

Internet 城域出口链路流量测量与特征分析

张广兴^{1,2}, 张大方³, 谢高岗², 杨建华², 张敦行^{1,3}

(1. 湖南大学计算机与通信学院, 湖南长沙 410082; 2. 中国科学院计算技术研究所
下一代互联网研究中心, 北京 100080; 3. 湖南大学软件学院, 湖南长沙 410082)

摘要: 通过对某城域出口链路上实测的流量数据从出/入境方向、字节/数据包等不同角度进行了逐层逐协议的分析, 同时采用了多种 Hurst 指数估计方法考察了流量的自相似特征. 发现相比先前的监测结果, 流量特征在构成和趋势等方面都发生了显著的变化; 并且, 网络流量的自相似程度甚至自相似特征是否依然存在都变得不再确定.

关键词: 城域网; 流量测量; 流量特征; 自相似; Hurst 指数

中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112 (2007) 11-2092-06

Internet Traffic Measurement and Characteristic Analysis on Output Link of Metro Area Network

ZHANG Guang-xing^{1,2}, ZHANG Da-fang³, XIE Gao-gang², YANG Jiang-hua², ZHANG Dun-xing^{1,3}

(1. College of Computer and Communication, Hunan University, Changsha, Hunan 410080, China;
2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China;
3. Software School, Hunan University, Changsha, Hunan 410080, China)

Abstract: The nature of the increasing network traffic has taken much more changes. In order to find out how these changes impact on the traffic character, we make a comprehensive analysis by using the traffic data collected with NetTurbo on an ISP WAN link. The result shows that compared to the previous analysis there are so many un-neglectable differences in the traffic characterization. More significantly, we found that despite its wide use, there is still much confusion regarding the identification of self similarity in recent real network traffic data and whether the self similarity still exists is uncertain, which is much more difference from the previous result.

Key words: WAN; traffic measurement; traffic character; self similarity; Hurst parameter

1 引言

网络流量特征是协议设计、设备开发、流量工程与网络性能提升的基础. 自 1994 年 W. Leland 等人首次发现了网络流量具有自相似性特征后, 国内外大量的科研力量投入到自相似和与自相似密切相关的长相关、重尾分布、分形和多重分形等网络流量特征的分析和研究中, 取得了一系列重要的发现和成果^[1-6]. Hurst 指数是衡量自相似程度的重要参数^[1]. Kihong Park、Mark E. Crovella、Ericsson 等分别论证了被传输文件大小的重尾分布、Web 文件的大小分布特征、TCP 拥塞控制和具有重尾特性的上层协议的共同作用等因素是造成流量自相似原因^[4-6]; Thomas Karagiannis 总结了自相似和 Hurst 指数的估计方法, 并通过对仿真数据的分析指出不同的

估计方法会因受到隐藏在流量序列中的周期性、趋势性和噪声性等的影响而存在结果上的偏差^[7].

现有的对网络流量自相似特征的研究主要集中在对 HTTP 和 FTP 等传统业务的分析, 并且所关注流量的构成是以 TCP 为主体的. 然而, 近年来网络规模与带宽不断扩大, 基于 IP 的数据、音/视频, 尤其是以 P2P 为基础的实时流媒体、文件共享等大量新业务层出不穷, 网络流量呈现复杂的非线性增长趋势, 其构成等相比先前有了显著的不同^[8-11]. 为了考察这些变化对网络流量特征造成的影响以及网络流量的自相似程度是否有所有改变(甚至自相似是否仍旧存在), 本文通过 NetTurbo 长期监测运营商城域网出口链路所得的流量数据, 区分出/入境方向、字节/数据包进行了逐层逐协议的流量特征分析, 同时采用了多种 Hurst 指数估计方法考察了流

量的自相似程度.分析发现网络的流量特征发生了很大的变化,并且这些变化使得网络流量的自相似程度甚至自相似是否仍旧存在都变得并非如先前一样确定.

2 相关知识

自相似(self-similar)概念最早由 Mandelbrot 在上个世纪 70 年代提出,其描述了随机过程在时间或空间尺度上进行缩放时某些特征的不变性.具有恒定均值和方差的二阶平稳随机过程 $X = (X_1, X_2, X_3, \dots)$,若在分布上满足 $X(at) = a^H X(t)$ (其中, $a > 0$ 且被称作尺度因子, H 叫做自相似指数,即 Hurst 指数),则称其在统计意义上是自相似的^[8].Hurst 指数描述了自相似特征,用来衡量过程的自相似程度.对于自相似过程有 $0.5 < H < 1$, H 的值越大则表明过程的自相似程度越高.而自相似程度越高表明网络流量的突发程度越高^[1~3,7];特别的,在排队系统中 $H = 0.75$ 是一个突变点,在 $H > 0.75$ 后排队性能会急剧下降^[2].Hurst 指数的估计结果直接影响到网络流量建模的准确程度和高速网络传输中流量控制和统计复用的性能.

当前存在有多种 Hurst 指数估计方法,每种方法都有一定的适用条件和范围^[7,12~15].而在应用多种分析方法得出的估计结果中选取相对较为集中且接近的 Hurst 指数,可以相对准确地近似表示流量的自相似程度^[14~16].因此,本文应用 S 语言在开源软件 R^[22] 中实现并综合应用到了涵盖了时域、频域和小波分析的多种估计方法: M. S. Taqqu、V. Teverovsky 和 W. Willinger 给出的聚和方差估计法(AggVar)、分形维度估计法(Higuchi)、残差估计法(Peng)、R/S 估计法、周期图估计法(Per)、Whittle 近似极大似然估计法(Whittle)等^[15],以及 Patrice Abry 和 Darryl Veitch 提出的基于小波分析的估计方法(Wavelet)^[12],等等.

3 测量实验与流量特征分析

我们在国内某运营商的城域网出口处部署了自主研发的分布式宽带网络监测分析系统 NetTurbo. NetTurbo 是一套完整的可提供 2-7 层网络性能测试与流量分析的分布式系统.它由支持多种速率及接口的分布式探针 NetProb 和集中式分析子系统 NetCom 组成. NetTurbo 在被监测链路上的部署情况如图 1 所示,其部署在某城域网的出口链路上,通过分光的形式 7*24 小时的监测和分析被监测链路上的流量和性能信息,位于网管中心 NetCom 周期性地各探针上得到的监测和分析结果采集过来保存分析.监测时间从 2005 年 07 月 14 日到 2005 年 09 月 23 日,本文选取了自 2005 年 08 月 28 日 0 时至 2005 年 9 月 03 日 24 时完整一周的流量数据

进行分析.为了描述方便,用 LF05 表示所选取的流量数据集,In/Out 表示入/出境流量、All 表示出入境流量总和;Ratio 表示特定协议或应用的 In、Out 和 All 分别占特定协议或应用的 All 的百分比;Percent 表示特定协议或应用的 In、Out 和 All 分别占 SUM 所对应 In、Out 和 All 的百分比.

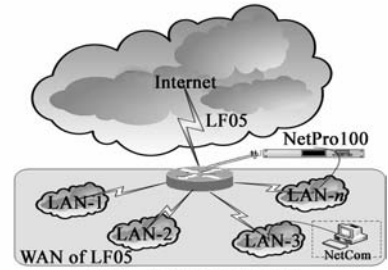


图 1 监测系统部署图

表 1 字节流量信息摘要

字节流量比例信息			
Bytes	In/Ratio/Percent	Out/Ratio/Percent	All/Ratio/Percent
TCP	39,144/78.3%/80.10%	9,848/19.7%/73.46%	48,992/100%/78.98%
UDP	9,685/73.16%/19.82%	3,533/26.84%/26.35%	13,238/100%/20.91%
ICMP	28/54.3%/0.06%	23/45.1%/0.17%	51/100%/0.08%
ESP	14/87.5%/0.03%	2/12.5%/0.01%	16/100%/0.03%
SUM	48,829/77.12%/100%	13,402/21.18%/100%	62,231/100%/100%
Unit	GB/Null/Null	GB/Null/Null	GB/Null/Null

表 2 数据包大小信息摘要

Packet Size Summary Description			
Size	In/Mean/Max/Med/Min	Out Mean/Max/Med/Min	All Mean/Max/Med/Min
TCP	705/850/713/471	390/603/400/182	602/689/600/423
UDP	350/454/349/210	279/441/267/189	328/428/318/228
ICMP	88/109/88/75	112/813/85/74	100/424/87/75
ESP	268/1311/222/122	393/1459/212/98	282/1308/240/123
SUM	587/704/594/443	353/520/363/187	511/605/509/395
Unit	Bytes/Bytes/Bytes/Bytes	Bytes/Bytes/Bytes/Bytes	Bytes/Bytes/Bytes/Bytes

注:包大小不包括数据链路头(一般为 14Bytes)和 IP 头(一般为 20Bytes)的长度

在 LF05 中,总计一周内 82,895M 个数据包产生了 48,829GB 的流量.链路平均比特速率为 824Mbps,其中入境方向为 646Mbps,出境方向为 178Mbps;平均包速率为 209,103pps,其中入境方向为 141,348pps,出境方向为 67,755pps;平均数据包大小为 511Bytes,其中入境方向为 587Bytes,出境方向为 353Bytes.从表 1 可以看出,字节分布上,在传输层 TCP 仍旧是流量的主体,约占到了 78.89%,UDP 约占到了 20.91%.其他的如 ICMP、ESP 等传输层协议总计所占百分比不足 0.50%.数据包分布与字节分布类似.相比文[8~10,14,17]等早期的城域网流量监测结果,UDP 流量有较大的增长.如文[17]指出约 95%字节和 90%的数据包由 TCP 构成,而 UDP 仅占到了总字节的 5%、总数据包的 10%.同时,LF05 中字节流量与数据包流量的变化在出入境方向上是非对称的.入境方向平均链路利用率约为 64.65%,出境方向仅约为 17.76%,两者的比约为 3.64:1(包速率比约为 2.09:1).相比文[9,10,14,17]等,LF05 的平均数据

包大小几乎没有变化.然而根据表2可发现,对于不同方向上的不同传输层协议,平均数据包大小却存在较大的差异性.例如,TCP和UDP皆是入境方向的要大于出境方向的(其中TCP入/出境方向的比例约为1.80:1,UDP约为1.25:1),而ICMP和ESP则相反.并且,尽管总的来看TCP的平均数据包大小是大于UDP的,但在不同方向上表现并非完全相同,其中入境方向上TCP的约为UDP的2倍,而出境方向上TCP的仅约为UDP的1.39倍.

图2给出的是在一周内LF05链路上的TCP、UDP以及传输层协议总和在出境方向、入境方向和出入境方向总和上的数据包速率变化曲线.图3则是给出了完整一天内TCP和UDP的出/入境数据包速率变化曲线.从图2可看出,不论是对TCP、UDP,还是传输层协议总

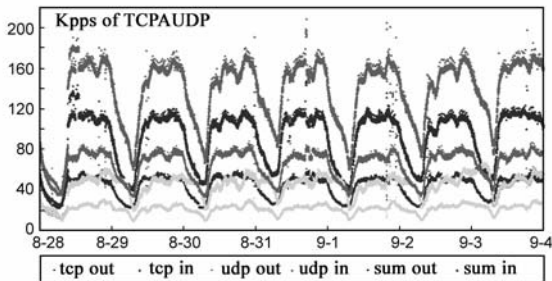


图2 In/Out TCP/UDP/SUM的数据包速率变化

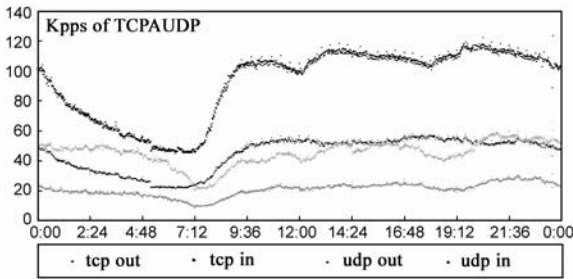
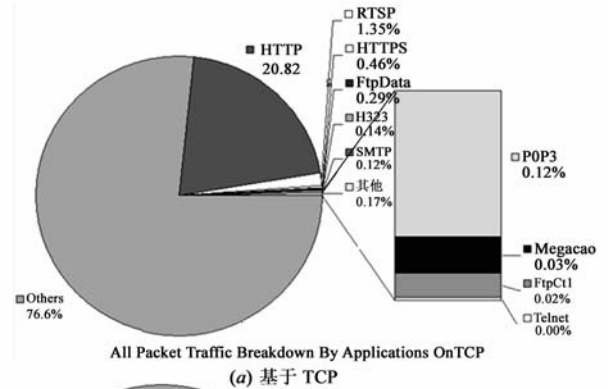


图3 In/Out TCP/UDP在2005-08-30的数据包速率变化

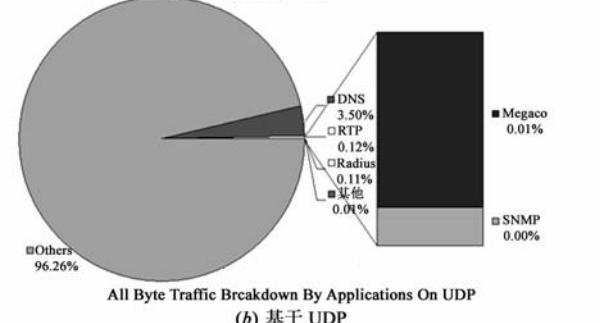
和,在不同方向上,尽管流量速率的变化趋势都大致相同,但入境方向的绝对值要远大于出境方向的;流量速率都呈现出明显的“天特征”变化趋势;TCP和UDP的数据包大小亦存在“天特征”的变化趋势.由图3可知,TCP与UDP的数据包速率都是在每天的0点开始相对缓慢地下降,至凌晨7点左右达最低点(称这段时间为流量相对下降区域);而后开始迅速上升,在凌晨9点左右,到达一个相对稳定的值(称这段时间为流量相对上升区域);接着其围绕这一值随时间推进进行幅度较小的波动,且这种波动一直延续到24点,即第二天的0点(称这段时间为流量相对稳定区域).值得注意的是,流量相对稳定区域内,在12点和18点这两个时间点附近分别存在一处较为明显的流量降至相对最低点的情况.与文[9,10,14,17]等给出的监测结果明显不同的

是,流量相对稳定区域拉长了(长度约为15小时).而如在文[17]给出的流量变化曲线中,流量相对稳定区域仅存在于12点到18点左右的长度约为6个小时的时间段内,其他时间段内流量则相对明显地表现为上升区域或下降区域.

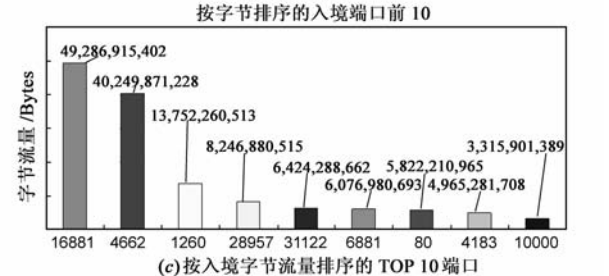
进一步分析基于TCP和UDP的应用层流量构成,如图4(a)和(b)所示,尽管在由传统应用(如基于TCP的HTTP、FTP和Telnet以及基于UDP的DNS等)所产生的流量占有一定的比例,但他们已经不再是应用层流量的主体.例如,在TCP上层有约76.67%的数据包是由非传统应用产生;而在UDP上层有约96.26%的字节是由非传统应用产生.这同样与文[9,10,14,17]等中给出的监测结果不相同.如在文[17]中指出,由HTTP、FTP、SMTP和DNS等传统应用所产生的流量分别约占到总字节流量的75%、5%、5%和1%,总数据包流量的70%、3%、5%和3%,且其他由不可识别应用所产生的流量也怀疑是和HTTP相关的;在文[10]中,尽管有约27%的流量是由非传统应用产生,但HTTP和FTP等传统应用所产生的流量仍保持在了约70%左右.简言之,



(a) 基于TCP



(b) 基于UDP



(c) 按入境字节流量排序的TOP10端口

图4 基于TCP和UDP的应用层流量构成

在早期应用层流量构成中,传统应用占据了主体.而本文给出的监测结果和文[11]等其他新近的一些研究中都发现,由传统应用所产生流量的比例日益减少,取而代之的是大量的 P2P 文件共享等非传统应用.如图 4(c)给出了在某段时间内的 TOP 10 端口排名情况,其中 80 为 HTTP 默认端口、16881 和 6881 为 BT 默认端口、4662 为 eMule 默认端口,BT 和 eMule 皆为目前在国内流行的 P2P 文件共享应用.

表 3 TCP 流量的相关系数

TCP Correlation Coefficient List						
Entry	Corr(bps, pps)			Corr(in, out)		
	In	Out	All	bps	pps	size
TCP	0.9903	0.4766	0.9864	0.2881	0.898	-0.5216
Other	0.9825	0.4276	0.9636	0.2303	0.8768	-0.015
HTTP	0.9552	0.8327	0.9604	0.8141	0.8285	0.0877
RTSP	0.9784	0.999	0.9629	0.671	0.6816	0.0241
HTTPS	0.9447	0.805	0.9404	0.6269	0.895	0.0702
SMTP	0.5951	0.7261	0.7077	0.3231	0.835	0.1202
FtpData	0.9167	0.7266	0.9244	-0.0775	0.1219	-0.3051
FtpCil	0.7641	0.917	0.8318	0.4079	0.5751	0.0288
H323	0.8927	0.5773	0.8905	0.1531	0.7353	-0.2196
POP3	0.9671	0.5823	0.9309	0.2701	0.6201	-0.0964
Megacao	0.8551	0.7184	0.8613	0.0895	0.4781	N/A
Telnet	0.7374	0.5576	0.7459	0.1163	0.6464	N/A
Unit	Null	Null	Null	Null	Null	Null

对 LF05 进行相关性分析可以发现:在不同方向上,字节速率和包速率的相关性是不同的,入境方向要远强于出境方向.如在表 3 中,入境方向上 TCP 的为 0.99,UDP 的为 0.95;而在出境方向上,他们分别仅为 0.48 和 0.43.特别的,对于 HTTP 等传统应用,在出境和入境方向上都表现出字节速率和包速率的强相关性;而对于非传统应用(Others),仅在入境方向上存在较强的相关性,在出境方向上则很弱.此外,出和入境方向上 TCP 数据包速率存在有较强的相关性,字节速率则表现很弱;而对于 UDP 而言,则是字节速率的相关性要大于数据包速率的;传统应用的流量主体 HTTP(s),其出和入境方向的字节速率和数据包速率都表现出较强的相关性;而对于大量的非传统应用,则仅是在数据包速率上表现出较强的相关性.另外,出和入境方向的数据包大小的相关性也是非常弱的.

4 Hurst 指数分析

接下来,应用上文提到的估计方法从多个角度对 LF05 进行自相似程度估计与分析,并与早期的分析结果进行比较.表 4 中列出的对典型流量数据(文[1]所给出,已被证明具有自相似特征,用 Leland 表示)估计得出的 Hurst 指数与文[1]中给出的估计结果基本相同,表明综合应用上文给出的多种 Hurst 指数估计方法可

以正确进行 Hurst 指数估计.LF05 相比早期的分析结果^[1,14],不论是对字节流量还是数据包流量,大多数估计方法得出的 Hurst 指数都明显的要大了很多,且大部分接近甚至超过了 1.而根据自相似的定义,Hurst 指数是应当小于 1 的,当其大于 1 时,自相似是否存在是不确定的.值得注意的是,观察表 4 中给出的结果,需要事先假定流量序列为 fGn 或 $fARIMA$ 过程的 Whittle 近似极大似然估计法(其对于早期的流量数据可进行合理的估计^[1~5,14])并不适用于 LF05.对于 Whittle 方法而言, $method = fgn$ 时若估计得出的 Hurst 指数为 0.999 则表示其不可用;而 $method = farma$ 时若估计得出的 Hurst 指数大于 1 或者相对非常的小,则表明被分析序列不是自相似的^[16].这也表明:尽管对于早期的监测结果, fGn 或 $fARIMA$ 过程可在一定程度上对具有自相似特征的网络流量进行刻画和描述,但对于新近的 LF05 链路上监测所得的流量数据,其不再适用.这恰恰也正是上文中所提到的构成、变化趋势以及相关性等流量特征都发生了显著变化而所产生的一个不可忽视的结果.

表 4 LF05 和 Leland 流量数据的 Hurst 指数估计

The Estimation of the Hurst Parameter of the LangFang&Leland Traffic Info					
Method	LangFang		Leland		Notes
	Byte	Pkt	Byte	Pkt	
AggVar	0.9820	0.9849	0.8169	0.8486	
DiffAggVar	1.3907	1.4594	0.9001	0.8624	
AggAbsVal	0.9830	0.9879	0.8043	0.8422	Moment - 1
	0.9820	0.9849	0.7910	0.8556	Moment - 2
	0.9812	0.9828	0.7721	0.8616	Moment - 3
	0.9810	0.9816	0.7557	0.8618	Moment - 4
	0.9813	0.9813	0.7043	0.8584	Moment - 5
Higuchi	0.9838	0.9834	0.9883	0.9664	
Peng	1.6380	1.6380	0.8187	0.8512	Method-mean
	1.7931	1.8078	0.8803	0.8533	Method = median
R/S	1.0043	1.0067	0.7902	0.8781	
	1.3319	1.2735	0.8213	0.8882	Method-per
Periodogram	0.9356	0.9291	0.8150	0.8614	Method = cum
	0.9014	1.2496	0.8161	0.9014	
BoxedPeriodogram	0.9999	0.9990	0.7737	0.8295	Method-fgn
	0.0088	1.2880	1.0802	1.0589	Method-Farma
Wavelet	1.4290	1.4352	0.8600	0.8610	

如表 5(a)和(b)所示,进一步对不同协议/应用流量进行 Hurst 指数分析,我们发现:不论是 TCP,还是 UDP;不论是从字节角度,还是从数据包角度;不论是在出境方向上,还是在入境方向上;不论是对于传统应用,还是对于非传统应用,应用多种方法估计得出的 Hurst 指数都接近甚至大于 1.尤其对比文[4,5]中的结果和表 5(a),对于在早期的流量监测分析中一直被认为具有自相似特征的传统应用 HTTP 业务,在对 LF05 进行分析时所得出的结果大不相同.先前得到的 HTTP 流量的 Hurst 指数多是落在 0.70 到 0.90 之间^[4,5],而在

表 5 对特定协议/应用流量的 Hurst 指数估计 (α) TCP

The Estimation of the Hurst Parameter of the LangFang&Leland Traffic Info							
Method	Pkt Series			Byte Series			Notes
	In	Out	All	In	Out	All	
AggVar	0.9851	0.9776	0.9850	0.9837	0.9043	0.9819	
DiffAggVar	1.4628	1.2720	1.4441	1.4003	1.3488	1.4622	
AggAbsVal	0.9877	0.9863	0.9878	0.9851	0.9014	0.9831	Moment - 1
	0.9851	0.9776	0.9849	0.9837	0.9043	0.9819	Moment - 2
	0.9841	0.9691	0.9827	0.9835	0.9053	0.9811	Moment - 3
	0.9844	0.9581	0.9815	0.9839	0.9038	0.9809	Moment - 4
	0.9852	0.9427	0.9812	0.9845	0.9009	0.9811	Moment - 5
Higuchi	0.9666	0.9664	0.9665	0.9664	0.9673	0.9666	
Peng	1.5443	1.3163	1.5266	1.5377	1.4599	1.5377	Method-mean
	1.5255	1.4848	1.4800	1.4313	1.5148	1.3578	Method = median
R/S	1.0143	1.0260	1.0154	1.0117	1.0261	1.0093	
Periodogram	1.3485	1.0656	1.2735	1.3558	1.3148	1.3319	Method-per
	0.9362	0.9133	0.9291	0.9323	0.9184	0.9356	Method = cum
BoxedPeriodogram	1.3532	1.0133	1.2691	1.1635	1.1695	1.1445	
Whittle	0.9990	0.9990	0.9990	0.999	0.999	0.999	Method-fgn
	2.3852	1.2728	1.2880	2.7595	1.0435	0.0888	Method-Farma
Wavelet	1.4469	1.3000	1.4351	1.4391	1.4441	1.4289	

(b) HTTP

The Estimation of the Hurst Parameter of the LangFang&Leland Traffic Info							
Method	Pkt Series			Byte Series			Notes
	In	Out	All	In	Out	All	
AggVar	0.9724	0.9552	0.9741	0.9805	0.9642	0.9805	
DiffAggVar	1.1491	0.9526	1.1826	1.5415	1.3289	1.5321	
AggAbsVal	0.9839	0.9855	0.9845	0.9832	0.9744	0.9831	Moment - 1
	0.9724	0.9552	0.9741	0.9805	0.9642	0.9805	Moment - 2
	0.9562	0.8816	0.9614	0.9806	0.9581	0.9805	Moment - 3
	0.9328	0.7837	0.9431	0.9815	0.9508	0.9812	Moment - 4
	0.9039	0.7038	0.9187	0.9826	0.9402	0.9821	Moment - 5
Higuchi	0.9671	0.9666	0.9669	0.9670	0.9677	0.9671	
Peng	1.3132	1.1050	1.2930	1.4563	1.2578	1.4756	Method-mean
	1.4394	1.4095	1.4563	1.3446	1.2660	1.3424	Method = median
R/S	1.0440	1.0968	1.0354	1.0659	1.0979	1.0570	
Periodogram	1.1352	0.9288	1.0835	1.2153	1.0636	1.2092	Method-per
	0.9030	0.8808	0.9021	0.9219	0.9339	0.9234	Method = cum
BoxedPeriodogram	1.1981	1.1531	1.1727	1.0796	0.9584	1.0676	
Whittle	0.999	0.999	0.999	0.999	0.999	0.999	Method-fgn
	2.9995	1.0264	1.2804	1.0972	1.2291	1.0972	Method-Farma
Wavelet	1.2322	1.1547	1.2508	1.3978	1.2539	1.4140	

LF05 中的得到的 Hurst 指数则绝大多数是接近甚至超过 1 的. 这与对 LF05 上总的字节流量和总的数据包流量进行估计分析时所得出的结论基本相同. 而且在进行更详细的协议/应用划分, 从不同角度上来进行 Hurst 指数分析时, 所得结论基本相同, 自相似是否存在充满了不确定性. 而 Hurst 指数大于 1 是序列中有非平稳性特征存在的结果^[7,14]. 我们猜测, 网络规模的不断增大、网络带宽的日益增加和新的网络应用不断涌现给网络流量带来了非平稳特征, 而正是非平稳性的存在使得在先前普遍认为的网络流量的自相似特性在当前变得

不再那么确定.

5 结论

本文应用 NetTurbo 在某城域出口链路上实际采集的流量数据, 逐层对构成、变化趋势、相关性等方面特征做了分析. 同时, 应用多种估计方法从多角度给出 Hurst 指数估计结果, 并与早期监测结果进行了比较. 研究发现, 相比先前的结论, 当前的网络流量特征发生了较为显著的变化, 主要表现在:

在传输层上, TCP 流量所占比例下降, 而 UDP 则变得不可忽略; 在应用层上, 非传统应用成为了流量的主体, 而 HTTP 传统应用对总流量的贡献比例变小;

“天特征”仍旧存在, 但是流量的相对上升区域缩短, 相对平稳区域明显拉长; 同时, 不同方向上的字节速率和数据包速率的变化皆存在有一定的差异性;

应用层上, 除 FTP Control 外, 皆存在入境方向上的数据包速率和字节速率的相关性很强, 而出境方向上的数据包速率和字节速率的相关性很弱的现象;

对于传统应用和非传统应用, 不同方向上的字节速率和数据包速率的相关性表现不尽相同; 入境字节/数据包速率和出境字节/数据包速率的相关性亦是如此;

不论在哪个层面上, 出入境方向上的数据包大小的相关性都是很弱.

网络流量自相似的程度, 甚至自相似特征是否仍然存在, 都变得不确定.

尤其最后一点的发现, 这是与先前对网络流量进行分析时可较为确定地得出其具有自相似特征的情况不尽相同的. 下一步, 将对此展开更加深入的研究工作.

参考文献:

- [1] W E Leland, et al. On the self-similar nature of ethernet traffic [J]. IEEE/ACM Transactions on Networking, 1994. 2, 2(1): 1 - 15.
 - [2] V Paxson, et al. Wide-area traffic: The failure of poisson modeling [J]. IEEE/ACM Transactions on Networking, 1995. 6, 3(3): 226 - 244.
 - [3] 张连芳, 薛飞, 舒炎泰, 李星, 王永铭. CERNET 网络业务的自相似性及性能分析 [J]. 天津大学学报 (自然科学与工程技术版), 2000, 33(3): 367 - 370.
- Lianfang Zhang, et al. Self-similarity of CERNET traffic and

- their performance evaluation[J]. Journal of Tianjin University (Science and Technology), 2000, 33(3): 367 - 370. (in Chinese)
- [4] Kihong Park, et al. On the relationship between file sizes, transport protocols, and self-similar network traffic[A]. Proceedings of the International Conference on Network Protocols[C]. Washington, DC, USA: IEEE Computer Society, 1996. 171 - 180.
- [5] Mark E Crovella, Azer Bestavros. Self-similarity in world wide web traffic: Evidence and possible causes[J]. IEEE/ACM Transactions on Networking, 1997, 5(6): 835 - 846.
- [6] A Veres, M Boda. The chaotic nature of TCP congestion control[A]. Proceedings of IEEE INFOCOM[C]. Tel-Aviv, Israel: IEEE, 2000. 1715 - 1723.
- [7] Thomas Karagiannis, et al. Long-range dependence: Ten years of Internet traffic modeling[J]. IEEE Internet Computing, 2004, 8(5): 57 - 64.
- [8] S McCreary, K Claffy. Trends in wide area IP traffic patterns: A view from Ames Internet Exchange[A]. ITC Specialist Seminar on IP Traffic Measurement, Modeling, and Management[C]. Manterey, California: CAIDA, 2000. 11.
- [9] M Fomenkov, et al. Longitudinal study of Internet traffic from 1998-2001: A view from 20 high performance sites[R]. American: CAIDA, 2003. 4.
- [10] C Fraleigh, et al. Packet-level traffic measurement from the Sprint IP backbone[J]. IEEE Network Magazine. 2003, 17(6): 6 - 16.
- [11] Thomas Karagiannis, et al. Is P2P dying or just hiding[A]. IEEE Globecom 2004[C]. Dallas, TX: IEEE, 2004. 11.
- [12] Patrice Abry, Darryl Veitch. Wavelet analysis of long-range dependence traffic[J]. IEEE Transactions on Information Theory, 1998, 44(1): 2 - 15.
- [13] Matthew Roughan, Darryl Veitch. Measuring long-range dependence under changing traffic conditions[A]. Proceedings of INFOCOM99[C]. New York, NY: IEEE, 1999.
- [14] Trang Dinh Dang. New Result in Multifractal Traffic Analysis and Modeling[D]. University of Technology and Economics Budapest, Hungary: High Speed Networks Laboratory, Department of Telecommunications and Telematics Budapest, 2002.
- [15] M S Taqqu, et al. Estimators for long-range dependence: An empirical study[J]. Fractals, 1995, 3(4): 785 - 798.
- [16] W Willinger, et al. Theory and Applications of Long-Range Dependence[M]. Boston: Birkhauser, 2003. 625 - 716.
- [17] K Thompson, et al. Wide-area Internet traffic patterns and characteristics[J]. IEEE/ACM Transactions on Networking, 1997, 3(3): 10 - 23.

作者简介:



张广兴 男, 1978 年生于河北, 湖南大学计算机与通信学院博士研究生. 主要研究方向为网络监测、流量测量与特征分析.

E-mail: guangxing@ict.ac.cn



张大方 男, 1959 年生于上海, 博士, 教授, 博士生导师, CCF 高级会员. 主要研究领域为可信系统与网络、容错计算.