

基于小波的多尺度网络流量预测模型

洪 飞¹⁾ 吴志美²⁾

¹⁾(北京航空航天大学计算机学院 北京 100083)

²⁾(中国科学院软件研究所多媒体通信与网络研究中心 北京 100080)

摘 要 通过把 ARIMA 线性预测方法引入小波域内, 提出一个基于多重分形小波模型的网络流量预测模型, 通过对真实网络流量的仿真实验, 结果表明该模型能够对网络流量进行比较精确的预测。

关键词 网络流量; 小波; ARIMA 模型; MWM; 预测

中图法分类号 TP393

Multiscale Network Traffic Prediction Model Based on Wavelet

HONG Fei¹⁾ WU Zhi-Mei²⁾

¹⁾(School of Computer Science, Beihang University, Beijing 100083)

²⁾(Multimedia Communication and Network Research Center, Institute of Software, China Academy of Sciences, Beijing 100080)

Abstract Scaling and multiscale behaviours, such as the long-range dependence, the self-similarity, and the multi-fractal have been commonly viewed as the most significant characteristics of the network traffic today. The capacity planning theory for network requires accurate modelling of the incoming traffic, as well as accurate predictions of its future behaviour. The wavelet-based approach is a natural way to provide multiscale prediction to applications. With introduction of the ARIMA linear prediction in the wavelet domain, a novel network traffic prediction model is presented based on the multi-fractal wavelet model, which is shown to be the accurate model of the multiscale network traffic. The simulation results with the real traffic traces show the accuracy and efficiency of the model.

Keywords network traffic; wavelet; ARIMA model; MWM; prediction

1 引 言

当前, 网络流量被公认最为重要的统计特征是尺度和多尺度行为特性, 如长相关性、自相似性以及多重分形性^[1~6], 这种特性不仅仅存在于 Internet 网络中, 同时存在于无线网络、Ad Hoc 网络以及卫星网络中. 网络流量这种复杂的行为特性通常表现在大多数时间尺度和统计阶上的突发性. 这使得网

络性能要比传统的符合高斯和短相关特性下的网络性能差许多. 从而, 尺度和多尺度网络流量的建模和预测对于网络容量规划、异构网络的性能保证有很重要的意义.

文献[7]最早提出了对 IP 网络流量的预测, 基于线性预测模型, 给出了基于 NSFNET 聚集流量的预测值. 得到的结论是基于网络流量的时间序列可以通过低阶的 ARIMA 模型精确建模, 提供精确度很高的预测值. Sang 和 Li 在文献[8]中分析了基

于 ARMA 和 MMPP 模型的多尺度网络流量预测的相关问题。

对于网络流量所表现出的多尺度本质,有些学者研究利用小波的尺度分析本质特性对网络流量进行预测.这里所描述的小波方法主要是根据较为简单的小波基,如 Haar 小波,本质上是一个二项尺度分析过程.就小波分析在网络流量尺度分析中的应用,Abry 等给出了一个较为详细的讨论^[9],同时给出了一种基于小波的 HURST 指数估计方法^[10].Feldmann 等进一步将小波分析推广到多尺度研究,也就是网络流量的多重分形特性^[11],同时研究了基于这些特性的网络控制机制和 TCP 拥塞控制机制^[12].Riedi 等提出了如何用小波拟合网络流量^[13],所提出的多重分形小波模型能够很好地在统计上拟合真实网络流量.本文工作主要在于如何通过基于小波的网络模型对网络流量进行预测.

第 2 节引入基于小波的多尺度分析;第 3 节给出基于小波的预测模型;第 4 节给出基于真实网络流量的仿真实验结果;最后给出相应结论和下一步的研究内容.

2 网络流量的多尺度分析

小波分析所固有的尺度特性很自然地成为进行多尺度预测的有效方法.二项尺度分析方法给出了一种研究网络流量多尺度特性的直观表示.在恰当小波基的选择下,网络流量在不同尺度下的特性可以得到很好的刻画,因此,该方法为研究这种尺度行为提供了一个有力的工具.

Mallat^[14],Daubechies^[15]和 Abry^[10]等利用小波方法做了一些研究工作.离散小波变换(DWT)是一维离散信号 $X(k)$ 的一种数学变换.直观上,离散小波变换把一个一维的信号分解为两个一维信号,分别是时间上平移后的逼近信号和尺度(频率)上缩放后的细节信号.基于选定的尺度函数 ϕ_0 和小波函数 ψ_0 ,构建带通小波函数基 $\psi_{j,k}$ 和低通尺度函数基 $\phi_{j,k}$,这样原始信号可以通过小波基的平移和收缩来表示,也就是逼近信号可以通过一系列尺度系数来表示,细节信号可以通过一系列小波系数来表示.小波函数基和小波函数以及尺度函数基和尺度函数的关系如下:

$$\begin{cases} \phi_{j,k}(t) = 2^{-j/2} \phi_0(2^{-j}t - k) \\ \psi_{j,k}(t) = 2^{-j/2} \psi_0(2^{-j}t - k) \end{cases}, k \in Z \quad (1)$$

为了得到原始信号的精确多尺度分析,所选择的函数 ϕ_0 和 ψ_0 都必须具有高阶矩并且能够构成一个无条件 Riesz 基.关于小波函数和尺度函数的详细说明可以参考 Daubechies 文章^[15].多尺度分析是由 Mallat^[14]最先提出,原始信号通过离散小波变换可以分解为逼近信号和细节信号,通过分析逼近信号 $approx_j$,可以得到随着尺度 j 增加,逼近信号不仅含有的采样值比较少,而且是对原始信号一个较为粗略的逼近.每一个相近的尺度逼近信号都含有上一个逼近尺度一半的采样点,同时保持了其一般的频谱信息.即使随着尺度 j 增加,逼近信号的采样点越来越少,逼近信号的时间周期是不变的.通过分析细节信号,可以知道从逼近信号中提取出的信息就是下一个尺度的细节信号,也就是 $approx_j = approx_{j-1} - detail_j$.可以这样认为,存在一系列自包含的子空间组成 $\{V_j\}_{j \in Z}$,也就是满足: $V_j \subset V_{j-1}$,多尺度分析是把信号 $X(k)$ 投影到每一个逼近的子空间 V_j .从而逼近信号可以通过下面的关系得到:

$$approx_j(t) = \{Proj_{V_j} X(k)\}(t) = \sum_k a_x(j, k) \phi_{j,k}(t) \quad (2)$$

系数 $a_x(j, k)$ 通过信号 $X(k)$ 和 $\phi_{j,k}$ 的内积定义:

$$a_x(j, k) = \langle X(k), \phi_{j,k} \rangle \quad (3)$$

同样的,细节信号可以通过下面的关系得到

$$detail_j(t) = (Proj_{W_j} X(k))(t) = \sum_k d_x(j, k) \psi_{j,k}(t) \quad (4)$$

系数 $d_x(j, k)$ 通过信号 $X(k)$ 和 $\psi_{j,k}$ 的内积定义

$$d_x(j, k) = \langle X(k), \psi_{j,k} \rangle \quad (5)$$

基于以上的多尺度分析,一个信号可以通过逼近信号以及相应的细节信号无损地表示出来,即

$$X(k) = approx_j(t) + \sum_{j=0}^J detail_j(t) \quad (6)$$

3 网络流量预测模型

多重分形小波模型是基于 Haar 小波的网络流量模型,这里 Haar 小波函数和尺度函数构成了一个简单的小波正交基.在小波变换中,所采用的尺度函数 $\phi(t)$ 、小波函数 $\psi(t)$ 如下所示:

$$\phi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ 0, & \text{否则} \end{cases}, \quad \psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{否则} \end{cases} \quad (7)$$

从而 Haar 小波变换的尺度系数和小波系数可以通过以下迭代公式计算得到

$$\begin{cases} u_{j-1,k} = 2^{-1/2}(u_{j,2k} + u_{j,2k+1}) \\ \tau_{j-1,k} = 2^{-1/2}(u_{j,2k} - u_{j,2k+1}) \end{cases} \quad (8)$$

相应的小波逆变换,也就是公式(6),可以通过尺度系数从逼近系数和小波系数得到,也就是

$$\begin{cases} u_{j,2k} = 2^{-1/2}(u_{j-1,k} + \tau_{j-1,k}) \\ u_{j,2k+1} = 2^{-1/2}(u_{j-1,k} - \tau_{j-1,k}) \end{cases} \quad (9)$$

应用基于 Haar 的小波模型到正定平稳具有长相关性信号 $X(t)$,为了保证信号的正定性,也就是 $u_{j,k} \geq 0, \forall j, k$, 要满足 $|\tau_{j,k}| \leq u_{j,k}, \forall j, k$. 假设 $A_{j,k}$ 是在区间 $[-1, 1]$ 符合 beta 分布的随机变量,并且通过以下的递归过程定义小波系数 $\tau_{j,k} = A_{j,k}u_{j,k}$, 从而,式(9)可以表示为

$$\begin{cases} U_{j,2k} = 2^{-1/2}(1 + A_{j+1,k})U_{j+1,k} \\ U_{j,2k+1} = 2^{-1/2}(1 - A_{j+1,k})U_{j+1,k} \end{cases} \quad (10)$$

为了把小波的多尺度分析应用到网络流量的预测中,这里主要考虑经过二项采样得到的离散信号 $X(k)$. 从而 $X(k)$ 可以很方便地通过 Haar 小波进行多尺度分析. 分析预测主要是通过逼近信号 $approx_j$ 进行. 在每一个尺度,根据二项关系对逼近信号进行预测,也就是对逼近信号前半部分进行线性预测的参数拟合估计,而后通过估计参数预测网络流量的后半部分. 对于多重分形小波模型,也就是逼近信号通过尺度系数表示,对逼近信号的预测也就是对尺度系数的预测,可以认为

$$approx_j = U_{j,k}, \quad j=0, 1, 2, \dots \quad (11)$$

目前所采用的大多数预测模型主要是线性模型. 在这些模型中,最为常用的是自回归模型 AR (Autoregressive)、滑动平均模型 MV (Moving Average) 和自回归滑动模型 ARMA (Autoregressive Moving Average).

假设时间序列 $X(t)$ 是一个 ARMA(p, q) 过程,如果 $X(t)$ 是平稳的,如果对每一个 t :

$$X(t) - \phi(1)X(t-1) - \dots - \phi(p)X(t-p) = Z(t) + \theta(1)Z(t-1) + \dots + \theta(q)Z(t-q) \quad (12)$$

这里 $Z(t) \sim WN(0, \sigma^2)$ 并且多项式

$$\begin{cases} 1 - \phi(1)Z - \dots - \phi(1)Z^p \\ 1 + \theta(1)Z + \dots + \theta(q)Z^q \end{cases} \quad (13)$$

没有共同的因子.

用 $\phi(\cdot)$ 和 $\theta(\cdot)$ 分别表示 P 阶和 Q 阶多项式算子, B 为前项移位算子,上述公式可以重写为如下形式:

$$\phi(B)X(t) = \theta(B)Z(t) \quad (14)$$

ARMA 模型假设数据是平稳的,如果时间序列显示可变性打破了平稳的假设,可以采用具体的方法用于产生平稳时间序列,最常用的方法是使用差分算子. 定义差度为 1 的差分算子 ∇ 为

$$\nabla X(t) = X(t) - X(t-1) = (1-B)X(t) \quad (15)$$

如果非平稳时间序列是时间的多项式方程,那么进行有限次差分之后可以使时间序列符合 ARMA 过程.

具有差分次数为 d 的 ARMA(p, q) 称为 ARIMA(p, d, q) 模型. 因此它具有这样一个形式:

$$\phi(B)(1-B)^d X(t) = \theta(B)Z(t) \quad (16)$$

这里 $Z(t) \sim WN(0, \sigma^2)$. 从而可以采用 ARIMA 模型预测网络流量的逼近部分 $approx_j$. 首先是最佳预测尺度的选择,可以根据文献[6]提出的最佳尺度选择方法选择所有尺度中预测结果最优的尺度. 其次是 ARIMA 模型参数的估计,采用逼近系数前半部分来估计 ARIMA 模型的参数,然后预测逼近系数后半部分. 设 $U_{j,k}^{pre}$ 是在逼近尺度下的预测值,参考式(10),可以得到

$$\begin{cases} U_{j,2k}^{pre} = 2^{-1/2}(1 + A_{j+1,k})U_{j+1,k}^{pre} \\ U_{j,2k+1}^{pre} = 2^{-1/2}(1 - A_{j+1,k})U_{j+1,k}^{pre} \end{cases} \quad (17)$$

从而根据式(6),可以得到相应的网络流量预测值.

4 仿真实验结果

在仿真实验中我们分析了一系列的公开网络流量测试集,包括各种网络业务,都获得了比较好的结果. 实验主要采用 AUCKLAND 网络流量测试集,流量踪迹来源于 NLNR 的 PMA 项目,是基于 Auckland 大学上行卫星线路通过 DAG3 系统捕获的 GPS 同步 IP 数据包头的踪迹. WAND 研究小组收集了很多代表性的网络聚集流量. 文献[6]对这一网络流量进行了较为详细的可预测性分析,主要结论如下:

(1)所有的预测率小于 1 的网络流量都具有可预测性,并且 80% 网络流量的预测率都偏离 1,预示着网络流量具有高可预测性.

(2)尽管线性预测算子存在多样性,对于网络流量来说,ARIMA 显现预测具有较好的预测效果.

(3)网络流量的可预测性以及预测精确度和尺度的选择有很大的关系,而且存在最佳预测尺度. 对于不同的网络流量,这个最佳预测尺度也是不

同的,可以说最佳预测尺度是网络流量本身的一个特性.

在这里只给出一个网络流量的预测结果. 为了便于小波分析的二尺度需要,选取的踪迹含有 4096000 个测量值,通过小波的多尺度分析,在最佳预测尺度 $j=12$ 下,得到相应的尺度系数为 1000 个点,如图 1(a)所示,将 1~250 个点用于 $ARIMA(p, d, q)$ 模型的参数估计,这里选取 $p=4, q=4, d=2$, 然后给出 251~1000 点的预测值以及相应的预测误差如图 1(b),图 1(c)所示. 从中可以看出在小波域内,尺度系数可以通过线性模型进行较为精确的预测. 根据式(17),可以得到网络流量的后 2048000 个点的预测值,如图 2(b)所示. 同时还给出了原始网络流量以及相应的预测误差,如图 2(a),图 2(c)所示. 从图中可以看出基于小波的网络流量预测得到了一个相当精确的预测值.

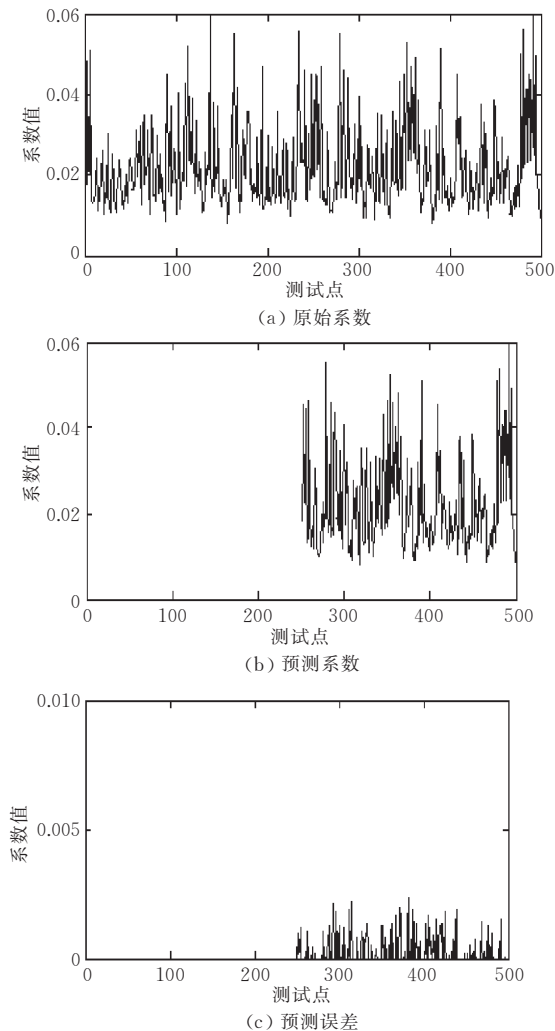


图 1 基于 ARIMA 的尺度系数预测

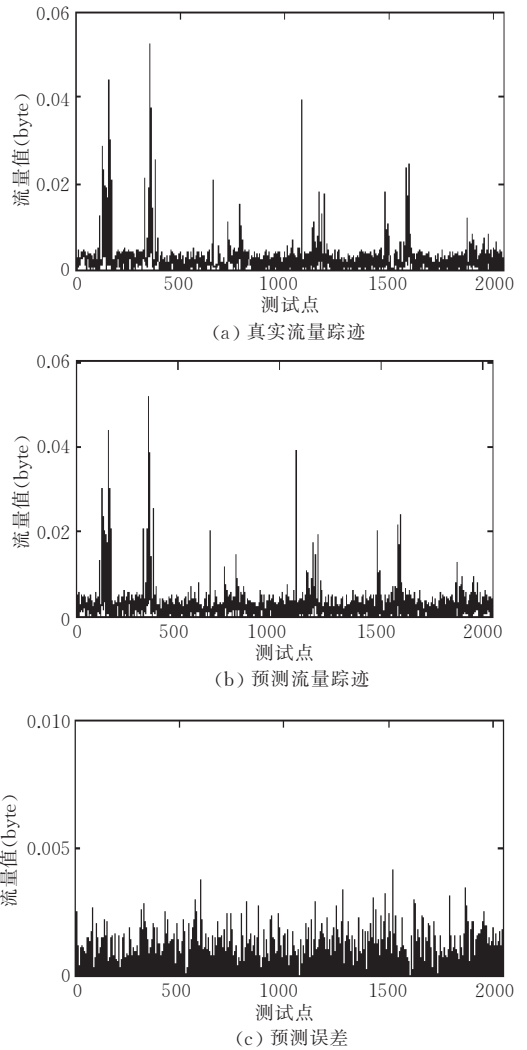


图 2 基于小波的网络流量预测

5 结 论

许多研究表明多尺度网络模型可以精确地把握网络流量的尺度本质,网络流量的预测对于网络容量规划、控制以及性能保证具有重要的意义. 因此我们根据网络流量的尺度本质和多尺度小波分析,给出了一个基于多重分形小波模型的网络流量预测模型. 由于小波所固有的尺度特性,仿真实验结果表明了该模型具有较高的精确度. 进一步研究可以确定自适应的预测尺度并应用到网络流量工程中.

参 考 文 献

- 1 Krunz M., Makowski A.. Modeling video traffic using M/G/infinity input processes: A compromise between markovian and LRD models. IEEE Journal on Selected Areas in Communica-

- tions, 1998, 16(5): 733~748
- 2 Leland W. E. , Taqqu M. S. , Willinger W. , Wilson D. V. . On the self-similar nature of ethernet traffic. *IEEE/ACM Transactions on Networking*, 1994, 2(1): 1~15
 - 3 Park K. , Kim G. , Crovella M. . On the effect of traffic self similarity on network performance. In: *Proceedings of SPIE International Conference Performance and Control of Network Systems*, Dallas, USA, 1997, 168~175
 - 4 Park K. , Willinger W. . *Self-Similar Network Traffic and Performance Evaluation*. Wiley-Interscience, 2000
 - 5 Paxson V. , Floyd S. . Wide-area traffic: The failure of poisson modelling. *IEEE/ACM Transactions on Networking*, 1995, 3(3): 226~244
 - 6 Konstantina Papagiannaki, Nina Taft, Zhang Zhi-Li, Christophe Diot. Long-term forecasting of Internet backbone traffic: Observations and initial models. In: *Proceedings of INFOCOM*, London, UK, 2003, 753~764
 - 7 Groschwitz N. K. , Polyzos G. C. . A time series model of long-term NSFNET backbone traffic. In: *Proceedings of IEEE ICC*, Pittsburgh, PA, 1994, 234~238
 - 8 Sang A. , Li S. . Predictability analysis of network traffic. In: *Proceedings of INFOCOM*, TelAviv, Israel, 2000, 342~351
 - 9 Abry P. , Veitch D. , Flandrin P. . Long-range dependence: Revisiting aggregation with wavelets. *Journal of Time Series Analysis*, 1998, 19(3): 253~266
 - 10 Abry P. , Flandrin P. , Taqqu M. S. , Veitch D. . Self-similarity and long range dependence through the wavelet lens. In: *Doukhan Paul, Oppenheim Georges, Taqqu Murad S. eds. Long-Range Dependence: Theory and Applications*. Birkhauser, 2002, 342~360
 - 11 Feldmann A. , Gilbert A. C. , Willinger W. . Data networks as cascades: Investigating the multifractal nature of internet WAN traffic. In: *Proceedings of ACM SIGCOMM*, Vancouver, Canada, 1998, 25~38
 - 12 Feldmann A. , Gilbert A. , Huang P. , Willinger W. . Dynamics of IP traffic: A study of the role of variability and the impact of control. In: *Proceedings of the ACM SIGCOMM*, Cambridge, MA, 1999, 28~39
 - 13 Riedi R. , Crouse M. , Ribeiro V. , Baraniuk R. . A multifractal wavelet model with application to network traffic. *IEEE Transactions on Information Theory*, 1999, 45(3): 992~1019
 - 14 Mallat S. . Multiresolution approximation and wavelets. *Transactions American Mathematics Society*, 1989, 2(3): 69~88
 - 15 Daubechies I. . *Ten Lectures on Wavelets*. USA: Society for Industrial and Applied Mathematics, 1999
 - 16 Yi Qiao, Jason Skicewicz, Peter Dinda. Multiscale predictability of network traffic. NWU, USA; Technical Report NWU-CS-02-13, 2002



HONG Fei, born in 1974, Ph. D. , lecturer. His research interests focus on network traffic model, WSN, WMN.

WU Zhi-Mei, born in 1942, professor, Ph. D. supervisor. His research interests focus on network access system.

Background

This project is supported by the National Basic Research Program of China (973 Program) under grant No. G1998030407 on high level network design and performance analysis.

Scaling and multiscale behaviours have been commonly viewed as the most significant characteristics of the Internet traffic today. Modelling and prediction of scaling and multiscale traffic is thus of great importance for planning, dimensioning, control, and performance guarantee of various types of networks.

For the multiscale nature of network traffic, the wavelet-based approach is a natural way to provide multiscale pre-

diction to applications. Riedi, *et al.* have shown how to use wavelets to synthesize network traffic, computing results in an efficient manner that appear to match real network traffic traces visually and statistically. Here with introduction of the ARIMA linear prediction in the wavelet domain, authors' work focus on the network traffic prediction based on the multifractal wavelet model. The simulation results show the accuracy of their method. Further research will focus on the application to the network engineering tasks, like scheduling of maintenance windows or large database network backups.