PAPER

A Timescale Decomposition Approach to Network Traffic Prediction*

Guoqiang MAO[†a)],

SUMMARY The presence of the complex scaling behavior in network traffic makes accurate traffic prediction a challenging task. Some conventional prediction tools such as the recursive least square method are not appropriate for network traffic prediction. In this paper we propose a timescale decomposition approach to real time traffic prediction. The raw traffic data is first decomposed into multiple timescales using the *à trous* Haar wavelet transform. The wavelet coefficients and the scaling coefficients at each scale are predicted independently using the ARIMA model. The predicted wavelet coefficients and scaling coefficient are then combined to give the predicted traffic value. This timescale decomposition approach can better capture the correlation structure of the traffic caused by different network mechanisms, which may not be obvious when examining the raw data directly. The proposed prediction algorithm is applied to real network traffic. It is shown that the proposed algorithm outperforms traffic prediction algorithms in the literature and gives more accurate results.

key words: *traffic prediction, wavelet, timescale, traffic scaling*

1. Introduction

It is well known that some characteristics of Internet traffic fall beyond the conventional framework of Markov traffic modeling. Leland et al. demonstrated self-similarity in a LAN environment (Ethernet) [1]. Paxson et al. showed self-similar burstiness manifesting itself in pre-World Wide Web WAN IP traffic [2]. Beran et al. demonstrated self-similairty in variable-bit-rate (VBR) video traffic [3] and Crovella et al. showed self-similarity for WWW traffic [4]. Recent measurements and simulation studies further revealed that wide area network traffic has complex multifractal characteristics on small timescales, and is self-similar on large timescales [5], [6]. Cao et al. [7] and Furuya et al. [8] analyzed the small timescale behavior of traffic and found that WAN traffic tends to be Poisson on small timescales due to statistical multiplexing.

Accurate forecasting of the traffic is important in the planning, design, control and management of networks. Traffic prediction at different timescales has been used in various fields of networks, such as long-term traffic prediction for network planning, design and routing; and real-time traffic prediction for dynamic bandwidth allocation, and predictive and reactive traffic and congestion control. The presence of the complex scaling behavior makes the accurate forecasting of network traffic a challenging task. An implication of the self-similarity in network traffic is that the autocorrelation function $r(k)$ of the traffic rate decays hyperbolically rather than exponentially fast:

$$r(k) \sim C_r k^{-\beta}\ ,\ 0 < \beta < 1 \qquad (1)$$

where C_r is a positive constant and β is related to the Hurst parameter by $H = 1 - \beta/2$. Hurst parameter is a measure of the self-similarity. As a result the autocorrelation function is non-summable, i.e. $\sum_k r(k) = \infty$. Together with the slowly-decaying autocovariance function and the traffic non-stationarity, it suggests that some conventional prediction tools such as the recursive least square method [9] are not appropriate for network traffic prediction [10].

In this paper, we shall present a timescale decomposition approach to traffic prediction. In addition to the characteristics of the applications generating the traffic, traffic variations at different timescales are caused by different network mechanisms. Traffic variations at small timescales (i.e. in the order of ms or smaller) are caused by buffers and scheduling algorithms etc. Traffic variations at larger timescales (i.e. in the order of 100ms) are caused by traffic and congestion control protocols, e.g. TCP protocols. Traffic variations at even larger timescales are caused by routing changes, daily and weekly cyclic shift in user populations. Finally long-term traffic changes are caused by long-term increases in user population as well as increases in bandwidth requirement of users due to the emergence of new network applications. This fact motivates us to decompose the traffic into different timescales and predict traffic independently at each timescale. The proposed timescale decomposition approach to traffic prediction allows us to explore the correlation structure of network traffic at different timescales caused by different network mechanisms, which may not be easy to investigate when examining the raw data directly.

The rest of the paper is organized as follows: in section 2, we shall review existing work in the area; in section 3, we shall introduce the use of the *à trous* Haar wavelet transform to decompose the traffic into different timescales; in section 4 the prediction algorithm will be introduced; simulation results using real traffic

[†]The author is with the School of Electrical and Information Engineering, the University of Sydney

a) E-mail: guoqiang@ee.usyd.edu.au

*This research is supported by Australian Research Council Discovery Project DP0559248.

traces are given in section 5 and finally some conclusions and further work are summarized in section 6.

2. Related Work

Some algorithms have been proposed in the literature for real-time traffic prediction, which include traffic prediction based on the FARIMA (fractional autoregressive integrated moving average) model [11], the neural network approach [12], [13] and method based on the α-stable model [14], [15], etc. Traffic prediction based on the FARIMA model relies on the accurate estimation of the Hurst parameter. Despite a number of estimators reported in the literature, the accurate estimation of the Hurst parameter remains a difficult problem even in off-line conditions [16]. The presence of non-stationarity and complex scaling behavior in network traffic makes the situation even worse. Therefore, traffic prediction based on the FARIMA model is not suitable for real applications. Traffic prediction using the neural network approach can be quite complicated to implement. The accuracy and applicability of the neural network approach to traffic prediction is limited [13]. Finally, traffic prediction based on the α-stable model has the same problem as traffic prediction based on the FARIMA model, which relies on the accurate estimation of the Hurst parameter. Moreover, the α-stable model is based on a generalized central limit theorem and its application is limited by that. It might achieve a good performance in heavy traffic or when there is a high level of traffic aggregations. However when traffic conditions deviate from that, the performance may be poor. Furthermore, the α-stable model is a parsimonious model, which may not be able to capture the complex scaling behavior of the traffic.

Earlier work also exists on using wavelet for traffic prediction. In [17], Wang et al. use the Daubechies 40 wavelet filter to decompose the raw traffic rate into the wavelet coefficients and the scaling coefficients at one scale only and predict the wavelet coefficients and scaling coefficients using the recursive least square algorithm. The predicted wavelet and scaling coefficients are combined to give the predicted traffic rate. Their approach suffers from the fundamental flaw that in a Daubechies 40 wavelet, the wavelet coefficients and the scaling coefficients at the current time t rely on the future values of the raw signal. Therefore their algorithm actually cannot do any prediction. In [18], Papagiannaki et al. use the $\grave{a}-trous$ wavelet transform for long-term traffic prediction. By using a B_3 spline filter, they decompose the raw signal, i.e. an exponentially weighted moving average of the $10s$ link utilization measurements, into six scales. The wavelet coefficients and the scaling coefficients at each scale are predicted using the ARIMA model. The predicted wavelet coefficients at scale 3 and the predicted scaling coefficients at scale 6, i.e. the largest timescale, are used to construct the predicted traffic value. The B3 spline filter used in their approach suffers the same problem as the Daubechies 40 wavelet filter in that the wavelet coefficients and the scaling coefficient at the current time t rely on the future values of the raw signal. To solve the problem, Papagiannaki et al. use the weekly standard deviation of the wavelet coefficients at scale 3 to replace the wavelet coefficient at scale 3 and use the weekly average value of the scaling coefficients at scale 6 to replace the scaling coefficient at scale 6 for long-term traffic prediction. Their approach may distort the physical meaning of the wavelet transform and is custom made for the traffic being analyzed, which do not have general significance. Moreover, their focus is on the prediction of long term traffic trend, which allows them to discard traffic variations at small scales (i.e. wavelet coefficients at small scales). The same method is not applicable to real time prediction of the instantaneous traffic rate, where the small scale traffic component constitutes a substantial part of the instantaneous traffic rate.

Wavelet transform has been widely used in traffic analysis and modeling, in this paper we use a special kind of redundant wavelet transform, i.e. the $\grave{a}-trous$ Haar wavelet transform, which is particularly suited for traffic prediction, and the ARIMA model for traffic prediction. A major advantage of the $\grave{a}-trous$ Haar wavelet transform is the calculation of the scaling coefficients and wavelet coefficients at the current time t uses information before time t only. Hence it solves the problem in [17], [18].

3. Wavelet Traffic Decomposition

Wavelet has many advantages when used for traffic analysis. Fundamentally, this is due to the non-trivial fact that the analyzing wavelet family itself possesses a scale invariant feature, a property not shared by other analysis methods. Quite different kinds of scaling features can be analyzed by the same technique.

Wavelet analysis is based on the decomposition of a signal using orthogonal bases[†]. Discrete wavelet transform (DWT) consists of the collection of coefficients

$$c_J(k) =< X, \varphi_{Jk}(t) >,\ d_j(k) =< X, \psi_{jk}(t) >,\ j,k \in Z, \tag{2}$$

where $< *, * >$ denotes inner product, $\{d_j(k)\}$ are the wavelet coefficients and $\{c_J(k)\}$ are the scaling coefficients. Equation (2) compares the signal X to be analyzed with a set of analysis functions

$$\psi_{jk}(t) = 2^{-j/2}\psi(2^{-j}t - k), \tag{3}$$

which is constructed from the mother-wavelet $\psi(t)$ by

[†] Some other wavelet bases exist, such as semi-orthogonal or bi-orthogonal wavelet bases. However in this research, we only consider orthogonal bases.

a time-shift operation and a dilation operation. The mother wavelet is a band-pass or oscillating function, hence the name "wavelet". Function $\varphi_{Jk}(t)$ is a time shifted version of the mother scaling function $\varphi_J(t)$: $\varphi_{Jk}(t) = \varphi_J(t-k)$. $\varphi_J(t)$ is a low-pass function which can separate the large timescale (low frequency) component of the signal. Thus wavelet transform decomposes a signal into a large timescale approximation and a collection of details at different smaller timescales. Theoretically the scale j can span from $-\infty$ to ∞. For practical signals, i.e. network traffic, we limit the scale to $0 \sim J$, where scale J is the largest timescale and scale 0 is the smallest timescale.

Define a dilated and shifted function $\varphi_{jk}(t)$ of $\varphi(t)$ as

$$\varphi_{jk}(t) = 2^{-j/2}\varphi_(2^{-j}t-k). \tag{4}$$

Denote the subspace spanned by the basis functions $\{\varphi_{jk}, k \in Z\}$ as V_j and the subspace spanned by the basis functions $\{\psi_{jk}, k \in Z\}$ as W_j. Multiresolution analysis (MRA) requires the subspaces satisfy

$$V_J \subset V_{J-1} \subset \cdots \subset V_0 \quad \mathit{and} \quad V_{j+1} \bigoplus W_{j+1} = V_j. \tag{5}$$

Equation (5) implies that we can zoom into any timescale that we are interested in and use the coefficients of the wavelet transform to directly study the scale dependent properties of the data. For example, if we fix a scale j and investigate certain statistics of the wavelet coefficients at that scale across time we can obtain information about the scaling behavior of the signal as a function of j (the global-scaling behavior). Alternatively, if we fix a point in time t and examine how the wavelet coefficients within the cone of influence of t change across scales as we examine finer and finer scales, we can determine the local irregularity (the local scaling behavior) of the signal about the point t. Moreover the analysis of each scale is largely decoupled from that at other scales [6]. Refer to [19], [20] for details of the wavelet theory.

The roles of the mother scaling function $\varphi(t)$ and the mother wavelet function $\psi(t)$ can also be represented by a low-pass filter h and a high pass filter g. Thus the analysis and synthesis of a signal $x(t)$ can be implemented efficiently as a filter bank [19]. The approximation at scale j, $c_j(t)$ is passed through the low-pass filter h and the high pass filter g to produce the approximation $c_{j+1}(t)$ and the detail $d_{j+1}(t)$ at scale $j+1$. At each stage, the number of coefficients at scale $j+1$ is decimated into half of that at scale j, due to downsampling. This decimation reduces the number of data points to be processed at the larger timescales and removes the redundancy information in the wavelet coefficients and the scaling coefficients. Decimation allows us to represent a signal X by its wavelet and scaling coefficients whose total length is the same as the original signal. However decimation has the undesirable effect that we cannot relate information at a given time point at different scales in a simple manner. Moreover, while it is desirable in some applications (e.g. image compression) to remove the redundancy information, in time series prediction the redundancy information can be used to improve the accuracy of the prediction.

In this paper, we use a redundant wavelet transform, i.e. the $\grave{a}-trous$ wavelet transform [21], to decompose the signal. The $\grave{a}-trous$ wavelet transform is a non-decimated wavelet transform which produces smoother approximations of the signal. Using the $\grave{a}-trous$ wavelet transform, the scaling coefficients at different scales can be obtained as:

$$c_0(t) = x(t) \tag{6}$$

$$c_j(t) = \sum_{l=-\infty}^{\infty} h(l)c_{j-1}(t+2^{j-1}l). \tag{7}$$

where $1 \le j \le J$, and h is a low-pass filter with compact support. The wavelet coefficients at scale j are given by:

$$d_j(t) = c_{j-1}(t) - c_j(t). \tag{8}$$

The set $\{d_1, d_2, ..., d_J, c_J\}$ represents the $\grave{a}-trous$ wavelet transform of the signal up to the scale J, and the signal can be expressed as a sum of the wavelet coefficients and the scaling coefficients:

$$x(t) = c_J(t) + \sum_{j=1}^{J} d_j(t) \tag{9}$$

Many wavelet filters are available, such as Daubechies' family of wavelet filters, $B3$ spline filter, etc. Here we choose the Haar wavelet filter to implement the $\grave{a}-trous$ wavelet transform. A major reason for choosing the Haar wavelet filter is the calculation of the scaling coefficients and the wavelet coefficients at time t uses information before time t only. This is a very desirable feature in time series prediction. The Haar wavelet uses a simple filter $h = (1/2, 1/2)$ [19], [20]. The scaling coefficients at the higher scale can be easily obtained from the scaling coefficients at the lower scale:

$$c_{j+1,t} = \frac{1}{2}c_{j,t-2^j} + \frac{1}{2}c_{j,t}. \tag{10}$$

The wavelet coefficients can then be obtained from Equation (8).

4. The Prediction Algorithm

In this section, we use the aforementioned $\grave{a}-trous$ Haar wavelet decomposition for traffic prediction. Instead of predicting the original signal directly, we predict the wavelet coefficients and the scaling coefficients independently at each scale and use the wavelet coefficients and the scaling coefficients to construct the predicted value of the original signal.

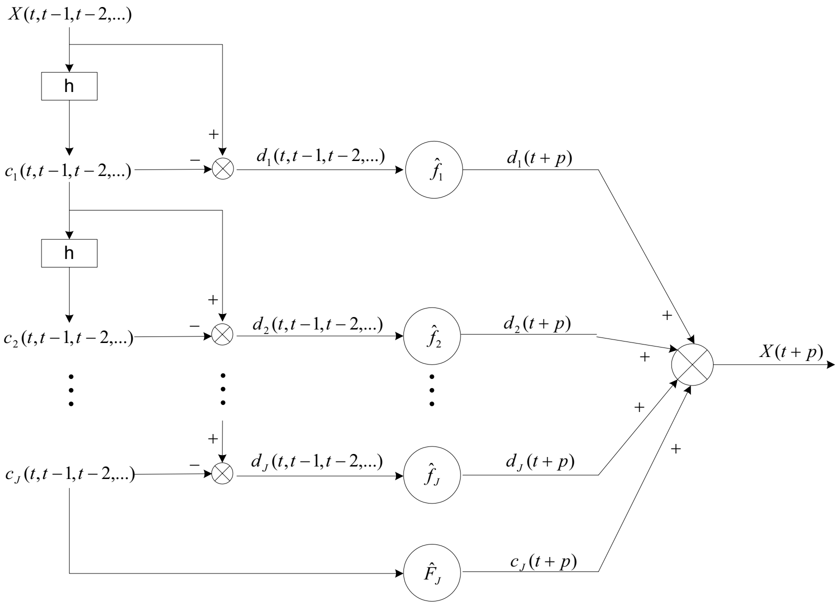

Fig. 1 Architecture of the prediction algorithm

Fig. 1 shows the architecture of the prediction algorithm. Coefficient prediction can be represented mathematically as

$$\widehat{c}_J(t+p) = \widehat{F}_J(c_J(t), c_J(t-1), ..., c_J(t-m)), \tag{11}$$

$$\widehat{d}_j(t+p) = \widehat{f}_j(d_j(t), d_j(t-1), ..., d_j(t-n_j)), \tag{12}$$

where m and n_j is the number of coefficients taken for prediction and p is the prediction depth. In this paper, we only consider one-step prediction, i.e. $p = 1$. Multistep prediction can be achieved by using the predicted value as the real value or by aggregating the traffic into larger time interval.

$ARIMA(p, d, q)$ model is used for prediction. An ARMA(p,q) (autoregressive moving average) model can be represented as:

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \tag{13}$$

where Z_t is a Gaussian distributed random variable with zero mean and variance σ^2 and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors [22]. Equation (13) can also be written in a more concise form as:

$$\phi(B)X_t = \theta(B)Z_t, \tag{14}$$

where ϕ and θ are polynomials of degree p and q respectively and B is the backward shift operator:

$$B^j X_t = X_{t-j}, j = 0, 1, ... \tag{15}$$

ARMA model assumes that the time series are stationary. For nonstationary time series, differencing operation can be used to remove the non-stationary trend in the time series. We define the lag-1 difference operator ∇ by

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t. \tag{16}$$

An ARIMA(p,d,q) model is an ARMA(p,q) model that has been differenced d times:

$$\phi(B)(1 - B)^d X_t = \theta(B)Z_t. \tag{17}$$

Fig. 2 and Fig. 3 shows the wavelet coefficients and the scaling coefficients of an hour-long LAN traffic trace. The time series being analyzed is the data rate of the LAN trace measured in byte/s during 1s measurement intervals. The details of the traffic trace will be introduced later. A visual inspection of the scaling coefficients and the wavelet coefficients indicates that the wavelet coefficients can be reasonably treated as a stationary time series with zero mean. Therefore wavelet coefficients can be modeled using the ARMA(p,q) model, or equivalently the ARIMA(p,0,q) model. However there is significant non-stationarity in the scaling coefficients. This non-stationarity becomes more obvious when examining the scaling coefficients over a longer time period as shown in Fig. 4. Therefore for the scaling coefficients it is more appropriate to use the ARIMA(p,d,q) model.

Box-Jenkins forecasting methodology is used to establish the ARIMA(p,d,q) model for prediction at each scale. Box-Jenkins methodology involves four steps [22]:

- The first step is the tentative identification of the model parameters. This is done by examining the sample autocorrelation function and the sample partial autocorrelation function [22] of the time series X.
- Estimation step. Once the model is established,

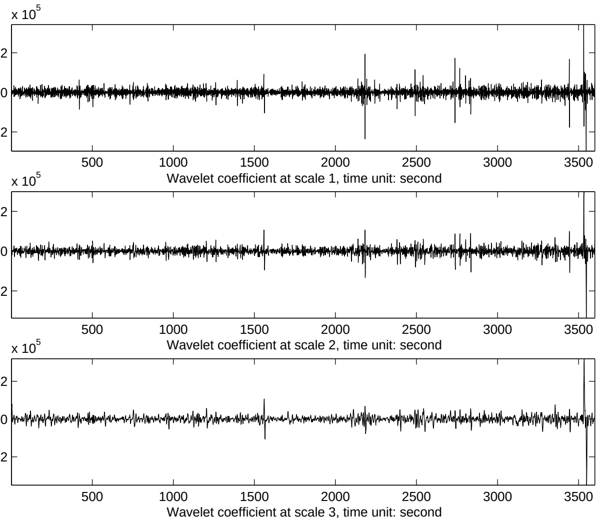

Fig. 2 Wavelet coefficients from scale 1 to 3

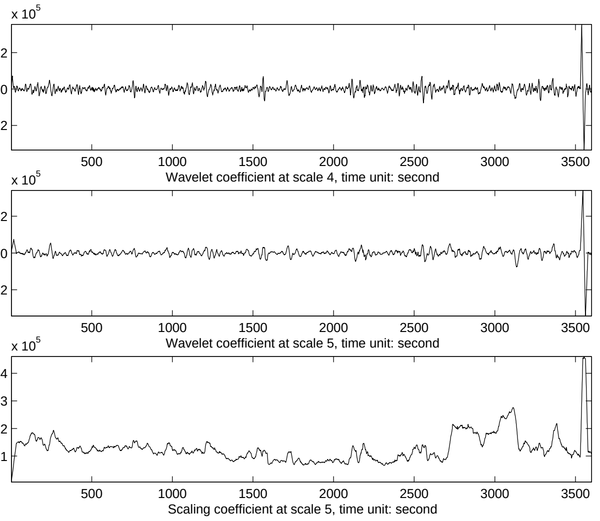

Fig. 3 Wavelet coefficients at scales 4 & 5 and Scaling coefficients at scale 5

the model parameters can be estimated using either a maximum likelihood approach or a least mean square approach. In this paper both the maximum likelihood approach and the least mean square approach were tried and their results are almost exactly the same. Thus we stick to the least mean square approach for its simplicity.

- Diagnostic check step. Diagnostic checks can be used to see whether or not the model that has been tentatively identified and estimated is adequate. This can be done by examining the sample autocorrelation function of the error signal, i.e. the difference between the predicted value and the real value. If the model is inadequate, it must be modified and improved.
- When a final model is determined, it can be used to forecast future time series.

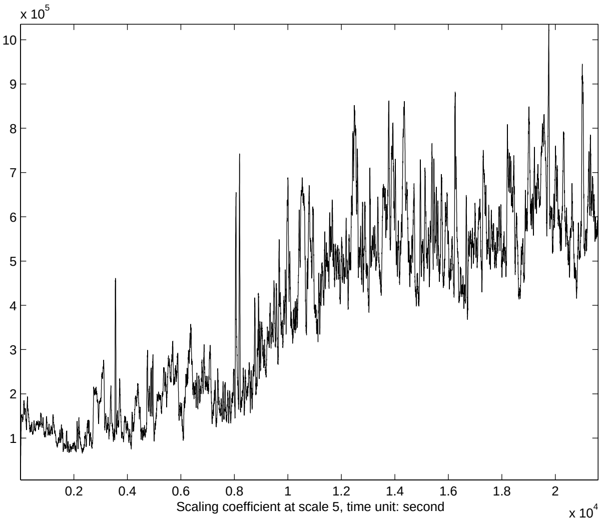

Fig. 4 Scaling coefficients at scale 5 over a 6-hour period

5. Simulation

In this section, we apply the proposed method to the prediction of real network traffic. The traffic traces used were collected by WAND research group at the University of Waikato Computer Science Department. It is the LAN traffic at the University of Auckland on campus level. The traffic traces were collected between 6am and 12pm from June 9, 2001 to June 13, 2001 on a 100Mbps Ethernet link. IP headers in the traffic trace are GPS synchronized and have an accuracy of $1\mu s$. More information on the traffic trace and the measurement infrastructure can be found on their webpage: http://atm.cs.waikato.ac.nz/wand/wits/auck/6/. Fig.5 shows the traffic rate of the traffic trace measured between 6am and 12am on June 12, 2001. The traffic rate is measured on 1 second intervals. Five traffic traces are available. Table 1 shows information of the traffic traces. The Hurst parameters of the traffic traces are also shown in the table for reference. The Hurst parameters are obtained using the method in [23].

We use the traffic rate measured in the previous 1s time intervals to predict the traffic rate in the next second. Prediction over a longer or a shorter time interval can be achieved by reducing the length of the time interval or by multistep prediction. To validate the performance of the proposed prediction model, one of the traffic traces (i.e. trace 4) was picked randomly to establish the prediction model and the prediction model is then applied to the other traffic traces for prediction.

Table 2 shows the model parameters of the ARIMA(p,d,q) model at each scale. Three scales are chosen. The choice on the number of scales is made based on the tradeoff between the model complexity

Table 1 Trace trace used in the simulation

Trace ID	File name	Measurement time	Duration	Hurst Parameter
1	20010609-060000-e0.gz	Saturday June 9, 2001	6am-12pm	0.935
2	20010610-060000-e0.gz	Sunday June 10, 2001	6am-12pm	0.838
3	20010611-060000-e0.gz	Monday June 11, 2001	6am-12pm	0.890
4	20010612-060000-e0.gz	Tuesday June 12, 2001	6am-12pm	0.937
5	20010613-060000-e0.gz	Wednesday June 13, 2001	6am-9am	0.945

Table 2 Model parameter of the prediction model

Scale	Model name	Model parameters ϕ	Model parameters θ	Noise σ^2
Wavelet coefficient 1	ARIMA(1,0,4)	$\phi_1 = 0.8842$	$\theta_1 = 1.311, \theta_2 = -0.2185,$ $\theta_3 = 0, \theta_4 = -0.1008$	2.147×10^9
Wavelet coefficient 2	ARIMA(4,0,4)	$\phi_1 = 1.443, \phi_2 = -0.4782,$ $\phi_3 = 0.04215, \phi_4 = -0.02682$	$\theta_1 = -0.04322, \theta_2 = 1.768$ $\theta_3 = 0.04953, \theta_4 = -0.7767$	5.847×10^8
Wavelet coefficient 3	ARIMA(4,0,8)	$\phi_1 = 1.384, \phi_2 = -0.435$ $\phi_3 = 0.02306, \phi_4 = -0.004911$	$\theta_1 = -0.1833, \theta_2 = -0.1531,$ $\theta_3 = -0.1824, \theta_4 = 1.751,$ $\theta_5 = 0.1789, \theta_6 = 0.1508,$ $\theta_7 = 0.1782, \theta_8 = -0.7583$	1.422×10^8
Scaling coefficient 3	ARIMA(2,1,8)	$\phi_1 = 0.508, \phi_2 = 0.02201$	$\theta_1 = -0.07853, \theta_2 = -0.08036$ $\theta_3 = -0.07985, \theta_4 = -0.08014,$ $\theta_5 = -0.07935, \theta_6 = -0.08083,$ $\theta_7 = -0.0796, \theta_8 = 0.9188$	1.348×10^8

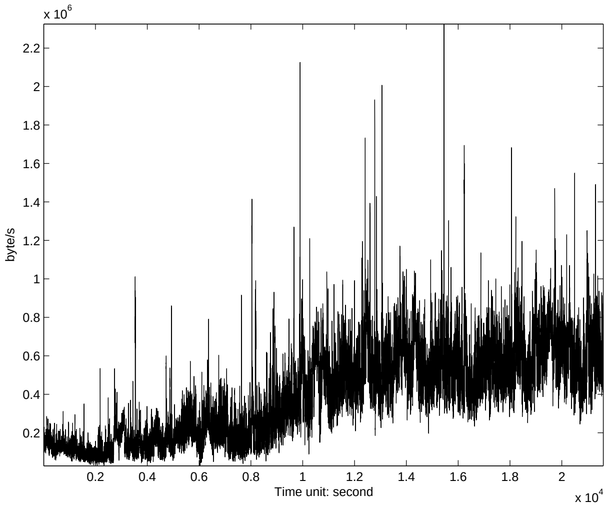

Fig. 5 Traffic rate of the LAN trace measured between 6am and 12am on June 12, 2001

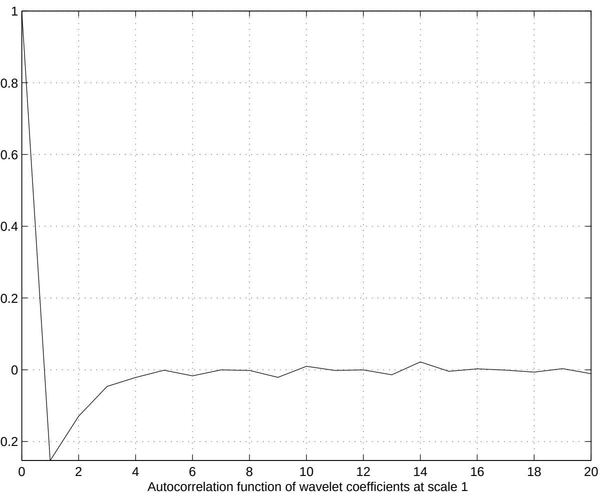

Fig. 6 Autocorrelation function of wavelet coefficients at scale 1

and accuracy. Further increase in the number of scales significantly increases the complexity of the algorithm but there is only a marginal increase in accuracy. As shown in the table, most noise in the model comes from the wavelet coefficients at scale 1. In comparison with the wavelet coefficients and the scaling coefficients at other scales, the wavelet coefficients at scale 1 has very weak autocorrelations and a white noise like power spectral density. It is almost like white noise. It is the wavelet coefficients at scale 1 that limits the overall performance that can be achieved by the prediction algorithm. Fig. 6 shows the autocorrelation function of the wavelet coefficients at scale 1.

The ARIMA models developed from trace 4 are then applied to the other traffic traces to establish the performance of the prediction algorithm. To measure the performance of the prediction algorithm, two metrics are used. One is the normalized mean square error (NMSE):

$$NMSE = \frac{\frac{1}{N}\sum_{n=1}^{N}(X(n) - \hat{X}(n))^2}{var(X(n))} \tag{18}$$

where $\hat{X}(n)$ is the predicted value of $X(n)$ and $var(X(n))$ denotes the variance of $X(n)$. The other is the mean absolute relative error (MARE), which is defined as:

$$MARE = \frac{1}{N}\sum_{n=1}^{N}\left|\frac{X(n) - \hat{X}(n)}{X(n)}\right| \tag{19}$$

Since the relative error may be unduely affected by vary small values of $X(n)$, to make meaningful observations, we only consider those samples of X which are not small than $E(X)$ when computing $MARE$. Table 3 shows the performance of the prediction algorithm. For comparison purpose, the performance of traffic prediction algorithms using the neural network approach and using ARIMA model without wavelet decomposition are also shown in the table. A number of neural network models with different number of input nodes, hidden nodes and transfer functions are evaluated, including those reported in [13], [24]. It is found that the 32-16-4-1 network architecture used in [24] gives the best performance. Hyperbolic tangent sigmoid transfer function is used in the hidden layer and linear transfer function is used in the output layer. The performance of the 32-16-4-1 neural network model is shown in Table 3 to represent the prediction performance using the neural network approach. To achieve a fair comparison, the same trace is used to train the neural network parameters. The very large data size in the training trace ensures the convergence of the neural network, which is also confirmed by a visual inspection of the error signal. The parameters of the ARIMA model without wavelet decomposition are $p = 1, d = 1, q = 4, \phi_1 = -0.146, \theta_1 = -0.274, \theta_2 = -0.270, \theta_3 = -0.113, \theta_4 = -0.055$.

As shown in Table 3, the proposed algorithm gives better performance than the neural network prediction in most cases except for trace 2, where the MARE metric of the neural network approach is slightly better than the proposed approach. However, the NMSE metric of the neural network approach is much worse than the proposed algorithm for trace 2. Since when training the neural network and estimating the ARIMA model parameters for the proposed prediction algorithm, the metric used is the mean square error, we conclude that the proposed algorithm performs better than the neural network prediction even for trace 2. The proposed algorithm outperforms the ARIMA model without wavelet decomposition both in NMSE and MARE.

Fig. 7 shows the absolute value of the autocorrelation function of the error signal for traffic trace 5 using the proposed algorithm and using the neural network prediction respectively. The autocorrelation function of the error signal using the proposed algorithm is much weaker than that using the neural network prediction and it dies down faster. This also indicates that the performance of the proposed algorithm is better than the neural network prediction. The autocorrelation function of the error signal for other traffic traces demonstrates similar characteristics.

As such, it can be concluded that the proposed algorithm achieves better performance than the neural network prediction. Moreover, only three scales are employed in the proposed prediction algorithm, which requires a memory length (here the memory length refers to the number of past raw data samples required for prediction) of about 16. In comparison, the neural network prediction requires a memory length of 32. The computation complexity using the proposed algorithm is also easier than that using neural network.

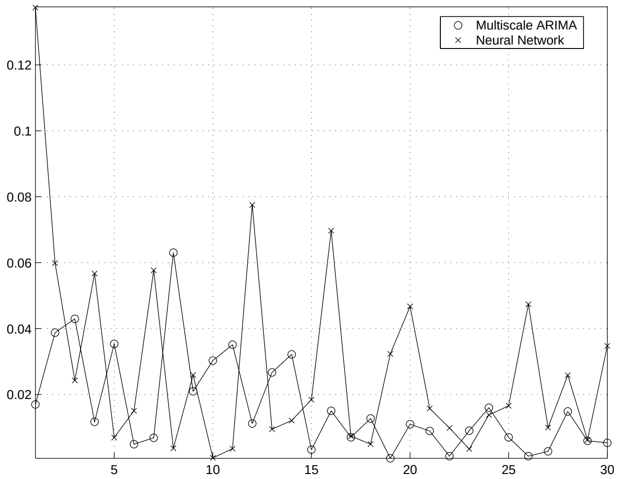

Fig. 7 Absolute value of the autocorrelation function of the error signal for trace 5 using multiscale ARIMA model and using neural network.

6. Conclusion and further work

In this paper we proposed a real-time network traffic prediction algorithm based on a timescale decomposition. The raw traffic data is first decomposed into different timescales using the *à trous* Haar wavelet transform. The prediction of the wavelet coefficients and the scaling coefficients are performed independently at each timescale using the ARIMA model. The predicted wavelet coefficients and scaling coefficient are then summed to give the predicted traffic value. As traffic variations at different timescales are caused by different network mechanisms, the proposed timescale decomposition approach to traffic prediction can better capture the correlation structure of traffic caused by different network mechanisms, which may not be obvious when examining the raw data directly.

The prediction algorithm was applied to the real network traffic. The performance of the prediction algorithm was compared with those using the neural network prediction and using ARIMA model without wavelet decomposition. It was shown that the proposed algorithm outperforms traffic prediction algorithm using the neural network approach and using ARIMA model without wavelet decomposition and gives more accurate prediction. The complexity of the prediction algorithm is also lower than that using the neural network. The autocorrelation of the error signal of the prediction algorithm is very weak, which is an indication of the good performance of the proposed algorithm.

Furthermore, the prediction model developed from a weekday traffic trace showed good performance when it was applied to traffic traces collected in both week-

Table 3 Performance of the prediction model

Trace ID	Multiscale ARIMA		Neural network		ARIMA	
	NMS	MARE	NMS	MARE	NMS	MARE
1	0.1319	0.1633	0.1603	0.1667	0.1860	0.2008
2	0.2296	0.2165	0.3168	0.2053	0.3485	0.2661
3	0.1507	0.1403	0.1565	0.1493	0.2190	0.1765
4	0.1592	0.1313	0.1622	0.1386	0.2299	0.1666
5	0.2197	0.1731	0.2258	0.1823	0.3167	0.2118

days and weekends where the traffic rate changes significantly. This demonstrated the good generalization capability of the proposed prediction algorithm. It is expected that some work will be done in the future to automate the parameter estimation process for the prediction model, which will enable the proposed algorithm to be easily used in a variety of environments.

References

[1] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," IEEE/ACM Transactions on Networking, vol.2, no.1, pp.1–15, 1994.

[2] V. Paxson and S. Floyd, "Wide area traffic: The failure of poisson modeling," IEEE/ACM Transactions on Networking, vol.3, no.3, pp.226–244, 1995.

[3] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," IEEE Transactions on Communications, vol.43, no.2/3/4, pp.1566–1579, 1995.

[4] M.E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," IEEE/ACM Transactions on Networking, vol.5, no.6, pp.835–846, 1997.

[5] A. Feldmann, A. Gilbert, W. Willinger, and T. Kurtz, "The changing nature of network traffic: Scaling phenomena," Computer Communication Review, vol.28, no.2, pp.5–29, 1998.

[6] P. Abry, P. Flandrin, M.S. Taqqu, and D. Veitch, "Wavelets for the analysis, estimation, and synthesis of scaling data," in Self-Similar Network Traffic and Performance Evaluation, ed. K. Park and W. Willinger, pp.39–88, John Wiley and Sons, Inc., 2000.

[7] J. Cao, W.S. Cleveland, D. Lin, and D.X. Sun, "The effect of statistical multiplexing on the long-range dependence of internet packet traffic," bell labs technical report, Bell Labs, 2002.

[8] H. Furuya, H. Nakamura, S. Nomoto, and T. Takine, "Local poisson property of aggregated ip traffic," IEICE Transactions on Communications, vol.E86- B, no.8, pp.2368–2376, 2003.

[9] E.C. Ifeachor and B.W. Jervis, Digital signal processing : a practical approach, Harlow : Prentice Hall, 2002.

[10] P.J. Brockwell and R.A. Davis, Introduction to Time Series and Forecasting, Springer-Verlag New York, Inc., 2002.

[11] Y. Shu, Z. Jin, L. Zhang, and L. Wang, "Traffic prediction using farima models," IEEE International Conference on Communications, pp.891–895, 1999.

[12] Y. Liang, "Real-time vbr video traffic prediction for dynamic bandwidth allocation," IEEE Transactions on Systems, Man and Cybernetics, Part C, vol.34, no.1, pp.32–47, 2004.

[13] J. Hall and P. Mars, "Limitations of artificial neural networks for traffic prediction in broadband networks," IEE Proceedings Communications, vol.147, no.2, pp.114–118, 2000.

[14] M. Lopez-Guerrero, J. Gallardo, D. Makrakis, and L. Orozco-Barbosa, "Optimizing linear prediction of network traffic using modeling based on fractional stable noise," 2001 International Conferences on Info-tech and Info-net, pp.587–592 vol.2, 2001.

[15] A. Karasaridis and D. Hatzinakos, "Network heavy traffic modeling using *alpha*-stable self-similar processes," IEEE Transactions on Communications, vol.49, no.7, pp.1203–1214, 2001.

[16] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-range dependence ten years of internet traffic modeling," IEEE Internet Computing, vol.8, no.5, pp.57–64, 2004.

[17] X. Wang, Y. Ren, and X. Shan, "Wdrls: a wavelet-based on-line predictor for network traffic," IEEE GLOBECOM '03., pp.4034–4038, 2003.

[18] K. Papagiannaki, N. Taft, Z.L. Zhang, and C. Diot, "Long-term forecasting of internet backbone traffic: observations and initial models," IEEE INFOCOM, pp.1178–1188, 2003.

[19] G. Strang and T. Nguyen, Wavelets and Filter Banks, Wellesley-Cambridge Press, 1996.

[20] I. Daubechies, Ten Lectures on Wavelets, Capital City Press, Montpelier, Vermont, 1992.

[21] M. Shensa, "The discrete wavelet transform: wedding the a trous and mallat algorithms," IEEE Transactions on Signal Processing, vol.40, no.10, pp.2464–2482, 1992.

[22] B.L. Bowerman and R.T. O'Connell, Time Series Forecasting - Unified Concepts and Computer Implementation, 2 ed., PWS publishers, 1987.

[23] D. Veitch and P. Abry, "A wavelet-based joint estimator of the parameters of long-range dependence," IEEE Transactions on Information Theory, vol.45, no.0018-9448, pp.878–897, 1999.

[24] Y. Liang and E. Page, "Multiresolution learning paradigm and signal prediction," IEEE Transactions on Signal Processing, vol.45, no.11, pp.2858–2864, 1997.

Guoqiang Mao received the Bachelor degree from Hubei Institute of Technology, China, in 1995, the Master degree from Southeast University, China, in 1998, and the Ph.D. degree from Edith Cowan University, Australia, in 2001. After graduation, he became a senior engineer at Intelligent Pixel Incorporation, Australia. He is now a lecturer at the University of Sydney. His research interests include network QoS, traffic measurement, analysis and modeling, performance analysis in both wired and wireless network and sensor network.