEE Trans. on Circ. & Syst. II*, vol.45, no. 2, pp.237–241, Feb. 1998.

[10] Y. Luan, “Multiresolutional Traffic Prediction: Combine RLS Algorithm with Wavelet Transform,” *Proc. on ICOIN2005*, pp.321–331, 2005.

[11] S. Haykin, *Adaptive Filter Theory*, 2nd ed., Prentice-Hall 1991.

[12] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd.ed, Academic Press, 1999.

[13] Y. Fan, “On the Approximate Decorrelation Property of the Discrete Wavelet Transform for Fractionally Difference Processes,” *IEEE Trans. on Inf. Theory,* vol.49, no.2, Feb. 2003.

[14] R. Lewis and G. C. Reinsel, “Prediction of Multivariate Time Series by Autoregressive Model Fitting,” *Journal of Multivariate Analysis*, vol.16, no.3, pp.393–411, 1985.

[15] NLANR PMA, http://pma.nlanr.net/