**Network Traffic Prediction Algorithm based on Wavelet Transform**

1 Xiaoling Tan, 2 Weijian Fang, 3 Yong Qu

*1, Chongqing Three Gorges University, Chongqing, China, xltan023@hotmail.com
2, Chongqing Three Gorges University, Chongqing, China, wanzhoufwj@yahoo.cn
3, Chongqing Three Gorges University, Chongqing, China, quyong98@163.com

**Abstract**

The features of dynamic, noise and instability, make the network traffic eruptive and unstable, and this obstructs the network traffic prediction. In order to figure out its characteristics and developing tendency accurately, the paper proposes a wavelet-transform-based prediction algorithm: Firstly, with the multi-resolution analysis of wavelet transform, the network traffic, which is difficult to be analyzed or modeled in the time domain, is divided into different bands of frequency by wavelet decomposition. Later, simplify and stabilize the divided traffic by denoising with different thresholds on the detail components of different sub-bands of frequency. At last, synthesizes the traffic prediction with the predictions of the denoised sub-traffic by Auto Regressive Moving Average Model. The algorithm improves the prediction accuracy significantly in practical modeling, especially in short-range traffic prediction.

**Keywords**: Network Traffic, Prediction, Wavelet Decomposition

1. Introduction

With the rapid developing of network technology and perfecting of Internet service [1, 2], the internet scale is bigger and bigger [3, 4]. Meanwhile, internet service and users’ requirement is more and more diversified and complicated [5]. The huge, isomeric, complicated and dynamic internet network owes a complicated operational mechanism, which makes network controlling, evaluating and analyzing difficult [6, 7]. In order to better understand the inner-internet dynamic behavior and the relevant factors, set up an effective model for network analyzing and prediction, and achieve for a better network protocol design, planning, optimization and service, a lot of researches have been taking place on IP network behaviors and performance [8, 9, 10]. Internetwork’s dynamic behavior could be fully reflected by its network traffic, while the traffic is one of the dominant factors of its dynamic behavior [11, 12]. As a result, network traffic is an entry point for network architecture analyzing, designing, improving, and understanding of its dynamic behavior by modeling and analyzing [2, 13].

Traffic prediction technology makes the advanced warning possible. Traffic prediction technology means the prediction of future traffic data by historical data analyzing and suitable traffic modeling [14, 15]. The traffic tendency could be revealed by the prediction [16, 17]. If the prediction data overpassed the threshold, warning would be activated and the network would be checked in advance to avoid the network failure [18, 19, 20].

In order to figure out its characteristics and developing tendency accurately, the paper proposes a wavelet-transform-based prediction algorithm: Firstly, with the multi-resolution analysis of wavelet transform, the network traffic, which is difficult to be analyzed or modeled in the time domain, is divided into different bands of frequency by wavelet decomposition. Later, simplify and stabilize the divided traffic by demising with different thresholds on the detail components of different sub-bands of frequency. At last, synthesizes the traffic prediction with the predictions of the denoised sub-traffic by Auto Regressive Moving Average Model. The algorithm improves the prediction accuracy significantly in practical modeling, especially in short-range traffic prediction.

2. Traffic modeling

Traffic model is used to predict network performance and evaluate access controlling. An accurate traffic model could reflect the statistical features of the real traffic. It should contain manageable parameters, and the parameter estimation should be as simple as possible [4].
The modeling consists of measuring-based traffic modeling and physical modeling. In this modeling, data is extracted and statistical analyzed from the real traffic without taking the network into consideration. The purpose of using statistics is to test and verify that if the data is corresponded with the hypothetical model. If no hypothesis was set before modeling, a most suitable model should be chosen according to DS. The researchers found that the eruption of traffic is relatively invariant to time scale. This self-similar phenomenon exists in many networks. Many self-similar traffic models, from fractional Brownian motion to chaotic model, appear in LAN, WAN, ATM and FON [5]. Although some features of traffic data are reflected in those models, the multi-fractal in small time scale is not expressed. The application of wavelet technology could make the modeling more accurate. Wavelet technology is sensitive to scale phenomena, by which the model of a network traffic could be worked out. What’s more, the wavelet technology could localize the specific signal in different time and scale, by which the multi-fractal of the traffic could be detected, identified and described.

This modeling is a structured modeling to network features. It is a model, usually a network-based one, combined with network configuration parameters (topological structure, congestion control, QoS, etc.) and traffic data. Closely connected with network’s physical features, it makes network improvement by traffic controlling possible. Meanwhile, it creates a new orientation for network plan and design. The physical modeling for self-similar traffic is set to work out the source of the self-similar traffic. Some said the source is relevant to the arrive mode of single DS, such as VBR (Variable bit rate) video [6]; and some regarded the heavy-tailed distribution of the transmitted size and the chaos of TCP congestion controlling as the source [7].

3. Proposed algorithm

Wavelet is a small wave. Wavelet analysis is regarded as a milestone of the development of Fourier Analysis. It is called Mathematic Microscope, for it could be used to decompose any functions by zooming. It is an analytical approach of time and frequency localization, in which the time and frequency windows are with a fixed acreage and flexible size. Being different from Fourier Transform, wavelet transform can be used to describe the frequency features of the signals in local area [8].

3.1. From Fourier transform to wavelet analysis

Network traffic is mutable and unstable signal. In order to know it, the studies should not only on its frequency, but also on its frequency in different moments. In other words, it should be focused on its overall and local features in time and frequency domains, which obviously Fourier Transform couldn’t approach to it, for it is can completely localized in frequency domain (decomposing signal to detail frequency) but do nothing in time domain. Actually signal is consisted with various frequency components. When the signal alternated sharply, a short time window needed for more frequency information; when the signal stays gentle, a long time window needed for the overall description. In short, it needs a flexible time window. But the size and shape of STFT (short-time Fourier Transform) window function is fixed. It couldn't meet the needs. Therefore, a technology differing from Fourier Transform and STFT is required. Wavelet analysis is suitable to the analysis of unstable signals. It is with the capability of presenting the localized features of signal in time and frequency domains, which is helpful for the detecting of signals’ transient and singularity. Wavelet transform de-correlates the LRD signal, by which the analyzing and modeling of Internet traffic in wavelet domain become easier than that in time domain. When the traffic is decomposed to different frequency band, it was simplified and smoothed. Hence, the wavelet analysis is adopted for traffic analysis [9].

3.2. Application of wavelet analysis theory in signal processing

Wavelet is the functions which could decompose and analyze signal in different scale of time and frequency. Wavelet transform and Fourier Transform are different by the function comparison. Fourier Transform function is complex exponential. It is general for the function and is nonzero in the whole time scale; while the wavelet transform is local. It only exists in a specific duration and the signal in a
specific moment could be observed then. What’s more, the scale of wavelet could be extended to the whole time domain for the observation of the total signal [10].

Square integrable function L2 (R),

Wavelet function is: set \( \psi (t) \) as a square integral function, which is \( \psi (t) \in L2 (R) \), if its Fourier Transform \( \Psi(\omega) \) satisfies:

\[
\int_{-\infty}^{\infty} \left| \Psi(\omega) \right|^2 d\omega < \infty
\]  

(1)

\( \psi (t) \) is called as a basic wavelet or Mother Wavelet, and the formula above is its admissible condition.

Definition of continuous wavelet basis functions:

With the dilation and shifting operations of basic wavelet function, set its elastic factor (scale factor) as a, shift factor as \( \tau \), the function after dilation and shift \( \psi_{a,\tau}(t) \) as, there is the following:

\[
\psi_{a,\tau}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t - \tau}{a} \right), a \in \mathbb{R}^+, \tau \in \mathbb{R}
\]  

(2)

\( \psi_{a,\tau}(t) \) is called as basic wavelet function depending on a and \( \tau \). Because a and \( \tau \) are continuous, \( \psi_{a,\tau}(t) \) is called as continuous basic wavelet function. Their origin is the same Mother Function of \( \psi(t) \). Basic wavelet function is finitude or finitude approaching in time and frequency domain. The function after the dilation and shifting operations is also localized in time and frequency domain.

Unfolding a \( f(t) \) in L2 (R) under wavelet basis, a CWT (Continuous Wavelet Transform) of \( f(t) \) is formed. The expression is:

\[
W_{\psi}(f)(a,\tau) = \left\langle f(t), \psi_{a,\tau}(t) \right\rangle
\]  

(3)

Every transform should be with a reverse transformation. If the wavelet satisfied with the admissible condition, continuous wavelet transform owes its reversion, which means the original signal could be recovered. The inverse transform formula is as the following [11]:

\[
X(t) = \frac{1}{C_\psi} \int_{0}^{\infty} \int_{-\infty}^{\infty} W_{\psi}(f)(a,\tau) \psi_{a,\tau}(t) d\tau
\]  

(4)

where \( C_\psi = \int_{0}^{\infty} \frac{\left| \Psi(a\omega) \right|^2}{a} da < \infty \) is the admissible condition of \( \psi (t) \).

The definition of continuous wavelet transform showed that a non-orthogonal over entirely basis is formed by scale and continuous wavelet transform function. It means CWT coefficient is with a great redundancy. In order to reduce the redundancy and get a group of orthogonal basis, the method of MRA (multi-resolution analysis) is proposed.

3.3. Mallat algorithm of wavelet transform

Wavelet decomposition could be realized by MRA-based Mallat algorithm. Amazingly, Mallat algorithm is the quick algorithm of dyadic discrete wavelet transformation. Its decomposition algorithm is as the following:

\[
c_{j+1} = Hc_j \quad d_{j+1} = Gc_j \quad (j = 0, 1, \ldots, J)
\]  

(5)

\( H \) and \( G \) is the orthogonal wavelet basis of Mallat. Its decomposition is the convolution and point sampling of \( H \) and \( G \). \( H \) is the low-pass filter, while \( G \) is the high-pass one. Figure 1 shows the process of decomposition. Set the sampling frequency of original signals as \( f_s \), the approximate part \( c_j \) of \( 0 ~ f_s/2 \) and detail part \( d_j \) of \( f_s/2 ~ f_s/2 \) would be obtained in the scales of \( 2^j \). Define \( c_0 \) as the original signal \( X \), and the \( X \) could be decomposed to \( c_J \) and \( d_1, d_2, \ldots, d_J \) (\( J \) is the maximum decomposition layer-number ), \( c_j \) is the scale coefficient, and \( d_j \) the wavelet coefficient. The approximate part and detail part are different parts in different bands of \( X \). From the view of band allocation, the process is the semi-dividing of band. The decomposed signal by Mallat algorithm could be reconfigured by reconfiguration algorithm:
\[ c_j = H^* c_{j+1} + G^* d_{j+1}, \quad j = J - 1, J - 2, \ldots, 0 \]  \hspace{1cm} (6)

The \( H^* \) and \( G^* \) are dual operators of \( H \) and \( G \). Signal points could be added by the reconfiguration. Figure 2 shows the process of reconfiguration. It reconfigures \( c_j \) and \( d_1, d_2, \ldots, d_J \), and obtains \( C_j \) and \( D_1, D_2, \ldots, D_J \), which is with the same points of the \( X \), and satisfies the following:

\[ X = D_1 + D_2 + \cdots + D_J + C_J \]  \hspace{1cm} (7)

As Figure 3 shows, the paper decomposed the traffic data from a university network center by db3 (Daubechies). The data is collected from 2:00am to 9:00am of the next day with an interval of 5min. Decomposing in \( N \) scales by Mallat algorithm, a group of approximate components \( C_j \) and \( N \) groups of detail components \( D_j \) (\( j = 1, 2, \ldots, N \)) are obtained. Its \( N \) is with the value when the variance \( C_j \) is minimum. The traffic is separated into 9 layers: approximate component of the third layer of Figure 4; detail component of the first layer of Figure 5; detail component of the second layer of Figure 6; and detail component of the third layer of Figure 7. It shows that when \( N \) is between 3 and 5, approximate signal got minor variance \( C_j \); and when \( N > 5 \), \( C_j \) rises, which might be caused by the feature of wavelet decomposition: the more separated, the more bands are, the more smooth and stable the components are, but greater the error would be. Hence the layer could not be too few or too much, and the paper chose \( N = 3 \).

The traffic contains necessary LF signals and its eruptive parts, which usually reflects some significant features and contains abundant HF signals. With different thresholds, the HF of noise in the decomposed detail components should be cleared.

![Figure 3. Network Traffic](image)

![Figure 4. The Approximate Component of the Third Layer](image)
ARMA (p, q) model of stable time series \( \{x_t \mid t=1, 2, \ldots, N\} \) is:

\[
x_t - \sum_{j=1}^{p} \phi_j x_{t-j} = \alpha_t - \sum_{j=1}^{q} \theta_j \alpha_{t-j}
\]  \hspace{1cm} (8)

The \( p \) and \( q \) are the orders of auto-regressive and the moving average part. The \( \phi_j \) and \( \theta_j \) are the coefficients of auto-regressive and the moving average part. By the algebraization, ARMA (p, q) model could be expressed by [13][14]:

\[
\phi(B)x_t = \theta(B)\alpha_t
\]  \hspace{1cm} (9)

Transforming \( \{x_t\} \) to \( \{\alpha_t\} \), which is:

\[
x_t = \frac{\theta(B)}{\phi(B)} \alpha_t = \sum_{j=0}^{\infty} G_j \alpha_{t-j} = G(B)\alpha_t, \quad G_0 = 1
\]  \hspace{1cm} (10)

The \( G_j \) is Green function, acting as power. \( x_t \) is described as a linear combination of the weighed past and present value of \( \alpha_t \), by which to explain how \( \alpha_t \) and its past value impact on \( x_t \).
By a stable non-parameter method, the denoised approximate components $C_j$ and detail components $D_j (j=1, \ldots, J)$ are tested. If the result was between the top floor of the given significant level of $\alpha = 0.05$, the series is not obviously under a underlying trend, and the hypothesis of the stationarity is acceptable. According to the time series theory, ARMA $(2n, 2n-1)$ could be chosen, whose $n$ is positive integer. And according with AIC formula, the orders are chosen as $p=2$ and $q=1$ [15].

By the ARMA $(2, 1)$ model and least-squares estimation, the following parameters are obtained:

For the approximate component at the third layer:

$$C_3 = 1.8794C_{r-3} - 0.92086D_{r-2} + a_{C_j} + 0.17761a_{D_{r-1}}$$

For the detail component at the first layer:

$$D_1 = 0.3737D_{l-1} - 0.65214D_{l-2} + a_{D_{l-1}} - 0.3219a_{D_{l-1}}$$

For the detail component at the second layer:

$$D_2 = 0.68391D_{r-2} - 0.68682D_{r-2} + a_{D_{r-2}} + 0.076438a_{D_{r-1}}$$

For the detail component at the third layer:

$$D_3 = 1.5538D_{r-3} - 0.87955D_{r-3} + a_{D_{r-3}} - 0.22513a_{D_{r-3}}$$

And then do the smooth linear minimum variance prediction by model recursion, expressing the Green function as $G_0, G_1, G_2, \ldots, G_i$, the followings obtained:

$$G_0 = 1$$

$$G_i = \phi_i - \theta_i$$

$$G_i = \phi_iG_{i-1} + \phi_{i-1}G_{i-2} \quad (i \geq 2) \quad (11)$$

The following vector recursive prediction formulas could be obtained by the thereom [18]:

$$\hat{x}_{t+1}(1) = \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-q+1} \end{pmatrix} = \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_q \end{pmatrix} + \begin{pmatrix} G_1 \\ G_2 \\ \vdots \\ G_q \end{pmatrix} \begin{pmatrix} x_{t+1} \\ x_{t+2} \\ \vdots \\ x_{t+q} \end{pmatrix} (12)$$

its $\phi_i = \begin{cases} \phi_i, & i = 1, 2, \ldots, p \\ 0, & i > p \end{cases}$ When $p \leq q$, the third term of the Formula (12) is zero.

The following recursion formulas could be obtained from ARMA $(2, 1)$:

$$\hat{x}_{t+1}(1) = (-G_1)x_t(1) + G_1x_{t+1} + \phi_2x_t$$

$$\hat{x}_{t+1}(k) = \phi_2\hat{x}_{t+1}(k-1) + \phi_2\hat{x}_{t+1}(k-2)$$

The mean-square error of step-$k$ prediction is only relevant to the steps, not with the time. It’s corresponding to the real that farther the time, greater the variance, and it is becoming unpredictable [16].

The result could be obtained by combining the predicted values of the components as formula (15) shows:

$$\hat{x}_{t+1} = \hat{x}_{t+1} + \sum_{j=t+1}^{q} \hat{d}_{j+1}$$

5. Result of prediction

Figure 8 shows the prediction of the traffic from 9: 00am to 11: 00am of the next day by ARMA $(2, 1)$ model. The prediction reflects the real traffic well, which proves the model is feasible. In nature, wavelet decomposition and reconstruction separates a set of signals of summarized information into $N + 1$ groups time series signals with different features.

The inner variation tendency of is reflected by approximation signal $C_j$; and the affection of stochastic disturbance is reflected by $d_j (j= 1, 2, \ldots, N)$. They are different. Different signals could be predicted by different parameters. The combination of the predictions is better than one prediction of
the whole. By the Mathematic Microscope advantage of wavelet transform, it has improved the predictive accuracy significantly. What’s more, ARMA model is not the only way. Different models are accessible to different components, which could live the traditional linear prediction model and other arithmetic’s.

In the wavelet prediction, the number of layers should be set properly. The more wavelet decomposition separated, the more are the bands, the more smooth and stable are the components, but greater the error would be. Hence the layer could not be too few or too much.

![Wavelet Decomposition Example](image)

### Figure 8. The Real Traffic and the Traffic Forecast

#### 6. Conclusion

In nature, wavelet decomposition and reconstruction separates a set of signals of summarized information into $N + 1$ groups of components with different features. The traffic tendency is reflected by approximate component; and the affection of stochastic disturbance is reflected by detail component. They are different. The eruption and noise could be separated effectively by de-noising HF part. By the features of wavelet transform, the model did the prediction and synthesizing separately to the components and improved the predictive accuracy significantly. It’s proved by experiment that it could make accurate prediction to network traffic, especially to the short-range traffic.

#### 7. Acknowledgment

The authors wish to thank the subsidizing from Chongqing Educational Program (2011-KS-016) and Chongqing Three Gorges University Science Research Program (11ZD-15).

#### 8. References


