

## Internet 流量模型分析与评述\*

张 宾<sup>1,2</sup>, 杨家海<sup>1,2+</sup>, 吴建平<sup>1,2</sup>

<sup>1</sup>(清华大学 信息网络工程研究中心, 北京 100084)

<sup>2</sup>(清华大学 清华信息科学与技术国家实验室, 北京 100084)

### Survey and Analysis on the Internet Traffic Model

ZHANG Bin<sup>1,2</sup>, YANG Jia-Hai<sup>1,2+</sup>, WU Jian-Ping<sup>1,2</sup>

<sup>1</sup>(The Network Research Center, Tsinghua University, Beijing 100084, China)

<sup>2</sup>(Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China)

+ Corresponding author: E-mail: yang@cernet.edu.cn

**Zhang B, Yang JH, Wu JP. Survey and analysis on the Internet traffic model. Journal of Software, 2011, 22(1): 115-131. <http://www.jos.org.cn/1000-9825/3950.htm>**

**Abstract:** The Internet traffic model is the key issue for network performance management, Quality of Service management, and admission control. The paper first summarizes the primary characteristics of Internet traffic, as well as the metrics of Internet traffic. It also illustrates the significance and classification of traffic modeling. Next, the paper chronologically categorizes the research activities of traffic modeling into three phases: 1) traditional Poisson modeling; 2) self-similar modeling; and 3) new research debates and new progress. Thorough reviews of the major research achievements of each phase are conducted. Finally, the paper identifies some open research issue and points out possible future research directions in traffic modeling area.

**Key words:** Internet traffic feature; network traffic modeling; self-similar; long-range dependence; fractal

**摘 要:** Internet 流量模型对网络性能管理、QoS、准入控制等都有很重要的意义和作用.首先总结了现阶段已发现的主要网络流量的特性及相关度量参数,概要地介绍网络流量建模的意义和分类,然后按照“传统-自相似-流量建模的新发展”这3个阶段阐述网络流量建模的发展历程与最新的研究成果,最后针对目前网络流量建模中存在的难点问题,展望了该领域未来的研究发展方向.

**关键词:** Internet 流量特征;网络流量模型;自相似;长相关;分形

中图法分类号: TP393 文献标识码: A

自互联网问世以来,关于网络流量的研究一直在不断的探索中.1993年,Leland等学者发现了数据网络的流量具有自相似的特性<sup>[1]</sup>,掀起了互联网流量特性与建模研究的热潮.随着互联网的快速发展、网络应用的多样化以及新型网络应用的快速部署,网络流量特性也随之不断发生变化.近年来,许多学者针对不断变化的流量新特

\* 基金项目: 国家重点基础研究发展计划(973)(2009CB320505); 国家科技支撑计划(2008BAH37B05); 国家高技术研究发展计划(863)(2008AA01A303, 2009AA01Z251)

收稿时间: 2010-06-10; 定稿时间: 2010-10-26

CNKI 网络优先出版: 2010-11-09 15:19, <http://www.cnki.net/kcms/detail/11.2560.TP.20101109.1519.000.html>

性,对网络流量的自相似性以及其它相关特性进行了重新审视.本文试图从一个较长的时间轴对互联网流量特性与建模研究工作进行比较系统的梳理,以期对从事相关研究工作的人员有所裨益.

传统电信网络的业务可以用经典的泊松过程进行很精确的描述,因此,早期学者们也尝试用泊松过程来描述互联网的流量行为.的确,根据许多现有的马尔可夫过程分析的结论和排队分析的一些基本方法,学者们能够在一定程度上对以近似泊松流作为业务源的排队系统进行性能分析,并且得到等待时间、队列长度等性能参数的近似结论.同时,泊松过程有着十分简单的相关结构,这使得泊松模型具有很好的性能.但是,泊松过程在大多数情况下并不能很好地描述互联网的流量行为.Leland 等人在 20 世纪 90 年代初发表的具有开创性意义的论文<sup>[1,2]</sup>中第一次明确提出了网络流量中存在着自相似现象,随后,Paxson 等人<sup>[3]</sup>对 WAN 流量、Klivansky 等人<sup>[4]</sup>对 NSFNET 流量和 Crovella 等人<sup>[5]</sup>对 WWW 流量的测试分析,均发现网络流量具有自相似特性,引发了流量模型从传统模型到自相似模型转变.21 世纪初,Karagiannis 等人<sup>[6]</sup>通过分析 Tier 1 ISP 的骨干链路流量,发现目前高带宽和高聚合的链路流量在亚秒尺度下近似泊松过程,从而引发了人们对网络流量特征及建模新的思索和争论.

本文首先总结现阶段已发现的主要网络流量的特性及相关度量参数,概要介绍网络流量建模的意义和分类,然后按照“传统-自相似-流量建模的新发展”这 3 个时段及分类阐述网络流量模型与最新的研究成果,最后展望该领域未来的研究发展方向,并针对目前网络流量建模中存在的难点问题及全文主要内容给出总结.

## 1 网络流量特性与相关度量参数

实际网络流量特性以及刻画这些流量特性的度量参数是流量建模的基础和依据,流量建模的目标是能够更好地反映实际流量的特征并且用于更好地指导实际的应用.当前,网络流量公认的、最重要的统计特征是大时间尺度下的自相似性和小时间尺度下的多分形性.这些特性不仅存在于互联网络中,同时也存在于 Ad Hoc 网络<sup>[7]</sup>以及卫星网络<sup>[8]</sup>中.

### 1.1 自相似和长相关

自相似(self-similar)是指局部的结构与总体的结构相比具有某种程度的一致性,自相似过程是在统计意义上具有尺度不变性的一种随机过程.从这一点上来说,自相似过程实际上是在随机过程中引入了分形的概念.网络流量具有长相关性,是相对于泊松等短相关(short range dependence,简称 SRD)模型而言的.从物理意义上看,长相关性(long range dependence,简称 LRD)反映了自相似过程中的持续现象,即突发特性在所有的时间尺度上都存在的现象,也称其为多尺度行为特性.长相关和自相似的具体定义和描述请参考文献[9-11].

自相似性与长相关性有着密切关系, $H \neq 1/2$  稳定的自相似随机过程是长相关的.但长相关过程不一定是自相似的.长相关性是渐近意义上的,它只描述了自相关函数在大延迟下的行为,而对任何固定有限大小延迟下的行为没有限制.而且,长相关性是侧重于业务量的统计特性,而自相似性则是偏重于数学表述上的.自相似过程的重要性还表现在,它为赫斯特效应(Hurst effect)这一经验规律提供了一种解释.Hurst 参数是表征自相似特性的一个重要参数,一般来说, $H$  值越大,自相似(长相关)程度越高,突发性也越强.常用的  $H$  值估计方法有以下几种(限于篇幅,不具体介绍,详见对应的参考文献):

- 方差-时间图(variance-time plot)法<sup>[12]</sup>;
- 聚合时间序列绝对值法(absolute values of the aggregated series)<sup>[12]</sup>;
- 差分方差方法(differencing the variance)<sup>[13]</sup>;
- Higuchi 方法<sup>[14]</sup>;
- 回归残差法(residuals of regression)<sup>[15]</sup>;
- R/S 图(rescaled adjusted range plot)法<sup>[16-18]</sup>;
- 周期图(periodogram)法及其改进方法<sup>[12]</sup>;
- Whittle 估计方法<sup>[19,20]</sup>;
- 小波分析估计法<sup>[21]</sup>.

## 1.2 多分形

在分形理论中,多分形又称作多重分形测度.对于许多非均匀的分形过程,一个维数无法描述其全部特征,需要采用多重分形测度或维数的连续谱来表示.Riedi 等人<sup>[22]</sup>通过对 TCP 流量的数值分析,指出长相关只是流量分形性质的一个方面(在较大的时间尺度上);在较小的时间尺度上,流量体现出更为复杂的变化规律,尤其是局部具有突发性.自相似在相对较大的时间尺度上刻画了流量的长相关特性,但无法描述业务的局部特征.因此,研究者又引入了多分形的概念.多分形延伸和细化了网络测量流量中的自相似行为(自相似即单分形),多分形性质体现了依赖于时间的尺度规律,在描述局部时间内的不规则现象时更加灵活.流量的分形特性已经广泛应用到许多工程问题中,如互联网流量建模、TCP 流量的拥塞控制、实时估计 LRD 过程的 Hurst 参数等.并且,分形理论和其他的数学理论相结合,如应用小波变换的自相似及多分形分析,在流量工程中正体现出越来越重要的作用.

## 1.3 周期及混沌

实际的网络流量除了人们熟知的自相似特性外,还有其他的一些流量特性不断地被研究人员提出并进行相关研究.如周期性<sup>[23]</sup>和混沌性<sup>[24-26]</sup>等.周期性变化特性反映的是网络流量时间序列随着时间的变化而表现出来的一种季节性的变化规律.所谓混沌就是指在确定性系统中出现的一种貌似无规则的、类似随机的现象,它普遍存在于自然界及人类社会中,是有序与无序的统一、确定性与随机性的统一.网络流量序列中存在着低维混沌,去掉长相关后的数据是一个随机流量数据,这反映了网络流量数据的混沌与网络流量的长相关是有密切联系的,从而为混沌与长相关的密切联系提供了一个有力的证据.因此,把混沌控制的思想和方法引入自相似业务流控制和建模预测是一个非常值得探索的研究方向.

## 2 网络流量建模概述

流量模型是流量行为特征的数学近似,网络流量建模的基本原则<sup>[10]</sup>是:以流量的重要特性为出发点,设计流量模型以刻画实际流量的突出特性,同时又可以进行数学上的研究.从理论角度来看,网络业务流的数学模型提供了对流量特性简明的、抽象化的描述,其价值在于能够提取出网络流量的一些重要特性,并给出一个明确的量化表示.随着网络规模的扩大和各种网络服务的广泛应用,建立一个能够准确、有效地描述网络流量特性的流量模型,对 QoS、网络性能管理、准入控制等都有很重要的意义和作用<sup>[9-11]</sup>.20 世纪 70 年代和 80 年代早期,由于当时网络的应用比较单一,数据传输量较小以及受到网络测量技术的限制,人们借鉴公共交换电话网络的模型,使用泊松模型来描述数据网络的流量,并取得不错的效果.进入 80 年代中后期,随着 FAX、数据网络和 Web 的出现,泊松过程已不能充分反映 Internet 业务流量的特性.人们随后又逐渐引入了 Markov 和回归等随机模型来描述网络流量.习惯上,人们称这些早期模型为传统网络流量模型,其共同特点就是所描述的业务序列具有短期相关性.20 世纪 90 年代以来,随着网络节点数呈指数增加和多媒体、视频、远程教育等网络应用的不断出现,流量特性变得更加复杂.由于不同的网络应用具有不同的流量特性,使得网络流量特性也发生了显著变化,这些都增加了网络流量特征化的困难.随着研究的深入,研究人员发现网络流量具有自相似特性,传统的流量模型已不能很好地描述网络的自相似性,而流量的自相似性又是网络的普遍属性并决定了网络的行为,因而基于自相似建模的研究便成为网络研究中的一个重要方向.2004 年,Karagiannis 等人<sup>[6]</sup>通过分析 Tier 1 ISP 的骨干链路流量,发现目前高带宽和高聚合的链路流量在亚秒尺度下近似泊松过程,从而引发了人们近年来对网络流量特征及建模的新的思索和争论.我们按照这 3 个发展的时段可以把网络流量模型<sup>[10,27]</sup>分为:

- 传统(短相关)模型(20 世纪 70 年代~1994 年);
- 自相似(长相关)模型(1994 年~2004 年);
- 流量建模的新发展(2004 年泊松回归引发的争论至今).

按照流量建模的方式,我们可以把目前的流量建模划分成两大范畴<sup>[28]</sup>:基于流的方法和基于包的方法.基于流的方法根据流的到达过程建模,基于包的方法又可以分为两类:一类是基于包的到达过程和包的长度分布来

建模,另一类是根据时间序列的流量大小分布特征来建模.为了更清楚地展现流量建模的发展过程,我们在后文中按时段发展顺序的分类方法进行阐述.

### 3 传统(短相关)模型

传统的网络流量模型一般是基于泊松过程的,这些模型产生的流量通常在时域上仅具有短相关性.随着时间分辨率的降低,即时间尺度变大,网络流量将趋于一个恒定值,即流量的突发性得到缓和.下面介绍几个经典的网络流量模型.

#### 3.1 泊松(Poisson)模型

泊松模型是 20 世纪初 Erlang 根据电话业务的特征提出来的<sup>[29]</sup>,最初用于电话网的规划和设计,可以较为准确地描述电话网中的业务特征并得到广泛的应用.在网络流量建模的早期,人们便使用泊松模型来研究网络流量.泊松模型即指在时间序列  $t$  内,包到达的数量  $n(t)$  符合参数为  $\lambda t$  的泊松分布,即

$$P_n(t) = \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad (n=0,1,2,\dots,N) \quad (1)$$

其相应的包到达的时间间隔序列  $T$  呈负指数分布,即  $F(t)=1-e^{-\lambda t}$ .其中,泊松过程的强度  $\lambda$  表示单位时间间隔内出现包数量的期望值,即包到达的平均速率,其值为  $\lambda=1/E(t)$ .泊松模型假设网络事件(如数据包到达)是独立分布的,并且只与一个单一的速率参数  $\lambda$  有关.泊松模型较好地满足了早期网络的建模需求,在网络设计、维护、管理和性能分析等方面发挥了很大的作用.然而,根据泊松流量模型,从不同的数据源汇聚的网络流量将随着数据源的增加而日益平滑,这与实际测试的流量是不符合的,因而该模型变得已不适于刻画实际的网络流量.

#### 3.2 马尔可夫(Markov)模型

对于一个给定的状态空间  $S=\{s_1,s_2,\dots,s_m\}$ , $X_n$  表示在  $n$  时刻状态的随机变量,如果  $X_{n+1}=s_j$  的概率只依赖于当前的状态, $\{X_n\}$  就形成了一个 Markov 链<sup>[30]</sup>.如果状态转换发生在离散时间序列  $(0,1,\dots,n,\dots)$ ,则称 Markov 链是离散的,否则称为连续的 Markov 链.Markov 属性意味着未来状态只依赖于当前状态,这使得描述一个状态持续时间的随机变量的分布呈指数分布(连续时间)或几何分布(离散时间).在一个简单的 Markov 流量模型中,每次状态转换代表一个新的到达,因此到达间隔呈指数分布.常见的 Markov 模型有以下几种:

- On-Off Source Model and IPP (the interrupted Poisson process) Model<sup>[31]</sup>;
- Alternating State Renewal Process<sup>[32]</sup>;
- Markov Modulated Poisson Process<sup>[33]</sup>;
- Markov Modulated Fluid Models<sup>[34,35]</sup>.

Markov 模型是利用某一变量的现在状态和动向去预测该变量未来的状态和动向的一种分析方法.Markov 模型在随机过程中引入相关性,可以在一定程度上捕获业务的突发性.同时,Markov 方法是一种具有无后效性的随机过程,应用十分广泛.Markov 模型的缺点是只能预测网络的近期流量,而且无法描述网络的长相关性.

#### 3.3 回归(regression)模型

在回归模型随机序列中,下一时刻的随机变量是由过去一个特定时间窗口中的随机变量以及一个白噪声移动平均值来决定的.常见的回归模型有以下几种:

- AutoRegressive (AR) Model<sup>[36]</sup>;
- Discrete AutoRegressive (DAR) Model<sup>[37-39]</sup>;
- AutoRegressive Moving Average (ARMA) Model<sup>[40,41]</sup>;
- AutoRegressive Integrated Moving Average (ARIMA) Model<sup>[42]</sup>;
- Transform-Expand-Sample (TES) Model<sup>[43-45]</sup>.

限于篇幅,这里仅简单介绍自回归(AR)模型.自回归模型是 Yule 在 1927 年为了预测太阳黑子的数目而提出来的,该模型采用自动回归法,强调时间序列未来的点数由同一时间序列过去的值来决定;在技术上,它采用线

性映射,用过去的值来映射未来的值,在给定的时间序列中选取函数的参数使得预测结果的误差最小。 $P$  阶自回归模型  $AR(p)$  随机变量的当前值  $X_t$  由过去  $p$  个值的线性组合加一个白噪声扰动项  $\varepsilon_t$  组成,形式如下:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2)$$

其中,  $\phi_j$  为实数.使用 AR 模型预测时,只需求解线性方程组,计算相对简单.因此,AR 模型的应用很广泛.尽管 AR 模型易于计算,但其自相关函数以指数形式衰减,所以不能很好地模拟比指数衰减要慢的自相关结构的流量.

### 3.4 传统模型的不足

传统流量模型的优点是相应的概率理论知识发展比较完善,队列系统性能评价易于数学解析.由于传统的业务模型只有短相关性,即在不同的时间尺度上有不同特性,从而无法描述网络的长相关性.从传统模型得到的结论是:这些模型仿真产生的业务,通常在时域仅具有短相关性,当业务源数目增加时,突发性会被吸收,聚合业务变得越来越平滑,不能反映业务突发性;而且,传统模型产生的业务流高频成分多而低频成分少,相关结构呈指数衰减,因而不能准确地描述流量自相似性.总结起来,有以下几点:

- 1) 实际的数据包和大部分连接的到达是相关联的,并不严格服从泊松分布;
- 2) 传统的业务模型只具有短相关性,而流量自相似性反映业务在较大时间尺度具有突发性,对缓存的占用比传统排队论的分析结果要大,这样会导致更大的延时.这说明泊松到达流量模型会降低网络的性能;
- 3) 对于传统模型,当业务源数目增加时,突发性会被吸收,聚合业务会变得越来越平滑,但却忽略了流量的突发性.

## 4 自相似(长相关)模型

自从 1994 年流量的自相似特性被发现后,各种基于自相似性的流量模型被不断地提出.基于网络流量的自相似性,有两类建模方式:一类是构造建模(物理模型),这类方式试图利用已知的传输知识来解释所观察到的数据特征,如由于资源共享而导致大量信源叠加的事实,这类建模方式中具有代表性的有重尾分布的 ON/OFF 模型、Alpha-Beta on/off 模型以及  $M/G/\infty$  排队模型;另一类是行为建模(统计模型),这类方法试图用数据拟合方法模拟所测量真实数据的变化趋势,代表模型有 FBM 模型和基于小波的模型等.

### 4.1 重尾分布的 ON/OFF 模型

模型定义为叠加大量的 ON/OFF 源,每个源都有两个周期交替的 ON 和 OFF 状态.在 ON 状态,数据源以连续的速率发送数据包;在 OFF 状态,不发送任何数据包.其中,每个发送源 ON 或 OFF 的时长独立地符合重尾分布(如 Pareto 分布).传统的 ON/OFF 模型假定 ON 态和 OFF 态的持续时间均以指数形式分布.扩展这种模型使 ON 态和 OFF 态的持续时间有无限的方差(即高可变性或 Noah 效应)<sup>[46,47]</sup>,这样,无数个源的叠加就呈现出长相关性(Joseph 效应)<sup>[48]</sup>.Alpha-Beta on/off 模型<sup>[49]</sup>在 ON/OFF 模型的基础上进一步把高速率、高容量的连接定义为 Alpha 流量,把低速率、低容量的连接定义为 Beta 流量.Alpha 流量占全部连接的很少一部分(少于 0.1%),而对整个流量的属性有很大的影响,Beta 流量基本上表现为高斯边缘分布.此模型分别用相应的 ON/OFF 模型生成对应的 Alpha-Beta 流量,然后合成.

用 ON/OFF 模型叠加产生自相似流量可以解释产生自相似的部分原因:经检测发现<sup>[5]</sup>,若文件大小符合重尾分布,则对应的文件传输均导致链路层的自相似性,而与所用的传输协议等相关较小.这种模型包含明确的物理意义,有助于深入地了解自相似的本质.其缺点在于,假设前提过于严格,即各个源端必须是独立同分布的,且输出速率为常数,而大多数网络业务的分布是无法建立在此前提上的.这些都使得它在实际应用中受到很大限制.

### 4.2 $M/G/\infty$ 排队模型

排队论的基本思想是 1910 年丹麦电话工程师 Erlang 在解决自动电话设计问题时开始形成的,当时称为话务理论.图 1 为一最简单的排队系统模型.排队系统包括 3 个组成部分:输入过程、排队规则和服务机构.排队系

统一般是以顾客相继到达系统的间隔时间分布、服务时间的分布和服务台数目为分类标志.现代常用的分类方法是英国数学家肯德尔提出的分类方法,即用肯德尔记号  $X/Y/Z$  进行分类. $X$  处填写相继到达间隔时间的分布, $Y$  处填写服务时间分布, $Z$  处填写并列的服务台数目.各种分布符号有: $M$ -负指数分布, $D$ -确定型, $E_k$ - $k$  阶埃尔朗分布, $GI$ -一般相互独立分布, $G$ -一般随机分布等.

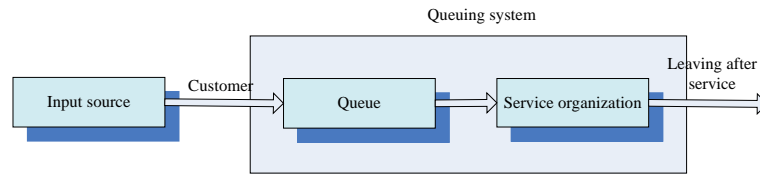


Fig.1 Queuing system model

图 1 排队系统模型

用  $M/G/\infty$  排队模型构造自相似序列的方法最早是由 Cox 提出来的,于 1998 年被 Krunzly 用于视频流量的建模<sup>[50]</sup>.结果显示,此模型能够较好地反映实际流量的排队性能. $M/G/\infty$  模型表示:输入顾客流服从参数为  $\lambda$  的 Poisson 过程(因  $M$  表示相继到达的时间间隔呈负指数分布),系统内有无穷个服务设备,每个服务设备的服务时间  $T$  服从独立同分布  $G$ . $M/G/\infty$  序列是指排队系统中的顾客总数在时间轴上构成的序列. $M/G/\infty$  模型可以通过选取不同的  $G$  使序列具有长/短相关的结构,系统的服务时间  $G$  服从 Pareto 分布的时候,顾客总数序列构成一个渐进自相似过程. $M/G/\infty$  序列无法直接用概率密度或分布函数描述.改进后的  $M/G/\infty$  模型<sup>[28]</sup>(包间隔即顾客流用 Pareto 分布代替指数分布)生成的流量更能反映真实流量的排队特性.

$M/G/\infty$  排队模型也是一种采用构造方式的自相似网络流量模型.由于现在 IP 网络设备都基于分组交换,并且在设备的接口上都采用了统计复用的实现方式,所以该模型的一个优点在于从排队系统的角度解释了网络流量产生自相似特性的原因;另外一个优点是该模型比较适合于分析自相似网络流量输入时的排队性能.但是,该模型假设了服务器一直处于忙期,主要凭借服务时间的随机性来描述自相似特性,因此对网络流量的突发性描述方面存在不足.

### 4.3 FBM/FGN模型

分形布朗运动(fractional Brownian motion,简称 FBM)是由 Manderbrot 和 Van Ness 提出的一种统计自相似过程的数学模型<sup>[51]</sup>,主要用于生成布朗运动过程.其定义如下:

定义(FBM). 设  $X(t)$  为一般布朗运动,称满足如下条件的随机过程  $X_H(t)$  为分形布朗运动:

$$X_H(0) = 0, X_H(t) - X_H(0) = 1/\Gamma(H+1/2) \left\{ \int_{-\infty}^0 [(t-s)^{H-1/2} - (-s)^{H-1/2}] dX(s) + \int_0^t [(t-s)^{H-1/2}] dX(s) \right\} \quad (3)$$

其中,  $t > 0, 0 < H < 1, X(s)$  为一般布朗运动.当  $H=1/2$  时,FBM 即为一般布朗运动.FBM 是一种不平稳的自相似过程,其自相似系数为  $H$ .FBM 是一个均值为 0 的连续高斯过程,其平稳增量过程是分形高斯噪声 FGN(fractional Gaussian noise).令  $Z_H(k) = X_H(k) - X_H(k-1)$ ,则  $Z_H(k)$  即为 FGN,FGN 是平稳的严格二阶自相似过程.

在此基础上,Norros 提出了一个自相似网络业务流模型<sup>[52]</sup>.令  $A_t^{(i)}$  为第  $i$  个信源在时间  $[0, t]$  内输入的网络业务流,其输入平均到达速率为  $m$ ,网络的聚合业务流的形式化表示如下:

$$A_t = mt + \sqrt{am} X_t, \quad t \in (0, +\infty) \quad (4)$$

$A_t$  表示到时刻  $t$  为止的所有网络业务流.其中,  $m$  为整个网络流量的平均到达速率,  $a > 0$  为方差系数,  $X_t$  为标准的分形布朗运动且其自相似系数  $H$  满足  $0.5 < H < 1$ .产生分形布朗运动的主要算法是 RMD<sup>[53]</sup>法,但此算法生成业务的 Hurst 系数与期望值不一致:当  $0.5 < H < 0.75$  时,其值偏大;而当  $0.75 < H < 1$  时,其值偏小;尤其是当  $H=0.5$  时,生成的业务数据与标准的布朗运动有较大偏差.另一种方法是通过对分形高斯噪声的频谱进行快速傅里叶逆变换而获得业务数据,所生成的业务源 Hurst 指数具有较好的一致性,而且业务数据样本的边缘分布非常接近高斯分布<sup>[54]</sup>.此外,还有采用小波变换<sup>[55]</sup>的方法和线性近似<sup>[56]</sup>的方法产生分形布朗运动.

FBM 模型能够描述网络业务流的自相似特性,只需要平均速率  $m$ 、方差  $a$  和 Hurst 参数 3 个参数就可以完整地刻画整个模型,在数学上有坚实的理论基础且比较好处理,因而可以很方便地应用于流量的实时仿真和特性分析.FBM 模型分析网络流量时也存在一些不足:由于 FBM 是严格自相似的过程,模型的参数较少使得其描述能力有限,可以用来对长相关数据进行建模,但无法描述业务的短相关特性,从而不能对既有长相关特性又有短相关性的流量准确建模;而且,FBM 模型带有高斯性,对于非负的信号(即非高斯性的信号)也不能很好地分析.

#### 4.4 FARIMA 模型

分形  $ARIMA(p,d,q)$ <sup>[57,58]</sup>过程(fractional autoregressive integrated moving average)是  $ARMA(p,q)$ 的一个扩展形式,在  $d$  为 0 时即为  $ARMA(p,q)$ 模型,其定义如下.

定义(FARIMA). 若均值为 0 的平稳随机序列  $\{X_k\}$  满足如下条件:

$$\begin{cases} \phi(B)\Delta^d X_k = \theta(B)\varepsilon_k, & d \in (-0.5, 0.5), B \text{ 是延迟算子 } BX_k = X_{k-1} \\ \phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p, \text{ 分形差分算子 } \Delta^d = (1-B)^d = \sum_k C_d^k (-B)^k \\ \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, C_d^k (-1)^k = \Gamma(-d+k)/(\Gamma(-d)\Gamma(k+1)) \end{cases} \quad (5)$$

则称  $X_k$  是  $FARIMA(p,d,q)$ 过程.其中,  $\Gamma$  为 Gamma 函数,  $\varepsilon_k$  为高斯过程.当  $k \rightarrow \infty$  时,其自相关函数为

$$\rho_x(k) \approx \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1} \quad (6)$$

因此,  $FARIMA(p,d,q)$  是二阶渐进自相似过程,且具有自相似参数  $H=d+1/2$ .  $FARIMA$  是一个时间序列模型,通过  $p,d,q$  这 3 个参数来控制自相关结构,用  $p+q+1$  个参数刻画样本中的短相关结构;采用  $d=H-0.5$  描述样本的长相关结构.参数  $d$  的取值区间不同,  $FARIMA$  过程的特性也不同.如果  $p=q=0$ ,即  $FARIMA(0,d,0)$ ,它是  $FARIMA(p,d,q)$ 过程的最简单的形式,一般称为分形差分噪声.事实上,当  $0 < d < 0.5$  时,  $FARIMA(p,d,q)$ 过程可以被看作是一个分形差分噪声  $FARIMA(0,d,0)$ 驱动的  $ARMA(p,q)$ 过程,其数学表达为

$$X_k = \phi^{-1}(B)\theta(B)Y_k \quad (7)$$

其中,  $Y_k = \Delta^{-d}\varepsilon_k$  是  $FARIMA(0,d,0)$ 中的分形差分噪声.

分形  $FARIMA(p,d,q)$ 算法其实就是先产生分形差分噪声  $FARIMA(0,d,0)$ ,然后利用分形差分噪声驱动  $ARMA$ 模型获得  $FARIMA$ 模型.实现分形差分算子是  $FARIMA$ 网络流量建模的一个关键,可以利用第 1.1 节的 Hurst 参数估计法间接地对  $d$  进行近似估计.

$FARIMA(p,d,q)$ 是一种渐近二阶自相似过程,可以有效地描述样本流量的长相关特性,同时也能很好地表示具有短相关结构的业务流量.但是,由于模型本身的复杂性和参数较多,计算量很大,算法复杂性为  $O(n^2)$ ,使其在实际应用中存在一定的局限性.

#### 4.5 基于小波的模式

小波变换是 20 世纪 80 年代后期在泛函分析、数值分析、逼近论和傅里叶分析基础上发展起来的一个应用数学分支.经过多年的发展,小波分析被广泛地应用于信号处理、图像处理、模式识别、数字水印等相关领域中.具有多分辨率,也叫多尺度的特点,可以由粗及细地逐步观察信号.小波分析是一种窗口大小(即窗口面积)固定但其形状可以改变、时间窗和频率窗都可以改变的时频局部化分析方法(即在低频部分有较高的频率分辨率和较低的时间分辨率,在高频部分具有较高的时间分辨率和较低的频率分辨率),可以根据实际分析需要自适应地调节时频窗口,能够聚焦到信号时域和频域的任意细节.自 20 世纪 90 年代开始,研究人员逐渐将其引入到网络模型的研究中,依靠它的多尺度特性来进一步揭示网络中的流量特征<sup>[59]</sup>.

随机信号  $X(t) \in L^2(\mathbb{R})$  的连续小波变换(continuous wavelet transform,简称 CWT)定义为

$$W_{ab} = \langle X, \psi_{a,b} \rangle = \int_{\mathbb{R}} X(t) \psi_{a,b}^*(t) dt = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} X(t) \psi^* \left( \frac{t-b}{a} \right) dt, \psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right), a > 0, b \in \mathbb{R} \quad (8)$$

称  $W_{a,b}$  为小波系数,  $\psi(t)$  为小波母函数,  $\psi(t)$  经过尺度为  $a$  的伸缩和参数为  $b$  的平移后得到正交基  $\psi_{a,b}(t)$ , 则

随机信号  $x(t)$  可由其小波系数重构为

$$X(t) = \frac{1}{C_v} \int_0^{+\infty} \int_{-\infty}^{+\infty} W_{a,b} \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \frac{da}{a^2} db \quad (9)$$

即小波变换可逆.经逆小波变换,信号  $x(t)$  被分解成为互为正交的小波基函数的线性组合,小波系数包含了随机信号  $x(t)$  的所有信息.由于连续小波变换是一种冗余变换,小波在空间两点之间的关联增加了分析的难度.因此,要把小波变换应用到实际的网络流量分析中,需要使用离散小波变换来更好地处理离散的数据信号.离散小波变换(discrete wavelet transform,简称 DWT)就是对连续小波变换中的伸缩因子  $a$  和平移因子  $b$  进行离散化处理,通过对放大倍数的改变,可以实现对信号的多尺度分析.最常用的方法是将信号通过一个高通的小波母函数  $\psi(t)$  和低通的尺度函数  $\phi(t)$ ,如公式(10)所示,然后求小波和尺度系数,如公式(11)所示:

$$\begin{aligned} \psi_{j,k}(t) &= 2^{j/2} \psi(2^j t - k) \\ \phi_{j,k}(t) &= 2^{j/2} \phi(2^j t - k), j, k \in \mathbb{Z} \end{aligned} \quad (10)$$

$$\begin{aligned} W_{j,k} &= \int X(t) \psi_{j,k}(t) dt \\ U_{j,k} &= \int X(t) \phi_{j,k}(t) dt \end{aligned} \quad (11)$$

称  $W_{j,k}$  为小波系数(wavelet coefficient),表示信号在区间  $[k2^{-j}, (k+1/2)2^{-j}]$  和  $[(k+1/2)2^{-j}, (k+1)2^{-j}]$  的均值差,提供尺度  $j$  到  $j+1$  的细节信息;  $U_{j,k}$  为小波尺度系数(scaling coefficient),表示信号在区间  $[k2^{-j}, (k+1)2^{-j}]$  的均值,可用在尺度  $j$  上逼近信号.重构信号(小波离散逆变换)的具体形式为

$$X(t) = \sum_k U_{j_0,k} \phi_{j_0,k}(t) + \sum_{j=j_0}^{\infty} \sum_k W_{j,k} \psi_{j,k}(t) \quad (12)$$

常用的小波函数有 Haar 小波、Morlet 小波、Maar 小波等.常见的小波模型均基于 Haar 小波. Haar 小波函数和尺度函数构成了一个简单的小波正交基, Haar 小波的尺度系数和小波系数有如下关系(如公式(13)、公式(14)和对应的图 2(a)、图 2(b)<sup>[60]</sup>所示):

$$\begin{aligned} U_{j,k} &= 2^{-1/2} (U_{j+1,2k} + U_{j+1,2k+1}) \\ W_{j,k} &= 2^{-1/2} (U_{j+1,2k} - U_{j+1,2k+1}) \end{aligned} \quad (13)$$

$$\begin{aligned} U_{j+1,2k} &= 2^{-1/2} (U_{j,k} + W_{j,k}) \\ U_{j+1,2k+1} &= 2^{-1/2} (U_{j,k} - W_{j,k}) \end{aligned} \quad (14)$$

用这样的方法来处理离散信号  $X(k)$ ,并假设  $X(k)$  的长度为  $2^n$ ,从而  $X(k)$  与最精细尺度系数关系如下:

$$X(k) = 2^{-n/2} U_{n,k}, k=0,1,\dots,2^n-1 \quad (15)$$

从上面的分析可以看到,小波变换具有对信号的自适应性,能够保持分析对象的尺度不变性.由于网络流量的自相似性是在统计意义上具有尺度不变性的一种随机过程,因此,小波变换在数学上具有其自身特有的优势,下面的小波模型都是建立在对网络流量多尺度分析的基础上.

#### (1) 小波域独立高斯 WIG(wavelet-domain independent Gaussian)模型<sup>[61]</sup>

小波域独立高斯 WIG 模型可以用来合成一个高斯过程,基本过程是按图 2(b)<sup>[60]</sup>所示逐步由  $j$  层的尺度系数和小波系数生成  $j+1$  尺度下的尺度系数,其具体合成方法如下:

- 1) 首先生成小波尺度系数树的根节点  $U_{0,0}$ ,它是一个高斯随机变量;
- 2) 然后生成尺度  $j$  上的各个小波系数  $W_{j,k}$ ,它们是相互独立的零均值高斯分布随机变量,只要在不同尺度  $j$  上小波系数的方差满足幂律衰减,就可以实现对分形布朗运动或是分形高斯噪声的合成<sup>[62]</sup>;
- 3) 要计算更小尺度上的尺度系数,可以用生成的上一级尺度系数和相应的小波系数由上面的变换公式得出.通过递归计算,直至求得最精细尺度  $n$  上的  $2^n$  个尺度系数为止,最终可以得到尺度系数序列  $\{U_{n,k}\}$ ,从而获得所求的合成长相关信号  $X(k)$ ,如公式(15)所示.

对于合成长度为  $N$  的信号,这种合成方法的计算复杂度只有  $O(n)$ . WIG 不仅能够表达随机过程中的长相关结构,还能表达其中的短相关结构. WIG 是一个单分形的加法模型,可以同时对流量的长、短相关特性进行描述,



其研究价值非常大.同时,WIG 的计算复杂度低,使其可以被方便地应用到实际中.但是,WIG 模型仍然是高斯的,对于突发的网络流量参数无法进行完整的描述;当方差大于平均值时,WIG 合成的数据会出现负数,这也与实际不相符.

(2) 多重分形小波模型 MWM(multi-fractal wavelet model)<sup>[62]</sup>

由于独立小波模型的高斯本性,使用 WIG 并不能对实际网络中的小时间尺度下的突发状况进行把握,并且由独立小波模型产生的信号量不能保证非负性.因此,独立小波模型并不能完全体现实际网络的真实特性.具有非高斯性质的 MWM 为了保证尺度系数的非负性,对 Haar 小波变换作了特殊的限制,将小波能量的衰减看作尺度的函数,用于网络业务的突发性建模,对小波系数和尺度系数增加一些限定条件,以此来保证信号的非负性.注意到小波逆变换迭代过程中,只要最粗糙尺度的尺度系数非负,并且满足 $|W_{j,k}| < U_{j,k}$ ,则迭代的每个公式都能保证尺度系数非负.MWM 模型中引入了因子  $A_{j,k}$ ,并使得

$$W_{j,k} = A_{j,k} \times U_{j,k} \tag{16}$$

如图 2(c)<sup>[60]</sup>所示,其中,  $A$  是 $[-1,1]$ 之间的独立随机变量,可以保证迭代中所有的尺度系数非负.MWM 模型产生模拟流量数据序列的过程可以简要描述如下:

- 1)  $j=0$ ,生成一个最粗糙的(根)尺度系数  $U_{0,0}$ ;
- 2) 在尺度  $j$  上,产生随机变量  $A_{j,k}$ (可选  $A_{j,k}$  为对称  $\beta$ 分布),并通过  $W_{j,k} = A_{j,k} \times U_{j,k}$  计算  $W_{j,k}, k=0,1, \dots, 2^j-1$ ;
- 3) 在尺度  $j$  上,用  $U_{j,k}$  和  $W_{j,k}$  由小波逆变换计算出尺度  $j+1$  的  $U_{j+1,2k}$  和  $U_{j+1,2k+1}, k=0,1, \dots, 2^j-1$ ;
- 4) 增加  $j$ ,重复步骤 2)、步骤 3),直至达到尺度  $j=n$  为止.

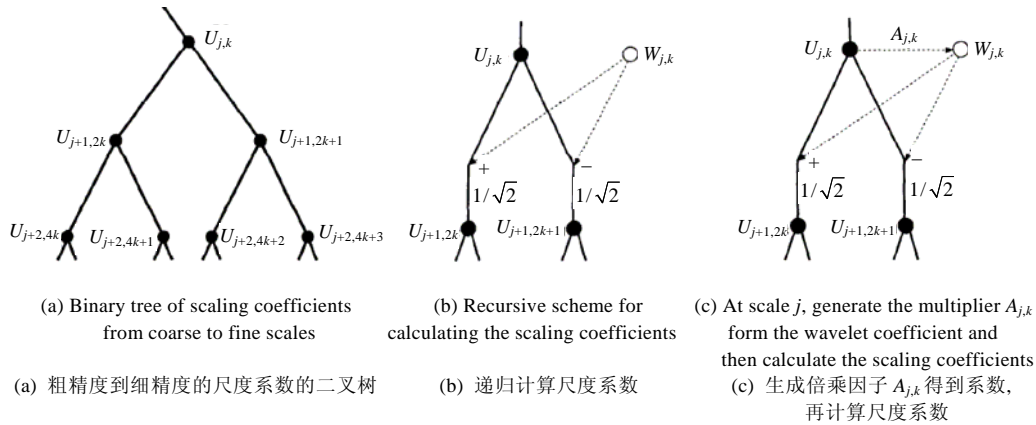


Fig.2  
图 2

MWM 模型是一个多分形的乘法模型,用较少的参数就能对网络流量中的短相关和长相关进行描述,还能匹配实际流量小尺度下的多分形特性,且能达到比较快速的收敛.其算法复杂度也是  $O(n)$ ,可以很好地匹配实际网络流量.不足之处是,小波变换系数并非在每个尺度下都独立,而且小波基的选取也影响模型的质量.

4.6 自相似模型小结

这一时期还有许多其他的自相似模型,如确定性的混沌映射模型<sup>[24-26]</sup>、基于 MMFM 的改进模型<sup>[63]</sup>、基于 WIG 的改进模型<sup>[60]</sup>、基于流的模型<sup>[64]</sup>、基于  $\alpha$ 稳态分布的模型<sup>[65]</sup>等,这里不再一一介绍.自相似流量模型与传统流量模型的不同之处在于:自相似模型是建立在网络特性的基础上,可以描述流量的突发性和长相关性,刻画了业务流量的自相似特性,有助于全面地认识网络业务流在各个方面的内在规律.表 1 对文中的自相似模型作了一个简单的对比.

**Table 1** Comparison of self-similar models**表 1** 常用几种自相似模型对比

Model	ON/OFF model	$M/G/\infty$ queuing model	FBM/FGN	$FARIMA(p,d,q)$	Wavelet based model
Physical meaning	Yes	Yes	No	No	Some models <sup>[68]</sup>
LRD	ON/OFF distributional function infinite variance	Service time G distributional function infinite variance	Yes	Yes ( $d \neq 0$ )	Yes
SRD	ON/OFF distributional function finite variance	Service time G distributional function finite variance	No	Yes ( $d = 0$ )	Yes
Fractal	Mono-Fractal	Mono-Fractal	Mono-Fractal	Mono-Fractal	MWM (multi-fractal)
Stability	Yes	No (most conditions)	Yes	Yes	Yes
Marginal distribution	Gauss	Gauss	Gauss	Gauss	Lognormal (MWM)
Complexity	$O(n)$ modelling complexity decided by source number	$O(n)$ modelling complexity decided by service number	$O(n)$	$O(n^2)$	$O(n)$

## 5 流量建模新发展

2004年, Karagiannis 等人<sup>[6]</sup>通过分析 Tier 1 ISP 的骨干链路流量发现, 目前高带宽和高聚合的链路流量在极小尺度下近似泊松过程, 从而引发了人们对网络流量特征及建模的新思索和争论. 之所以这样划分, 并不表示近时期的流量模型不具有自相似的特征, 主要是为了更清晰地了解近些年网络流量模型的发展情况.

### 5.1 泊松回归的争论

2004年, Karagiannis 等人<sup>[6]</sup>通过分析 Tier 1 ISP 的骨干链路流量发现, 目前高带宽和高聚合的链路流量在亚秒尺度下近似稳定泊松过程, 而在多秒尺度下流量表现出不稳定性. 这种时间相关的、不稳定的泊松过程, 与在大时间尺度上流量的长相关和自相似特性是一致的.

György 和 Tibor<sup>[66,67]</sup>基于 Karagiannis 等人的工作, 用新的方法重新分析了文献[6]中的流量. 尽管包间隔接近指数分布, 但由于 Kolmogorov-Smirnov 测试反驳了这样的包序列符合 Gamma 分布, 因此相应的流量不可能是泊松过程. 为了找到一个更好的分布来刻画包间隔分布, György 和 Tibor 选用 Lévy 过程族(泊松过程是限定间隔为独立、同质且指数分布的一个简单的 Lévy 过程, 改变这个限定可以得到许多 Lévy 过程来建模包间隔), 通过分析测试发现, 族中 Smoothly Truncated Lévy Flights(STLF)过程用于刻画包间隔分布较为合适. 通过调整 STLF 参数, 使其尽量符合真实流量. 分析调整后的参数显示, 真实流量接近泊松过程, 但仍然符合自相似特征. 这说明, 相比以往的流量, 目前的高带宽和高聚合的链路流量由于叠加了许多不同的流使得突发性明显减弱. 对于相同的流量, 针对时间序列的包长分布, György 和 Tibor 利用自相似模型和线性分形模型的分析验证了相同的情况. 即, 目前高速流量的相关性和突发性在下降.

Pierre 等人<sup>[68]</sup>通过分析 7 年贯穿太平洋主干链路的每天的流量数据发现, LRD 仍然保持强烈、持续和稳定, 聚合流量的边缘分布趋向高斯的演化在时间上保持稳定. 这些都说明, 即使随着实际链路的带宽和负载都增加了, 流量到达也不会是泊松过程, 泊松模型仍然难以刻画互联网络的流量特性. Rahman 等人<sup>[69]</sup>考虑到在目前高带宽网络下用传统的细粒度方法分析流量特性比较困难和费时, 提出了 Fuzzy Group Model, 用 Fuzzy 逻辑大大减少了数学运算的复杂性. 此模型定义了一个阈值, 按阈值把流量分成不同的组, 按每个组的大小把组划分为大、中、小, 按组间的间隔大小把间隔分为大、中、小, 然后按组大小和间隔大小用 Fuzzy 逻辑得出流量的参数  $R$  的值, 用以指示流量的排队特性.  $R \in (0, 1)$ ,  $R$  越大, 表示队列越长. 通过实验分析, 此模型的参数  $R$  比传统模型的参数  $H$  能够更好地反映流量的排队特性.

### 5.2 近年其他模型的发展

#### 5.2.1 流量预测模型

随着智能算法的不断发展, 其良好的非线性映射能力、灵活有效的学习方式在预测领域的应用中表现出较

大的优势和潜力,如神经网络理论、模糊理论、混沌理论等,已应用于通信、交通、气象、水文等多种预测领域.

#### (1) 基于神经网络的模型

人工神经网络(artificial neural network,简称 ANN)是基于生物学中神经网络的基本原理、按照控制工程的思路和数学描述的方法建立起来的数学模型.ANN 预测模型<sup>[70]</sup>是通过采集历史流量数据整理成神经网络的训练集,通过训练确定网络模型,并用该模型估计未来指定时间的流量.网络流量预测问题实质上是一个时序预测问题.就目前利用神经网络来进行预测而言,并没有通用的网络结构和算法.目前来说,神经网络技术还有很多有待于改进的地方,其性能还不十分稳定;而且预测时需要大量的训练样本和迭代,不断修正模型,从而增加了时间和空间复杂性.此外,单、多步长的预测关系、短长范围的预测关系及神经网络自身等还有很多工作有待研究.

#### (2) 混沌理论模型

混沌是一种复杂的运动形式,具有不可长期预报性、不可分解性和稠密的无穷多个周期轨道,如何判定时间序列的混沌特性一直是国内外学者研究的重点.网络业务流量具有混沌性,从混沌时间序列的角度来研究自相似的业务流速率是可行的.最早的模型是 Pruthi 等人提出来的确定性的混沌映射模型<sup>[24-26]</sup>,我国学者在这方面进行了大胆的尝试.陆锦军等人<sup>[71]</sup>在 wolf 算法的基础上提出了一个混沌模型,其基本思想是,基于最大 Lyapunov 指数的预测方法是直接根据数据序列本身所蕴含的规律来进行预测,不需要事先建立主观的分析模型,它具有精度高、可信度高的优点.基于混沌理论的网络流量预测是一个很有意义的尝试,随着对网络流量混沌现象本质的进一步深入研究,预测的可靠性和精度将更高,速度将更快.

#### (3) 模糊理论模型

模糊理论(fuzzy theory)是处理不确定性、非线性等问题的一种有力工具,比较适合于表达那些模糊或定性的知识.由于网络流量是一个非平稳的时间序列,而模糊理论能够对时间序列进行预测,因而可以把模糊理论引入到网络流量的建模和预测的研究中.此外,模糊理论与 ANN 结合组成模糊神经网络在流量预测中也有不少应用.王兆霞等人<sup>[72]</sup>在利用神经网络进行流量建模的过程中,在前端增加了模糊预处理过程,通过分类可以减少神经网络的输入节点,而使得神经网络系统可以处理定量和定性两方面的数据样本,提出了一种基于模糊神经网络的模型;另一方面,通过在 ANN 输出端增加一个模糊组件,可以对 ANN 预测出来的结果加入某些网络环境的定性因素来进行修正.

#### (4) 混合模型

单一的预测模型,如泊松模型、AR 模型、马尔可夫模型、ARMA 模型、小波模型等,由于只刻画了流量的泊松过程或自相似特性而不能很好地对业务流进行预测.采用混合的预测模型可以更加准确、全面地描述和预测实际网络流量.黄晓璐等人<sup>[73]</sup>引入半马尔可夫模型描述网络流量特性,通过忙阈值和闲阈值的设定将网络流量划分为 4 种状态:忙、空闲、上升和下降,研究各种状态下的网络流量特性及各种状态间的相互转换关系.通过网络协议性能分析,在一定的假设条件下推出 IP 网络流量在处于忙状态时服从几何布朗运动,在空闲状态下服从正态分布,在上升或下降状态下服从指数分布.洪飞等人<sup>[74]</sup>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.限于篇幅,其他文献这里不再一一列举.当应用混合模型来预测流量时,可以拟合多个模型的优点,更全面地刻画复杂的流量特性.基于此,可以认为,采用混合的预测模型可以更加准确、全面地描述和预测实际网络流量.目前,这方面的工作仍在探讨与研究中.采用混合模型得到的预测效果,相较于单一模型都具有良好的预测效果.但是,建立混合模型的主要难点在于确定合适的网络业务特性及其适用方法.

### 5.2.2 多分形模型

自相似在相对较大的时间尺度上刻画了流量的长相关特性,但无法描述业务的局部特征;在较小的时间尺度上(sub-RTT),流量体现出更为复杂的变化规律,尤其是局部具有突发性,这种细小尺度下的突发性可以用多分形描述.当  $a(t_0) > 1$  时,流量在  $t_0$  附近是低谷;当  $a(t_0) < 1$  时,流量在  $t_0$  附近是尖峰.对网络流量进行高频细节上的

研究表明<sup>[75]</sup>,此时已无法用单一的 Hurst 指数描述流量在处处发生的突变程度.传统的多重分形时间维度分析方法无助于确定自相似和多重分形在何种尺度出现转变,进而无助于进一步分析多重分形的成因.当时间尺度趋于微小时,流量更多地体现了多重分形特性.

根据 Cascade 模型<sup>[76]</sup>中对于 TCP 协议的特性对流量由粗粒度到细粒度的行为进行了阐述,相关研究证明,出现转变的尺度可能与 TCP/IP 协议从应用层到传输层的工作机制有关.此后,又有一些研究提出自己的看法,认为工作在 RTT 时间粒度上的 TCP 协议,其使用慢启动等拥塞控制机制来调节网络面临压力时的发包速率,是造成细节处的局部突发不规律的主要原因.总的来说,网络协议的分级级联是多重分形的重要原因.除了前述的多重分形小波模型 MWM(multi-fractal wavelet model)<sup>[62]</sup>以外,Gabriel 等人<sup>[77]</sup>用多分形布朗运动(multifractal Brownian motion,简称 mBm)代替分形布朗运动(FBM),从而扩展了 Norros 提出的 FBM 模型<sup>[52]</sup>.扩展后的 mBm 模型既能匹配流量在小尺度下的多分形,又能匹配流量在大时间尺度下的长相关.我国学者王升辉等人<sup>[78]</sup>通过分析树型多重分形结构的相关性发现,多重分形可以把非平稳且具有长相关和分形特性的网络流量序列转化为可用短相关(SRD)模型表示的序列组.利用多重分形这种将时间序列分解为多层的能力,提出了一种结合多重分形的 FIR 神经网络流量预测模型(multifractal FIR network,简称 MF-FIR).MF-FIR 合理地利用了流量序列的 LRD 信息,具有很好的多步预测性能,可以满足通信系统在线预测的要求.

## 6 展 望

从流量模型的发展过程来看,流量模型的发展基本遵循一个“短相关-自相似-多分形”的发展历程,不同的模型所描述的流量特征的本质是不同的.短相关模型描述流量中包的到达是一个泊松过程,刻画了流量的短相关特性,因此适合于小缓冲区的排队系统分析;自相似模型描述流量自相似和长相关特性,刻画流量的长相关特征,因此适合于大缓冲区的排队系统分析;多分形模型更多地用于流量的细节分析,流量预测模型主要用于预测未来到达流量的大小等等.

到目前为止,Internet 网络流量建模已经取得了一定的研究成果,但是与理想和适用的要求还存在很大的差距.总结前面所述的互联网络流量模型可以看到:(1) 度量参数相对单一,目前的网络模型大部分基于流量时间序列的自相似特性及其度量参数  $H$  上;(2) 基于小尺度的流量行为研究和相关的应用尚处于初始阶段,目前主要是小尺度下的多重分形模型的研究;(3) 目前的流量模型侧重于数学模型描述流量特征,而物理模型的发展还远远不够,难点在于如何把负载、网络架构、协议、传输技术和用户行为等相关特征组合起来,建模成一个能够持续、有效地描述网络系统的物理模型;(4) 大部分模型是为了符合某些实际流量的特征,如自相似等,而真正适合某些具体应用的模型的发展还不够,难点在于如何结合具体的应用来选择合适的尺度或用其他方式来建模,而不局限于表征自相似程度的  $H$  参数上;(5) 流量模型的复杂程度与描述流量特征的准确性之间存在一定的矛盾,难点在于取一个很好的折衷来用尽量简单的模型描述流量的特征.

基于流量模型发展上的以上不足,我们从以下几个方面来展望流量模型的发展:

### (1) 新的流量特性及其度量参数的发现

目前的网络模型大部分基于流量时间序列的自相似特性及其度量参数  $H$  上,实际流量中到底存在哪些未知特性仍有待探索;而对于已发现的特性,如自相似和长相关特性、小尺度下的多分形性、混沌和周期特性等来说,研究这些特性的生成机理和相互关系以及如何刻画,并利用这些特性对实际网络性能进行改善等,也都是今后的研究重点.另外,目前的网络流量模型生成的流量均基于流量的时间特性,而没有考虑其空间特性,即生成的流量不含空间信息(如每个包的五元组信息).2008年,Fonseca 等人<sup>[79]</sup>用信息熵的方法发现流量空间上的信息单元存在长程互相关特性,不仅为流量建模开辟了新的方向,也为新的应用领域指引了新的方向.即新模型产生的流量除了和时间序列上符合已有特性外,每个包要与实际流量一样含有空间信息.在空间上信息符合长程互相关特性、含有空间信息的流量模型必然极大地拓展了流量模型的应用领域.

### (2) 网络流量的小尺度行为研究和相关的应用

网络技术日新月异,各种新型的网络业务也在不断涌现,这都使网络行为的复杂性不断提高;并且由于用户

形态的不同,包括无线的接入等问题都已经慢慢出现.因此,网络流量的小尺度行为的研究将是高性能协议设计、网络设备开发、网络设计与规划、网络管理与操作的基础.从目前的情况来看,网络拥塞的原因有很多,我们可以结合网络行为对网络拥塞的成因进行研究,从而修正和设计新的网络协议,并且将网络队列分析和网络实时预测等研究结合到网络的实际控制之中,以提供更优良的网络服务.这也必将是我们今后的研究方向.

### (3) 新的物理模型的发展

目前的物理模型发展还比较滞后,由于物理模型能够很好地解释网络特性的成因,对应于网络的行为,将来应该有更好的、更实用的、能够解释实际网络行为的物理模型被发展起来.而正是这些网络行为,导致了流量的自相似和小时间尺度上的多分形.模型不仅要能够很好地解释动态的网络行为和相应流量特征的关系,还应能解释网络拓扑结构、协议、传输技术和用户行为等流量特征的相关关系.相信随着空间信息自相似特性的发现,必然会有新的、相应的物理模型被发展起来以解释这些特性.一种简单的思路是拓展现有的重尾分布 ON/OFF 模型,使其在 ON 态发包时含有五元组信息,信息在生成时在空间上符合重尾或幂律分布(如 Pareto 分布).这样生成的流量不仅在时间序列上具有自相似性,且其空间的信息结构也具有自相似性.

### (4) 新的应用领域及相关模型的发展

目前,流量模型主要用在 QoS、网络性能管理、准入控制领域,这种局限性与已有的流量模型不含空间信息有关.如果将来的流量模型生成的流量与实际流量一样含有类似特性的空间信息,必将极大地扩展流量模型的应用领域,如在网络设备设计、未来互联网设计、网络运行管理与异常监测、网络应用的优化设计等方面都会取得实际的应用.另外,在应用深度上也必定加强,如排队应用中被用于分析包的丢失率和延迟,却很少用于抖动方面的研究,新的模型必然会弥补这些方面的不足.新的流量建模方法一方面依赖于新的流量特性的发现,另一方面也与应用紧密相关.基于目前的应用和流量特性,勇于创新,用新的思路针对某类具体应用发展更适合其应用、更具适用价值的模型和度量指标,也是一个很好的发展方向.一个很好的例子就是第 5.1 节提到的针对排队的 Fuzzy Group 模型<sup>[69]</sup>,所提出的参数  $R$  比传统模型的参数  $H$  能够更好地反映流量的排队特性.另外,用混合模型弥补单一模型的不足也是很值得研究的方向.

### (5) 模型简单和精确性的发展

网络流量建模还存在很多有待于进一步解决的问题,如,网络流量模型的参数越多,模型就越精确,越接近真实的网络流量源;但同时,模型也越繁琐和复杂,理论分析和计算机仿真也会越困难.然而,理想的流量模型是尽量简单,参数尽量少且容易估算,在实际的网络流量建模中往往要对网络流量模型的精确性和复杂度加以折衷.未来的流量建模领域的发展必定是跨领域、跨学科间的交叉融合,如将 Internet 网络流量与其他网络的流量结合起来研究,提取其共性,并分析其差异,将不同网络的静态特征、动态网络行为、网络流量特征结合起来研究,研究它们之间的普遍特征和相互联系,都是将来很值得深入研究的方向.

## 7 总 结

本文针对互连网络流量建模这一研究领域,总结了现阶段已发现的主要网络流量的特性及相关度量参数,概要介绍了网络流量建模的意义和分类,然后按照“传统-自相似-泊松模型回归引发的争论”这 3 个时段及分类阐述了相关的网络流量模型的发展.本文以一个长的时间轴对流量模型的发展进行了深入的阐述,重点分析了发展过程中较有代表性的模型,分析比较了各种模型的优缺点.最后分析了流量模型发展上的不足和基于这些不足流量模型可能的发展方向.

人们常说“互联网是一个活生生的动物”,相应地,“互联网的流量也是一个活生生的动物”.随着流量特性的发展和发掘,相应的流量模型也随其发展,大致经历了“短相关-自相似-多分形”的发展过程.然而这个过程不是一成不变的,也远不是终点.随着流量特征的变化和新特征地发现,新的流量模型必将继续涌现.

**致谢** 感谢安常青、王会、陈福老师详细和中肯的修改意见;感谢张威博士关于小波理论的讨论;感谢实验室其他同学在写作过程中给予的无私支持和帮助.

**References:**

- [1] Leland WE, Taqqu MS, Willinger W, Wilson DV. On the self-similar nature of Ethernet traffic. *ACM SIGCOMM Computer Communication Review*, 1993,23(4):183–193. [doi: 10.1145/167954.166255]
- [2] Leland WE, Taqqu MS, Willinger W, Wilson DV. On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans. on Networking*, 1994,2(1):1–15. [doi: 10.1109/90.282603]
- [3] Paxson V, Floyd S. Wide-Area traffic: The failure of Poisson modeling. *IEEE/ACM Trans. on Networking*, 1995,1(3):226–244.
- [4] Klivansky S, Mukherjee A, Song C. On long-range dependence in NSFNET traffic. Technical Report, GIT-CC-94-61, Atlanta: Georgia Institute of Technology, 1994.
- [5] Crovella ME, Bestavros A. Self-Similarity in World Wide Web traffic: Evidence and possible causes. *IEEE/ACM Trans. on Networking*, 1997,5(6):835–846. [doi: 10.1109/90.650143]
- [6] Karagiannis T, Molle M, Faloutsos M, Broido A. A nonstationary Poisson view of Internet traffic. In: *Proc. of the 23rd Annual Joint Conf. of the IEEE Computer and Communications Societies*. Hong Kong: IEEE, 2004. 1558–1569. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1354569](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1354569)
- [7] Liang Q. Ad hoc wireless network traffic-self-similarity and forecasting. *IEEE Communications Letters*, 2002,6(7):297–299. [doi: 10.1109/LCOMM.2002.801327]
- [8] Ilow J, Leung H. Self-Similar texture modeling using FARIMA processes with applications to satellite images. *IEEE Trans. on Image Processing*, 2001,10(5):792–797. [doi: 10.1109/83.918571]
- [9] Mark C, Balachander K. *Internet Measurement: Infrastructure, Traffic and Applications*. Chichester: John Wiley & Sons Ltd., 2006.
- [10] Yang JH, Wu JP, An CQ. *The Theory and Applications of Internet Measurement*. Beijing: The People's Posts & Telecom Press, 2009 (in Chinese).
- [11] Sadrer KR, Haverkort BR. A validation of the pseudo self-similar traffic model. In: *Proc. of the Int'l Conf. on Dependable Systems and Networks*. Bethesda: IEEE, 2002. 727–734. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1029018](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1029018)
- [12] Taqqu M, Teverovsky V, Willinger W. Estimators for long-range dependence: An empirical study. *Fractals*, 1995,3(4):785–798. [doi: 10.1142/S0218348X95000692]
- [13] Teverovsky V, Taqqu M. Testing for long-range dependence in the presence of shifting means or a slowly declining. *Journal of Time Series Analysis*, 2001,18(3):279–304. [doi: 10.1111/1467-9892.00050]
- [14] Higuchi T. Approach to an irregular time series on the basis of the fractal theory. *Physica D*, 1988,31:277–283. [doi: 10.1016/0167-2789(88)90081-4]
- [15] Peng CK, Buldyrev SV, Havlin S, Simons M. Mosaic organization of DNA nucleotides. *Physical Review E*, 1994,49(2): 1685–1689.
- [16] Pacheco JCR. Behavior of R/S statistic implementations under time-domain operations. In: *Proc. of the 3rd Int'l Conf. on Electrical and Electronics Engineering*. Veracruz: IEEE, 2006. 1–4. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=4018020](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4018020)
- [17] Julio C, Pacheco R, Roman DT. A tool for long-range dependent analysis via the R/S statistic. In: *Proc. of the 15th Int'l Conf. on Computing*. Mexico City: IEEE, 2006. 361–366. <http://www.computer.org/portal/web/csdl/doi/10.1109/CIC.2006.16>
- [18] Julio C, Pacheco R, Roman DT, Vargas LE. R/S statistic: Accuracy and implementations electronics. In: *Proc. of the 18th Int'l Conf. on Communications and Computers*. Puebla: IEEE, 2008. 17–22. <http://www.computer.org/portal/web/csdl/doi?doc=doi/10.1109/CONIELECOMP.2008.14>
- [19] Beran J. *Statistics for Long-Memory Processes*. New York: Chapman & Hall/CRC Press, 1994.
- [20] Fox R, Taqqu MS. Large-Sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 1986,14(2):517–532. [doi: 10.1214/aos/1176349936]
- [21] Wornell GW, Oppenheim AV. Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. on signal processing*, 1992,40(3):611–623. [doi: 10.1109/78.120804]
- [22] Riedi RH, Vehel JL. Multifractal properties of TCP traffic: A numerical study. Technical Report, RR-3129, INRIA, Rocquencourt, 1997.
- [23] Ikoma N, Shimizu T, Imazu K, Yana K. Decomposition of network traffic of dialup access into trend and periodic components based on nonstationary Bayesian model. In: *Proc. of the SICE Annual Conf*. Fukui: IEEE, 2003. 2830–2833.

- [24] Erramilli A, Singh RP, Pruthi P. Chaotic maps as models of packet traffic. In: Proc. of the 14th ITC. 1994. [http://thyroid.org/ann\\_mtg/ITC2010/index.html](http://thyroid.org/ann_mtg/ITC2010/index.html)
- [25] Erramilli A, Singh RP, Pruthi P. Modeling packet traffic with chaotic maps. In: Proc. of the ICC. 1995. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.46.8102>
- [26] Erramilli A, Singh RP. An application of deterministic chaotic maps to model packet traffic. *Queueing Systems*, 1995,20(1,2): 171–206. [doi: 10.1007/BF01158436]
- [27] Adas A. Traffic models in broadband networks. *IEEE Communications Magazine*, 1997,35(7):82–89.
- [28] Hassan H, Garcia JM, Bockstal C. Modeling Internet traffic: Performance limits. In: Proc. of the Int'l Conf. on Internet Surveillance and Protection. 2006. 7. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=1690391](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1690391)
- [29] Bonald T. The Erlang model with non-poisson call arrivals. *ACM SIGMETRICS Performance Evaluation Review*, 2006,34(1): 276–286. [doi: 10.1145/1140103.1140309]
- [30] Grassmann WK, Luo JX. Simulating Markov-reward processes with rare events. *ACM Trans. on Modeling and Computer Simulation*, 2005,15(2):138–154. [doi: 10.1145/1060576.1060578]
- [31] Heffes H, Lucantoni DM. A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communication*, 1986,4(6):856–868. [doi: 10.1109/JSAC.1986.1146393]
- [32] Yang XS, Petropulu AP. The extended alternating fractal renewal process for modeling traffic in high-speed communication networks. *IEEE Trans. on Signal Processing*, 2001,49(7):1349–1363. [doi: 10.1109/78.928689]
- [33] Shah-Heydari S, Le-Ngoc T. MMPP modeling of aggregated ATM traffic. *IEEE Canadian Conf.*, 1998,1(24-28):129–132.
- [34] Maglaris B, Anastassiou D, Sen P, Karlsson G, Robbins JD. Performance models of statistical multiplexing in packet video communications. *IEEE Trans. on Communications*, 1988,36(7):834–844. [doi: 10.1109/26.2812]
- [35] Sen P, Maglaris B, Rikli NE, Anastassiou D. Models for packet switching of variable-bit-rate video sources. *IEEE Journal of Selected Areas in Communications*, 1989,7(5):865–869. [doi: <http://dx.doi.org/10.1109/49.32350>]
- [36] Shim C, Ryoo I, Lee J, Lee S. Modeling and call admission control algorithm of variable bit rate video in ATM networks. *IEEE Journal of Selected Areas in Communications*, 1994,12(2):332–344. [doi: 10.1109/49.272884]
- [37] Cohen D, Heyman D. Performance modeling of video teleconferencing in ATM networks. *IEEE Trans. on Circuits and System for Video Technology*, 1993,3(6):408–20. [doi: 10.1109/76.260197]
- [38] Heyman D, Lakshman TV, Tabatabai A, Heeke H. Modeling teleconference traffic from VBR video coders. In: Proc. of the ICC. IEEE, 1994. 1744–1748.
- [39] Heyman D, Tabatabai A, Lakshman T. Statistical analysis and simulation study of video teleconference traffic in ATM networks. *IEEE Trans. on Circuits and System for Video Technology*, 1992,2(1):49–59. [doi: 10.1109/76.134371]
- [40] Bagnall AJ, Janacek GJ. Clustering time series from ARMA models with clipped data. In: Kim W, Kohavi R, eds. Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data. New York: ACM, 2004. 49–58. <http://portal.acm.org/citation.cfm?id=1014061>
- [41] Grunenfelder R, Cosmas JP, Manthorpe S, Odinma-Okafor A. Characterization of video codecs as autoregressive moving average processes and related queuing system performance. *IEEE Journal of Selected Areas in Communications*, 1991,9(3):284–93. [doi: 10.1109/49.76626]
- [42] Kohn R, Ansley CF. Estimation, prediction, and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, 1986,81(395):751–761. [doi: 10.2307/2289007]
- [43] Frost V, Melamed B. Traffic modeling for telecommunications networks. *IEEE Communications Magazine*, 1994,32(3):70–81.
- [44] Melamed B, Hill JR, Goldsman D. The TES methodology: Modeling empirical stationary time series. In: Swain JJ, Goldsman D, Crain RC, WiLson JR, eds. Proc. of the 24th Conf. on Winter Simulation. Arlington: ACM, 1992. 135–144.
- [45] Melamed B, Hill JR. Applications of the TES modeling methodology. In: Evans GW, Mollaghasemi M, Russell EC, Biles WE, eds. Proc. of the 25th Conf. on Winter Simulation. Los Angeles: ACM, 1993. 1330–1338.
- [46] Willinger W, Taqqu MS, Sherman R, Wilson DV. Self-Similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *ACM SIGCOMM Computer Communication Review*, 1995,25(4):100–113. [doi: 10.1145/217391.217418]

- [47] Mandelbrot B. Long-Run linearity, locally gaussian processes,  $H$ -spectra and infinite variance. *Int'l Economic Review*, 1969,10(1): 82–111.
- [48] Willinger W, Taqu MS, Sherman R, Wilson DV. Self-Similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level. *IEEE/ACM Trans. on Networking*, 1997,5(1):71–86. [doi: 10.1109/90.554723]
- [49] Sarvotham S, Riedi R, Baraniuk R. Network and user driven alpha-beta on-off source model for network traffic. *Computer Networks*, 2005,48(3):335–350. [doi: 10.1016/j.comnet.2004.11.024]
- [50] Krunz M, Makowski M. Modeling video traffic using  $M/G/\infty$  input processes. *IEEE Journal of Selected Areas in Communications*, 1998,16(5):733–748. [doi: 10.1109/49.700909]
- [51] Mandelbrot B. Self-Similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Trans. on Communication Technology*, 1965,13(1):71–90. [doi: 10.1109/TCOM.1965.1089090]
- [52] Norros I. A storage model with self-similar input. *Queueing Systems*, 1994,16(3,4):387–396. [doi: 10.1007/BF01158964]
- [53] Lau WC, Erramilli A, Wang JL. Self-Similar traffic generation: The random midpoint displacement algorithm and its properties. In: *Proc. of the IEEE ICC, Vol.1*. Seattle: IEEE, 1995. 466–472.
- [54] Paxson V. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review*, 1997,27(5):5–18.
- [55] Flandrin P. Wavelet analysis and synthesis of fractional brownian motion. *IEEE Trans. on Information Theory*, 1992,38(2): 910–917. [doi: 10.1109/18.119751]
- [56] Ledesma S, Liu D. Synthesis of fractional Gaussian noise using linear approximation for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review*, 2000,30(2):4–17
- [57] Beran J, Sherman R, Taqu MS, Willinger W. Long-Range dependence in variable-bit-rate video traffic. *IEEE Trans. on Communications*, 1995,43(2-4):1566–1579. [doi: 10.1109/26.380206]
- [58] Huang CC, Devetsikiotis M, Lambadaris I, Roger-Kaye A. Modeling and simulation of self-similar variable bit rate compressed video: A unified approach. *ACM SIGCOMM Computer Communication Review*, 1995,25(4):114–125. [doi: 10.1145/217391.217420]
- [59] Flandrin P. Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. on Information Theory*, 1992,38(2): 910–917. [doi: 10.1109/18.119751]
- [60] Ma S, Ji C. Modeling heterogeneous network traffic in wavelet domain. *IEEE/ACM Trans. on Networking*, 2001,9(5):634–649. [doi: 10.1109/90.958331]
- [61] Ma S, Ji C. Modeling video traffic in the wavelet domain. In: *Proc. of the 17th Annual IEEE Conf. on Computer Communication, INFOCOM*. San Francisco: IEEE, 1998. 201–208. [http://ieeexplore.ieee.org/xpl/freeabs\\_all.jsp?arnumber=659655](http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=659655)
- [62] Riedi RH, Crouse MS, Rbeiro VJ. A multifractal wavelet model with application to network traffic. *IEEE Trans. on Information Theory*, 1999,45(3):992–1018. [doi: 10.1109/18.761337]
- [63] Grossglauser M, Bolot JC. On the relevance of long-range dependence in network traffic. *ACM SIGCOMM Computer Communication Review*, 1996,26(4):15–24. [doi: 10.1145/248157.248159]
- [64] Barakat C, Thiran P, Iannaccone G, Diot C, Owezarski P. A flow-based model for Internet backbone traffic. In: *Proc of the 2nd ACM SIGCOMM Workshop on Internet Measurement*. New York: ACM, 2002. 35–47. <http://portal.acm.org/citation.cfm?id=637206>
- [65] Gallardo JR, Makrakis D, Orozco-Barbosa L. Use of  $\alpha$ -stable self-similar stochastic processes for modeling traffic in broadband networks. *Performance Evaluation*, 2000,40(1-3):71–98. [doi: 10.1016/S0166-5316(99)00070-X]
- [66] Terdik G, Gyires T. Internet traffic modeling with Lévy flights. In: *Proc of the 7th Int'l Conf. on Networking*. Singapore: IEEE, 2008. 468–473.
- [67] Terdik G, Gyires T. Lévy flights and fractal modeling of Internet traffic. *IEEE/ACM Trans. on Networking*, 2009,17(1):120–129. [doi: 10.1109/TNET.2008.925630]
- [68] Borgnat P, Dewaele G, Fukuda K, Abry P, Cho K. Seven years and one day: Sketching the evolution of Internet traffic. In: *Proc. of the 28th Conf. on Computer Communications (INFOCOM 2009)*. Rio de Janeiro: IEEE, 2009. 711–719.
- [69] Rahman A, Kennedy P, Simmonds A, Edwards J. Fuzzy logic based modeling and analysis of network traffic. In: *Proc of the 8th IEEE Int'l Conf. on Computer and Information Technology*. Sydney: IEEE, 2008. 652–657.



- [70] Liu J, Huang YL. Nonlinear network traffic prediction based on BP neural network. *Journal of Computer Applications*, 2007,27(7): 1770–1772 (in Chinese with English abstract).
- [71] Lu JJ, Wang ZQ. Prediction of network traffic flow based on chaos characteristics. *Journal of Nanjing University of Aeronautics & Astronautics*, 2006,38(2):217–221 (in Chinese with English abstract).
- [72] Wang ZX, Sun YG, Chen ZQ, Yuan ZZ. Study of predicting network traffic using fuzzy neural networks. *Journal on Communications*, 2005,26(3):136–140 (in Chinese with English abstract).
- [73] Huang XL, Min YY, Wu Q. A semi-Markov model for network traffic. *Chinese Journal of Computers*, 2005,28(10):1592–1600 (in Chinese with English abstract).
- [74] Hong F, Wu ZM. Multiscale network traffic prediction model based on wavelet. *Chinese Journal of Computers*, 2006,29(1): 166–170 (in Chinese with English abstract).
- [75] Gilbert AC. Multiscale analysis and data networks. *Applied and Computational Harmonic Analysis*, 2001,10(3):185–202. [doi: 10.1006/acha.2000.0342]
- [76] Feldmann A, Gilbert AC, Willinger W. Data networks as cascades: Investigating the multifractal nature of Internet WAN traffic. *ACM SIGCOMM Computer Communication Review*, 1998,28(4):42–55.
- [77] Bianchi GR, Vieira FHT, Lee LL. A novel network traffic predictor based on multifractal traffic characteristic. In: *Proc. of the Global Telecommunications Conf. (GLOBECOM 2004)*, Vol.2. Dallas Texas: IEEE, 2004. 680–684. <http://www.ieee-globecom.org/2004/committees.html>
- [78] Wang SH, Qiu ZD. Network traffic nonlinear prediction combined with mutifractal. *Journal of Communications*, 2007,28(2):45–50 (in Chinese with English abstract).
- [79] Fonseca N, Crovella M, Salamatian K. Long range mutual information. *ACM SIGMETRICS Performance Evaluation Review*, 2008, 36(2):32–37. [doi: 10.1145/1453175.1453181]

#### 附中文参考文献:

- [10] 杨家海,吴建平,安常青.网络测量理论与应用.北京:人民邮电出版社,2009.
- [70] 刘杰,黄亚楼.基于BP神经网络的非线性网络流量预测.计算机应用,2007,27(7):1770–1772.
- [71] 陆锦军,王执铨.基于混沌特性的网络流量预测.南京航空航天大学学报,2006,38(2):217–221.
- [72] 王兆霞,孙雨耕,陈增强,袁著祉.基于模糊神经网络的网络业务量预测研究.通信学报,2005,26(3):136–140.
- [73] 黄晓璐,闵应骅,吴起.网络流量的半马尔可夫模型.计算机学报,2005,28(10):1592–1600.
- [74] 洪飞,吴志美.基于小波的多尺度网络流量预测模型.计算机学报,2006,29(1):166–170.
- [78] 王升辉,裘正定.结合多重分形的网络流量非线性预测.通信学报,2007,28(2):45–50.



张宾(1976—),男,河南新乡人,博士生,主要研究领域为网络流量特征及建模,网络测量,异常检测.



吴建平(1953—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为计算机网络体系结构,网络协议测试.



杨家海(1966—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为计算机网络,网络管理与测量,协议工程学.