

网络流量预测中的时间序列模型比较研究

姜 明¹, 吴春明², 张 旻¹, 胡大民²

(1. 杭州电子科技大学软件与智能技术研究所, 浙江杭州 310018; 2. 浙江大学人工智能研究所, 浙江杭州 310027)

摘 要: 网络流量预测在新一代网络协议设计、网络管理与诊断、设计高性能路由器等方面都具有重要意义. 目前通常采用 ARMA 和 FARIMA 时序模型对网络流量序列进行拟合与预测, 但没有对时间尺度的大小与模型选择的关系进行研究. 本文对实际网络流量在不同时间尺度(毫秒、秒、分)下进行了流量预测建模并对预测性能进行比较, 分析表明使用时序模型进行流量预测时, 大时间尺度(分)流量预测较小时间尺度(毫秒、秒)具有更小的预测误差. 并且, 对于小时间尺度上的自相似流量序列, 自相似模型 FARIMA 并没有较其他时序模型有更好的预测性能.

关键词: 流量预测; 时序模型; 时间尺度

中图分类号: TP393

文献标识码: A

文章编号: 0372-2112 (2009) 11-2353-06

Research on the Comparison of Time Series Models for Network Traffic Prediction

JIANG Ming¹, WU Chun-ming², ZHANG Min¹, HU Da-min²

(1. Institute of Software and Intelligent Technology, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China;

2. AI Institute of Zhejiang University, Hangzhou, Zhejiang 310027, China)

Abstract: Network traffic prediction is very important in network protocol designing, network management and high performance routers designing etc. Currently, ARMA and FARIMA time series are the main models used to fit and predict the network traffic. But the relation between time scale and time series models hasn't been studied. The network traffic was modeled according to different time scales using the traffic trace data taken from the Internet traffic archive, and the prediction performance of those models was compared. The results show that time series models perform better under the large scale(minute) than under small time scales(millisecond and second), and the performance of self similarity model FARIMA shows no advantage over other models.

Key words: traffic prediction; time series models; time scale

1 引言

网络流量预测对于设计新一代网络协议、网络管理、设计高性能路由器和负载均衡器等网络硬件设备以及提高网络的 Qos 都具有重要意义^[1,2]. 流量预测通常采用的方法是利用时间序列模型, 对流量特性进行分析刻画和建模, 在此基础上进行流量预测. 目前的流量模型大致可分为平稳和非平稳两类. 平稳流量模型又可分为短相关和长相关两种模型^[3]. 短相关模型包括 Markov 模型^[4]、ARMA 模型^[5]等, 长相关模型包括 FARIMA 模型^[6]等. 其中 ARMA 模型、FARIMA 模型可实现预测功能, 非常具有实用价值. 但以往的文献^[5~8]大都只针对某一时间尺度的网络流量数据进行预测, 本文针对不同时间尺度(毫秒、秒、分)的实际网络流量, 使用传统时序模型和自相似模型, 分别对实际网络流量进行预测, 并比较各个时序模型的预测性能.

2 时序模型与预测

平稳时间序列是序列中不存在任何趋势性和周期性, 其统计意义就是一阶矩为常数, 二阶矩存在且为时间间隔的函数. 较常见的平稳时间序列模型有自回归模型(AR, Auto Regressive)、滑动平均模型(MA, Moving Average)以及自回归滑动平均模型(ARMA).

AR 模型是最常见的平稳时间序列模型, 可以表示为^[9]:

$$X_t = \varphi_1 X_{t-1} + \cdots + \varphi_p X_{t-p} + a_t \quad (1)$$

其中 a_t 为白噪声, $\varphi_i (1 \leq i \leq p)$ 为自回归系数.

如果时间序列当前时刻 t 的值 X_t 与其以前时刻的值 X_{t-1}, X_{t-2}, \cdots 无关, 而与其以前时刻 $t-1, t-2, \cdots$ 进入系统的扰动 a_{t-1}, a_{t-2}, \cdots 存在着某种相关关系, 那么这一类时间序列可用 MA 模型表示. MA(q) 模型可以表示为^[9]:

$$X_t = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \quad (2)$$

其中 a_t 为白噪声, $\theta_i (1 \leq i \leq p)$ 为滑动平均系数。

ARMA 模型描述的系统在时刻 t 的响应 X_t 不仅与其以前时刻的自身值有关,而且还与其以前时刻进入系统的扰动存在一定的依存关系. ARMA(p, q) 模型可以表示为^[9]:

$$X_t - \sum_{k=1}^p \varphi_k X_{t-k} = a_t - \sum_{k=1}^q \theta_k a_{t-k} \quad (3)$$

在这里我们引入后向算子 B , 它的运算有 $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$, 依此类推. 那么式(3)可以变换为:

$$\Phi(B) X_t = \theta(B) a_t \quad (4)$$

其中 $\Phi(B)$ 和 $\theta(B)$ 分别为后向算子 B 的 p 阶和 q 阶多项式

$$(1 - \varphi_1 B - \cdots - \varphi_p B^p) = \Phi(B) \quad (5)$$

$$(1 - \theta_1 B - \cdots - \theta_q B^q) = \theta(B)$$

实际中遇到的时间序列不一定是平稳的, 非平稳时序模型主要有自回归求和滑动平均模型 (ARIMA) 以及分数差分自回归求和滑动平均模型^[10~12] (FARIMA). ARIMA(p, d, q) 模型可以表示为:

$$\Phi(B)(1-B)^d X_t = \theta(B) a_t \quad (6)$$

其中 $1-B$ 为差分算子, d 为差分次数, 且取正整数. FARIMA(p, d, q) 模型和 ARIMA(p, d, q) 模型的唯一差别就是差分次数 d 可以取分数, 这时

$$(1-B)^d = \sum_{k=0}^{\infty} \Pi_k B^k \quad (7)$$

$$\text{其中: } \Pi_k = (-1)^k \binom{d}{k} = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} \quad (8)$$

FARIMA 模型的参数 d 可由式 $H = d + 0.5$ 求得, 其中 H 为序列的 Hurst 参数。

建立时序模型, 必须对模型参数进行估计, 以上介绍的各个时序模型在很多领域中都得到了广泛的应用, 而且都有一套成熟的关于建模和预测的方法. 其中参数估计方法有矩估计、最小二乘估计、极大似然估计等, 后两种方法可以得到较精确的参数值。

基于时序模型的预测同样有多种方法, 本文采用的是条件期望预测方法。

$$\bar{X}_t(l) = E(X_{t+l} | X_t, X_{t-1}, \cdots) \quad (9)$$

有关 X_t 与 a_t 的条件期望具有以下定则:

(1) 常量的条件期望是其本身;

$$E(X_k | X_t, X_{t-1}, \cdots) = X_k, (k \leq t)$$

$$E(a_k | X_t, X_{t-1}, \cdots) = a_k, (k \leq t)$$

(2) 未来扰动的条件期望为零, 即

$$E(a_k | X_t, X_{t-1}, \cdots) = 0, (k > t)$$

(3) 未来取值的期望为未来取值的预测值, 即

$$E(X_{t+l} | X_t, X_{t-1}, \cdots) = \bar{X}_{t+l}, (l > 1)$$

3 网络流量特性

网络流量序列会因网络环境(网络利用率、传输层协议等)、流量采集时间以及时间尺度等因素的不同而表现出不同的特性, 如短相关、长相关等. 研究发现实际的网络流量的相关特性是短相关和长相关^[13~15] 并存的, 还发现网络流量表现出自相似特性, 由于长相关是自相似特性的一个明显特征, 因此研究网络流量的自相似特性也就包含了其长相关特性. 值得注意的是, 自相似性包含了长相关特性, 但是长相关特性不一定包含自相似特性。

网络流量的自相似特性^[16~19] 可以定义如下: 考察一个广义平稳过程 $X = \{X_t, t = 1, 2, 3, \cdots\}$, 其中 X_t 表示网络中 t 到 $t+1$ 个时间单位内到达的网络业务数目(总字节数或总分组数等). 平稳时间序列 X 的期望为 $\mu = E(X)$, 方差为 $\sigma^2 = E[(X - \mu)^2]$, 自相关函数为 $r(k) = E[(X_t - \mu)(X_{t-k} - \mu)] / \sigma^2, k > 0$, 对每个 $m = 0, 1, 2, \cdots$,

$$X_k^{(m)} = \frac{X_{km-m+1} + X_{km-m+2} + \cdots + X_{km}}{m}, k = 1, 2, 3, \cdots \quad (10)$$

$X^{(m)} = \{X_k^{(m)}, k = 1, 2, 3, \cdots\}$ 是根据 X 得到的 m 阶聚合序列, 记 $\gamma^{(m)}$ 为 $X^{(m)}$ 的自相关函数. 如果对所有的 $m = 1, 2, 3, \cdots$, 都有 $\text{Var}(X^{(m)}) = \sigma^2 m^{-\beta}$ 且 $\gamma_k^{(m)} = \gamma_k$, 则称 X 为二阶自相似的, 其自相似系数为 $H = 1 - \beta/2$. 如果对足够大的 k 值, 有 $\gamma_k^{(m)} = \gamma_k, m \rightarrow \infty$, 则称 X 为渐进二阶自相似的。

4 网络流量预测

本节使用各种时序模型对实际网络流量数据进行建模和预测, 分析比较了各种时序模型的预测性能。

为了对预测效果进行评价, 引入了以下衡量指标, 其中 X_k 为序列实际观测值, \hat{X}_k 为 X_k 的预测值:

均方根误差 (RMSE):

$$RMSE = \sqrt{\frac{\sum_{k=1}^N (X_k - \bar{X}_k)^2}{N}} \quad (11)$$

相对均方根误差 (RRMSE):

$$RRMSE = \sqrt{\frac{\sum_{k=1}^N \left(\frac{X_k - \bar{X}_k}{\bar{X}_k} \right)^2}{N}} \quad (12)$$

显然这两项指标越小, 表明预测效果越准确。

流量预测实验的流量样本数据采用实际流量数据 LBL-tcp-3.tcp*, 该流量数据集为两个小时內总共采集

* 该流量数据可以从 <http://ita.ee.lbl.gov/html/contrib/LBL-TCP.html> 获得

的 1789995 个数据,每次采集都记录上一次采集时间到当前采集点间的数据字节数,将该数据集记为 TSa ,本文对该数据集进行操作,按 10ms 为采样间隔形成长度为 719999 的流量序列,记为 $ts10ms$,序列值的意义为上一次采集到这一次采集之间共 10ms 时间内总的字节数.同理,按照 100ms、1s、10s、100s 为时间间隔可以获得不同时间尺度的流量序列,分别记作 $ts100ms$ 、 $ts1s$ 、 $ts10s$ 、 $ts100s$.按照以上步骤,共获得 5 个流量序列,长度分别为 719999,71999,7199,719,71.实验过程使用如下序列进行预测:

$TS1 = ts10ms(1:2000)$,
 $TS2 = ts100ms(1:2000)$,
 $TS3 = ts1s(1:2000)$,
 $TS4 = ts10s$,
 $TS5 = ts100s$.

其中 $TS1$ 表示 $ts10ms$ 的前 2000 个数据, $TS2$ 表示 $ts100ms$ 的前 2000 个数据, $TS3$ 表示 $ts1s$ 的前 2000 个数据, $TS4$ 、 $TS5$ 分别表示 $ts10s$ 和 $ts100s$ 流量序列.

(a)基于传统时序模型的预测

本节使用 AR、MA、ARMA、ARIMA 时序模型对流量进行预测.预测过程中,对序列的后 30 个值进行预测,并且与实际观测值进行比较,预测结果如图 1 所示.

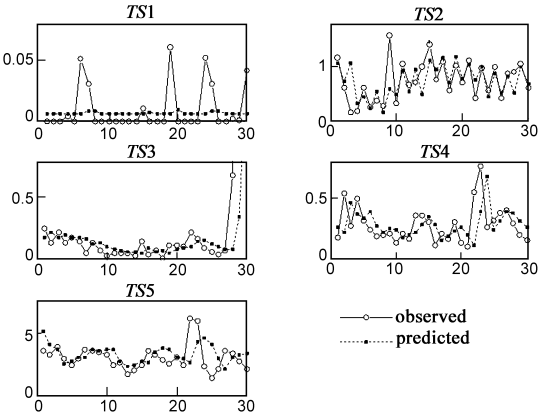


图1 预测结果

预测结果的 $RMSE$ 和 $RRMSE$ 指标如表 1 所示.

表 1 预测结果 $RMSE$ 和 $RRMSE$

序列	流量模型	$RMSE$	$RRMSE$
$TS1$	ARMA(0,1)	0.0185	较大*
$TS2$	高阶 AR	0.3084	0.4508
$TS3$	ARIMA(0,1,2)	0.3380	0.7437
$TS4$	ARIMA(0,1,3)	0.1711	0.4725
$TS5$	ARMA(0,1)	1.1930	0.3099

值得注意的是,由于不同序列的数量级不同,因此 $RMSE$ 值的绝对大小并不能说明预测结果的好坏,应该同时考虑 $RRMSE$ 值(下同).从图 1 与表 1 可以看出,对序列 $TS1$ 和 $TS3$ 的一步预测都有较大的误差,分析其

原因是由于序列的突发性, $TS1$ 在较短时间内在零值与较大值之间跳动,使得只记忆序列一期扰动的 ARMA(0,1)模型无法把握序列的突发特性, $TS3$ 在序列末尾处有一个较大突发,同样造成了较大的平均预测误差.

(b)基于自相似模型的预测

FARIMA 模型可看作一个分数差分噪声 FDN 驱动的 ARMA 过程,该模型既可以描述序列的长相关特性,又能描述序列的短相关特性,所以是一种自相似模型.实际的网络流量表现出自相似特性,从而使用 FARIMA 模型对流量序列建模,并基于此模型进行预测,以期获得较好的预测效果.

建立 FARIMA 模型需要对流量序列的 Hurst 参数进行估计.使用方差时间图法^[20](VT),R/S 法^[21,22],周期图法(PG)以及小波法^[23,24](AV)获得序列的 Hurst 参数,如表 2 所示.

表 2 不同时间尺度聚集序列的 Hurst 参数估计

序 列	VT	RS	PG	AV
$ts10ms$	0.6314	0.6235	0.5919	0.6022
$ts100ms$	0.9431	0.9277	0.9019	0.9960
$ts1s$	0.8636	0.8896	1.0160	0.9590
$ts10s$	0.8107	0.7986	0.9721	0.8271
$ts100s$	0.7787	0.6845	0.9991	0.6633

由于 Hurst 参数估计存在误差,而且从表 2 还可以看出不同方法估计出来的 Hurst 参数值也有所不同,但从估计值所处的区间来看, $ts100ms$ 、 $ts1s$ 和 $ts10s$ 具有较强的自相似特性,或者说具有较强的长相关特性.本文使用 RS 方法的估计值作为建模过程使用的 H 值.

本文参考了文献[25]提出的 FARIMA 模型提出以下基于 FARIMA 模型的网络流量预测算法.

FARIMA 单步预测算法:

输入:流量序列 X_t

输出:预测 \bar{X}_{t+1} .

第一步,流量序列 X_t 变换.首先对序列 X_t 去均值,再使用高阶自回归模型进行后向预报,形成新的序列 X'_t .根据式(9)进行 d 阶分数差分,获得新的平稳时间序列 W_t .

第二步,对 W_t 序列拟合 ARMA(p, q)模型,获得自回归系数 $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_p\}$ 和滑动平均系数 $\theta = \{\theta_1, \theta_2, \dots, \theta_q\}$,同时获取 ARMA 模型一步预测的模型残差序列 $resid = \{a_1, a_2, \dots, a_n\}$;

第三步,根据

$$W_t = (1 - B)^d X_t = \sum_{k=0}^{\infty} f(k) X_{t-k} \tag{13}$$

* $TS1$ 的 $RRMSE$ 值由于其存在较多的零值,使得 $RRMSE$ 值无法计算,但从直观上看,误差较大

可以得到

$$W_{t+1} = f(0)X_{t+1} + f(1)X_t + \cdots \quad (14)$$

从而有

$$\bar{X}_{t+1} = \bar{W}_{t+1} - \sum_{k=0}^{\infty} f(k)X_{t-k} \quad (15)$$

其中 \bar{W}_{t+1} 是 $\text{ARMA}(p, q)$ 模型对序列 W_t 的一步预测值, 可由第一步的结果获得. 根据以上预测公式, 可以获得 $\text{FARIMA}(p, d, q)$ 过程 X_t 的一步预测 \bar{X}_{t+1} .

按照上面的预测算法, 对实际流量序列 $TS2$ 、 $TS3$ 和 $TS4$ 进行预测. 预测结果如图 2 所示. 表 3 给出了预测结果的 $RMSE$ 和 $RRMSE$ 指标.

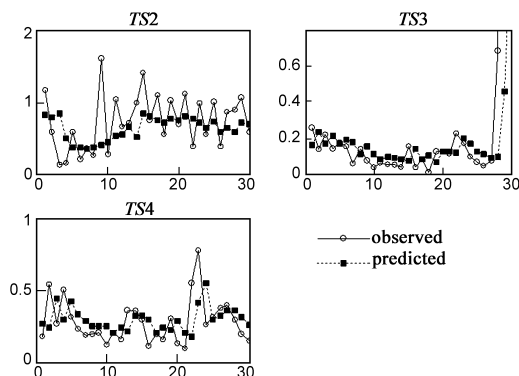


图2 FARIMA预测结果

表 3 FARIMA 预测结果

	拟合模型		$RMSE$	$RRMSE$
	d	ARMA 模型		
$TS2$	0.428	$\text{ARMA}(0, 1)$	0.3599	0.5095
$TS3$	0.39	$\text{ARMA}(0, 1)$	0.3414	0.8900
$TS4$	0.299	$\text{ARMA}(0, 1)$	0.1541	0.4795

(c) 预测结果分析之一

以上实验采用了传统的时间序列模型和 FARIMA 时间序列模型, 前者没有考虑流量的自相似特性, 后者有考虑, 通过对获得的流量序列的预测结果进行分析对比, 得到了以下结论:

第一, 基于时序模型的流量预测模型对较小时间尺度(毫秒级)的流量序列进行预测, 预测误差较大;

第二, 基于自相似性的 FARIMA 模型在网络流量预测过程中和其他时序模型相比, 并没有体现出优势;

本文通过大量实验进行分析, 认为基于时序模型的流量预测模型之所以不能对较小时间尺度的流量序列较好地预测, 原因在于小时间尺度下的网络流量具有很强的突发性、不确定性和复杂性, 难以把握其在统计上的规律特性, 而基于时序模型的预测的实质在于通过对历史数据的分析研究, 找出其特定的运行规律, 并据此对时间序列进行预测.

对于 FARIMA 模型在自相似网络流量预测过程中

的失效, 进一步分析认为导致这种结果的原因如下:

第一, 实际网络流量有很高的复杂性和不确定性, 网络流量本身呈现的自相似特性是一种渐近的二阶自相似特性, 并不是严格的二阶相似特性. 同时, 从 Hurst 参数的估计方法来看, Hurst 参数是对序列整体平均的自相似特性的描述, 因此这种整体的自相似特性不一定适用复杂网络流量未知时刻和已知时刻之间的相关性.

第二, FARIMA 模型的建模过程在于通过分数差分, 将具有长相关特性的序列转换成为短相关特性的序列, 再对获得的短相关序列进行 ARMA 建模. 在分数差分过程中, 由于样本序列有限, 通常使用后向预报法获得更长的样本序列, 而后向预报法一般基于高阶 AR 模型. 也就是说 FARIMA 模型是基于 ARMA 和高阶 AR 模型的, 而这两个模型的误差会在计算过程中累加.

综合考虑以上的实验结果, 本文认为自相似模型(FARIMA)在流量预测过程中并未体现出优势, 而且 FARIMA 模型由于其本身的复杂性, 使得建模和预测过程的计算复杂度都较高.

(d) 分钟时间尺度流量预测

以上都是对毫秒级和秒级的流量样本进行预测的, 接下来将对更大时间尺度(分钟级)的流量数据进行流量建模和预测, 分析时序模型对于更大时间尺度的流量序列的建模和预测效果.

流量样本数据采自 Brandeis 大学校园网的中心路由器于 2008 年 3 月 14 日 24 小时内的监测的流量数据, 按 5 分钟的时间尺度对该流量序列做聚集操作, 获得了用于建模的流量序列, 记为 TSb , 长度为 288.

对 TSb 进行 Hurst 参数估计, H 值并不在 0.5 与 1 之间, 因此认为该序列不存在自相似特性, 研究其自相关特性(见图 3), 发现 TSb 具有很强的长相关特性, 使用一阶差分对序列进行平稳化处理, 得到序列 TSb 一阶差分后的自相关函数, 发现一阶差分很好的去除了序列的长相关特性.

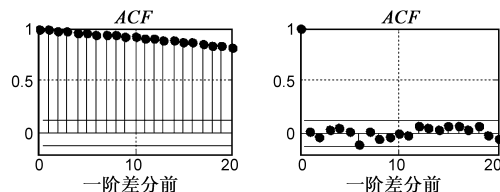


图3 TSb 的自相关函数值

使用 $\text{ARIMA}(p, 1, q)$ 模型对序列 TSb 进行建模, 获得了序列后 100 个数据的预测值, 如图 4 所示, 其 $RMSE$ 为 2.4935, $RRMSE$ 为 0.0192.

同时, 对序列的上升时段即第 101 到 200 间共 100 个数据进行预测, 图 5 给出了预测结果, 其 $RMSE$ 为 3.

1949, $RRMSE$ 为 0.0430.

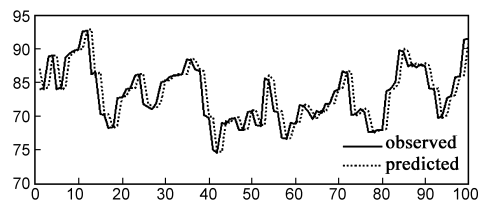


图4 TSb 一步预测结果, $TSb(189:288)$

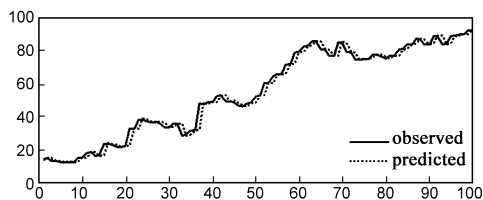


图5 TSb 的预测结果, $TSb(101:200)$

(e) 预测结果分析之二

从对分钟时间尺度流量预测实验结果中发现,时间序列模型对分钟级流量序列的预测达到了较好的效果,并且优于毫秒和秒级的预测.分析发现分钟时间尺度的流量序列不像较小时间尺度(毫秒和秒)下的流量序列那样具有较高的突发性和不确定性,也即分钟时间尺度的流量序列具有更好的平滑性,使得基于时序模型的流量模型能比较容易地把握流量序列的规律性.因此,本文认为,对于较大时间尺度的流量序列适合使用传统的时间序列模型进行建模分析,并且可以得到较好的预测效果.

5 结论

本文通过对流量序列按不同时间尺度(毫秒、秒、分)分别进行对比研究,发现使用时序模型进行流量序列的拟合与预测时,对较小时时间尺度(毫秒、秒)流量序列来说,流量预测结果并不理想,而对于较大时间尺度(分)的流量序列来说,流量预测能达到较好的效果.通过实验还发现,对于小时时间尺度上的自相似流量序列,自相似模型 FARIMA 并没有较其他时序模型有更好的预测效果.

从实际应用的角度来看,流量序列在小时间尺度上的建模与预测要求很高的实时性,而在大时间尺度上并不要求很强的实时性.因此,在小时间尺度上的流量建模和预测不适合采用复杂的流量模型,如何采用一种简单而有效的预测方法对小时时间尺度流量序列进行建模预测有很重要的应用价值,也尚待进一步研究.

参考文献:

[1] 王斌,戴琼海,丁嵘.基于 VBR 视频流量预测技术的 DVB-H 复用器调度算法[J].电子学报,2006,34(8),1522 – 1525.

Wang Bin, Dai Qiong-hai, Ding Rong. Scheduling algorithm based on VBR video traffic prediction for TS multiplexers in DVB-H networks[J]. Acta Electronica Sinica, 2006, 34(8), 1522 – 1525. (in Chinese)

[2] 蒋建国,苏兆品,李援,等.RTP/RTCP 自适应流量控制算法[J].电子学报,2006,34(9),1659 – 1662.

Jiang Jian-guo, Su Zhao-pin, Li Yua, et al. Adaptive flow control algorithm for RTP/RTCP [J]. Acta Electronica Sinica, 2006, 34(9): 1659 – 1662. (in Chinese)

[3] W Willinger, V Paxson, R H Riedi. Theory and Application of Long-Range Dependence[M]. Boston: Birkhauser, 2003. 625 – 716.

[4] H Heffes, D M Lucantoni. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance [J]. IEEE J. Select. Areas Commun, 1986, SAC-4: 856 – 868.

[5] 邹柏贤,刘强.基于 ARIMA 模型的网络流量预测[J].计算机研究与发展,2002,39(12):1645 – 1652.

Zou Bo-xian, Liu Qiang. ARMA-based traffic prediction and overload detection of network [J]. Journal of Computer Research and Development, 2002, 39(12): 1645 – 1652. (in Chinese)

[6] 李士宁,闫焱,覃征.基于 FARIMA 模型的网络流量预测[J].计算机工程与应用,2006,42(29):148 – 150.

Li Shi-ning, Yan yan, Qin Zhen. Network traffic forecast based on FARIMA model [J]. Computer Engineering and Applications, 2006, 42(29): 148 – 150. (in Chinese)

[7] Dongxu Shen, Joseph L Hellerstein. Predictive models for proactive network management: application to a production web server [A]. Hawaii: IEEE/IFIP Network Operations and Management Symposium [C]. 2000. 833 – 846.

[8] 吕军,李星.自适应网络流量线性预测算法及应用[J].计算机应用研究,2005,13(12):237 – 240.

Lv Jun, Li Xing. Network linear prediction based on adaptive filter theory [J]. Application Research of Computers, 2005, 13(12): 237 – 240. (in Chinese)

[9] 姚天任,孙洪.现代数字信号处理 [M]. 武汉:华中理工大学出版社,1999.

Yao Tian-ren, Sun Hong. Digital Signal Processing [M]. Huazhong University of Science and Technology Press, 1999. (in Chinese)

[10] Andrew, Sutcliffe. Time-series forecasting using fractional differencing [J]. Journal of Forecasting, 1994, 13(2): 383 – 393.

[11] Andel J. Long memory time series models [J]. Journal of Time Series Analysis, 1983, 22(4): 221 – 238.

[12] Jacek Ilow. Forecasting network traffic using FARIMA models with heavy tailed innovations [A]. IEEE International Conference on Acoustics, Speech, and Signal Processing [C]. Istanbul: IEEE Press, 2000. 3814 – 3817.

- [13] Sven A M, Harsha, Sirisena. The Influence of long-range dependence on traffic prediction[A]. IEEE International Conference on Communications [C]. Helsinki: IEEE Press, 2001. 1000 – 1005.
- [14] Karagiannis T, Faloutsos M, Riedi R. Long-range dependence: Now you see it, now you don't [A]. IEEE GLOBECOM' 2002 [C]. Taipei: IEEE Press, 2002. 2165 – 2169.
- [15] Beran J. Statistics Methods for data with long-range dependence[J]. Statistical Science, 1992, 7(4): 404 – 416.
- [16] Leland W E, et al. On the self-similar nature of Ethernet traffic[A]. ACM Sigcomm [C]. New York, ACM Press, 1993. 183 – 193.
- [17] Leland W E, Taqqu M S, Willinger W. On the self-similar nature of ethernet traffic (extended version) [J]. IEEE/ACM Transaction on Networking, 1994, 2(1): 1 – 15.
- [18] Grovella M, Bestavros A. Self-similarity in world wide web traffic: evidence and possible causes[J]. IEEE/ACM Transactions on Networking, 1997, 5(6): 160 – 169.
- [19] Willinger W, Taqqu M S, Sherman R. Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level[J]. IEEE/ACM Transactions on Networking, 1997, 1(15): 71 – 86.
- [20] Clegg, Richard G. A practical guide to measuring the hurst parameter[J]. International Journal of Simulation: Systems, Science & Technology, 2006, 2(7): 3 – 14.
- [21] Mandelbrot, Benoit B, Taqqu. Robust R/S analysis of long-run serial correlation[A]. Proceedings of the 42nd Session of the International Statistic Institute [C]. Manila: Philippine Organizing Committee, 1979. 69 – 99.
- [22] 郝清民. R/S 系列分析的非线性估计及应用[J]. 系统工

程理论与实践, 2005, 25(3): 80 – 85.

Hao Qing-min. Non-linear estimate and its application on R/S series analysis[J]. Systems Engineering-theory and Practice, 2005, 25(3): 80 – 85. (in Chinese)

- [23] Abry P, Veitch D. Wavelet analysis of long range dependent traffic[J]. IEEE Trans on Information Theory, 1998, 1(44): 2 – 15.
 - [24] Chen Huimin, Cai Hong, Li, Yanda. Self-similar traffic: hurst parameter estimation based on multiresolution sampling and wavelet analysis[J]. Acta Electronica Sinica, 1998, 26(7): 88 – 93. (in Chinese)
 - [25] 黄诒蓉. 中国股市分形结构: 理论与实证[M]. 广州: 中山大学出版社, 2006.
- Huang Yi-ron. Fractal Structure of China Stork Markets: Theroy and Demonstration[M]. Sun Yat-sen University Press, 2006. (in Chinese)

作者简介:



姜 明 男, 1974 年生于江苏如皋, 博士, 杭州电子科技大学软件与智能技术研究所副教授. 研究方向为网络服务质量、网络虚拟化和覆盖网.

吴春明 男, 1967 年生于浙江萧山, 博士, 浙江大学计算机学院教授, 博士生导师. 主要从事网络服务质量、可重构网络技术、网络虚拟化等领域方向的科学研究工作.

E-mail: wuchunming@cs.zju.edu.cn