

无线局域网聚集业务流非平稳性检验^{*}

冯慧芳¹, 苗文华¹, 舒炎泰²

(1. 西北师范大学 数学与信息科学学院, 兰州 730070; 2. 天津大学 计算机科学与技术学院, 天津 300072)

摘要: 无线网络业务特性分析是网络性能评价、规划、设计、管理和控制的基础。分析了真实无线局域网业务流的基本统计特性, 并应用单位根检验方法检验无线局域网中的聚集业务流的非平稳性。实验结果表明, 不论是下行的还是上行的聚集流在不同显著水平上存在非平稳性; 同时对这些聚集业务流的一阶差分序列进行检验, 结果表明差分序列都是平稳的, 这说明下行和上行的聚集业务流均是含有一个单位根的非平稳序列, 这为无线网络业务流的进一步建模提供了指导。

关键词: 无线局域网; 聚集业务流; 单位根检验; 非平稳性

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2010)01-0305-03

doi:10.3969/j.issn.1001-3695.2010.01.090

Testing non-stationarity of aggregate WLAN traffic

FENG Hui-fang¹, MIAO Wen-hua¹, SHU Yan-tai²

(1. College of Mathematics & Information Science, Northwest Normal University, Lanzhou 730070, China; 2. School of Computer Science & Technology, Tianjin University, Tianjin 300072, China)

Abstract: The study of WLAN traffic characteristics is not only important for network performance evaluation, but also imperative for the proper network design, management, and control. This paper employed actual wireless data that draw from well known archives of network traffic traces analysis the preliminary statistics and examined the non-stationarity of aggregate WLAN traffic by employing the unit-root tests. The results indicate that the upstream and downstream traffic are non-stationary at different levels of statistical significance. It also repeated the analysis for the first difference of these series. The findings show that these first difference series are stationary. These results show that aggregate traffic series are non-stationary and also integrated of order one. These provide guidance on modelling of wireless traffic in the future.

Key words: wireless local area network(WLAN); aggregate traffic; unit root tests; non-stationarity

0 引言

网络业务流特性是网络规划与设计、网络管理、网络性能分析和 QoS(quality of service)服务中所必须考虑的一个重要因素。只有正确地认识网络业务流的特性,才能建立精确的业务流模型,而精确的网络业务流模型可以刻画网络的动态特性,帮助人们设计更合理的网络拓扑结构、更好的网络协议、更高效的 QoS 服务保证手段、更好的队列管理和调度算法,保证网络高效、稳定、安全地运行。

与有线网络相比较,无线局域网有与其不同的某些特点,包括用户移动性、拓扑的动态变化、无线链路的不可靠,以及资源的匮乏等^[1]。因此有线网络中的业务流的建模方法在无线局域网中的有效性还有待验证^[2-4],不能将其直接应用到无线局域网中。另外,无线局域网的信道接入机制以及多径衰落等信道特性都会对业务流特性产生影响,深入理解无线局域网业务流的特征,对于分析与改进无线局域网传输协议性能,设计更好的 MAC(media access control)层机制和缓存调度算法都是十分重要的。因此,随着当前无线局域网应用的蓬勃发展,

为了有效利用无线局域网资源,优化网络性能,研究这一环境下的网络业务流特性问题是非常迫切和有意义的。

近年来,各国研究人员对各种无线网络业务流的统计特征也进行了一些研究。文献[5]用马尔可夫理论和小波理论研究了无线信道的分形特征,文献[6]研究了聚集业务流的非线性。Jiang 等人对 CDPD 网的业务数据的研究表明,该网络业务流具有自相似性^[7]。文献[8]对 Ad hoc 网络的业务流进行的分析也发现,Ad hoc 网络业务流呈自相似性和长相关性。文献[9]用 NS-2 仿真研究 WLAN 和 Ad hoc 网络中业务流的特性,仿真结果表明,单个流具有自相似性,而多个聚集流具有多重分形的特性。以上这些研究,主要是针对无线局域网中聚集业务流统计特性的研究,基本方法均采用聚集方差时间法、周期图法或 R/S(rescaled range analysis)方法计算无线网络业务流的 Hurst 参数,根据 Hurst 参数的值检验无线网络业务流的自相似性,结论是无线局域网聚集业务流具有自相似性。时间序列的自相似性与其平稳性(或非平稳性)之间没有必然关系,即自相似序列可以是平稳的,也可以是非平稳的,那么无线局域网聚集业务流是否非平稳?

收稿日期: 2009-05-24; **修回日期:** 2009-06-22 **基金项目:** 国家自然科学基金资助项目(60702038,90604013);国家“863”计划资助项目(2007AA01Z220);国家教育部科技创新工程重大项目培育资金资助项目(708024);西北师范大学科技创新工程骨干培育资助项目(NWNU-KJCXGC-03-52);甘肃省教育厅基金资助项目(0901-03)

作者简介: 冯慧芳(1971-),女,甘肃古浪人,副教授,博士,主要研究方向为计算机网络性能评价(hffeng@nwnu.edu.cn);苗文华(1984-),硕士研究生,主要研究方向为计算机网络性能评价;舒炎泰(1942-),男,天津人,教授,博导,主要研究方向为计算机网络。

1 单位根检验

平稳序列将围绕一个常数均值波动,并有向其靠拢的趋势,其均值与时间无关,而且其方差是有限的,并不随着时间的推移而产生系统的变化,而非平稳过程则不具有这个性质。目前普遍采用的单位根检验方法主要有两种:a)由迪基—富勒(Dickey-Fuller)提出的增项迪基—富勒(augmented Dickey-Fuller)检验,简称 ADF 检验^[10];b)由菲利普斯—皮荣(Phillips-Perron)提出的菲利普斯—皮荣检验,简称 PP 检验^[11]。本文中利用 ADF 检验对无线局域网业务流进行单位根检验。该方法的检验原理为:通过 n 次差分的办法将非平稳序列转换为平稳序列,具体方法是估计回归方程式

$$\Delta X_t = \alpha_0 + \alpha_1 t + \alpha_2 X_{t-1} + \sum_{i=1}^k \beta_{t-i} \Delta X_{t-i} + \mu_t \quad (1)$$

其中: $\{X_t, t=0, 1, \dots\}$ 为时间序列; α_0 为常数项; t 为时间趋势项; k 为滞后阶数(最优滞后阶数); μ_t 为残差项。该检验的原假设 $H_0: \alpha_2 = 0$;备择假设 $H_1: \alpha_2 \neq 0$ 。检验过程如下:a)根据所观测到的时间序列数据,利用最小二乘法估计式(1)中的回归系数 α_2 ,得到其估计值 $\hat{\alpha}_2$ 。b)仍根据最小二乘法估计 $\hat{\alpha}_2$ 的标准差 \hat{S}_{α_2} ,得 ADF 检验统计量 $\tau = \hat{\alpha}_2 / \hat{S}_{\alpha_2}$ 。c)将得到的 τ 值与其分布的给定显著水平 α 上的渐进临界值 $\tau(\alpha)$ 进行比较。若 $\tau > \tau(\alpha)$,则不拒绝 $\alpha_2 = 0$ 的原假设,即时间序列有一个单位根,序列是非平稳的;否则拒绝 $\alpha_2 = 0$ 的原假设,这时时间序列没有单位根,序列是平稳的(在给定显著水平 α 上)。

在 ADF 检验中,最优滞后阶数 k 可以采用赤池信息准则(AIC)确定^[11]。AIC 是确定最优滞后阶数的一种方法,是 20 世纪 70 年代日本学者赤池弘次(Akaike)提出一个基本信息量的定阶准则,按照使得该准则达到最小值来定阶。其基本原理是 $AIC(p) = n \ln(\sigma_p^2) + 2p$ 。其中: n 为时间序列的长度; σ_p^2 为残差的方差。按照 AIC 准则,使得 $AIC(p)$ 达到最小值的模型被认为是可以接受的好模型,即满足 $AIC(k) = \min\{AIC(p) | p = 1, 2, 3, \dots\}$ 的 k 就是最优滞后阶数。

在实际检验中,也可以采用相伴概率 p -value(本文采用 MacKinnon 单边检验的相伴概率)的值直接进行检验^[12], p -value 值也称为统计量精确置信水平,它可定义为拒绝原假设的最低置信水平, p -value 的值越小,越能拒绝原假设。用 p -value 值的优点是避免了在选择显著水平 $\alpha(1\%, 5\%, 10\%)$ 时的任意性。现在许多统计软件都能计算各种统计量的 p -value 值,所以可以直接用它替代检验统计量实际值进行检验,而不必去查阅有关统计表并比较临界值了。比如,在单位根检验中,如果 p -value 值在 0.01 ~ 0.99,则不拒绝 H_0 假设,即接受原假设,存在单位根,所以序列是非平稳的;否则拒绝 H_0 假设,序列是平稳的。

2 聚集业务流的非平稳性分析

本文分析中采用的真实无线局域网业务流来自国内外不同的无线局域网。其中一个是在 Stanford 大学 WaveLAN 无线局域网上采集的业务流 final.anon^[13];一个是 2001 年 8 月美国圣地亚哥召开的 ACM SIGCOMM'01 会议期间采集无线局域网业务流 trace.pcap^[14];还有一个是从天津大学计算机网络研究室 WLAN 实验床采集的无线局域网业务流 t030801。本文把三个业务流的下行和上行流分别按照不同的时间粒度进行聚

集。序列名称由三部分组成:原序列名 + D(或 U) + 聚集时间粒度,如 finalD10 s 表示 final 的下行业务按照 10 s 的时间粒度聚集的序列,finalU10 s 表示 final 的上行业务按照 10 s 的时间粒度聚集的序列。本文用该序列名字前面加 Δ 表示各个序列的一阶差分序列。

2.1 下行业务流的统计特性

表 1 给出了下行业务流的基本统计特性,包括均值、最大值、最小值、标准差、偏度(skewness)、峰度(kurtosis)和 Jarque-Bera 检验。偏度衡量序列分布围绕其均值的非对称性,如果序列的分布是对称的,skewness 值为 0;正的 skewness 值意味着序列分布有长的右拖尾,负的 skewness 值意味着序列分布有长的左拖尾。峰度度量序列分布的凸起或平坦程度,正态分布的 kurtosis 值为 3。如果 kurtosis 值大于 3,分布的凸起程度大于正态分布;如果 kurtosis 值小于 3,序列分布相对于正态分布是平坦的。Jarque-Bera 检验序列是否服从正态分布。

表 1 下行聚集业务流序列的基本统计量

业务流序列	均值	最大值	最小值	标准差	偏度	峰度	Jarque-Bera
finalD10s	58.223 5	755	0.000 0	115.835 0	2.054 6	6.137 4	0.000 0
finalD20s	271.359 3	2 893	0.000 0	311.023 7	0.836 1	4.076 5	0.0000
finalD30s	494.824 7	7 735	0.000 0	585.508 6	3.122 3	24.844 1	0.000 0
traceD10s	685.504 7	4 601	0.000 0	652.985 6	1.665 5	7.364 2	0.000 0
traceD20s	1 028.842 0	8 161	0.000 0	1 259.151 0	1.803 9	7.692 4	0.000 0
traceD30s	2 056.514 0	12 350	0.000 0	1 908.595 0	1.625 562	7.309 194	0.000 0
t030801D1s	55.322 0	167	0.000 0	32.928 2	0.326 3	2.881 1	0.000 1
t030801D10s	665.058 7	1 257	0.000 0	222.174 7	-0.429 7	3.533 8	0.0003
t030801D20s	1 326.580 0	2 252	4.000 0	424.870 1	-0.680 9	3.635 8	0.000 1

从表 1 可以看出,下行业务流序列具有不为零的偏度和较大的峰度, Jarque-Bera 检验也拒绝了正态分布的假设。

2.2 下行业务流的单位根检验

表 2 给出了下行业务聚集流的 ADF 检验结果,表中给出了 ADF 检验的统计量、最优 AIC 阶数和 p -value 的值。其既可以根据 p -value 的值检验序列的非平稳性,也可以根据 ADF 检验的统计量值检验序列的非平稳性。

表 2 下行聚集业务流分序列的 ADF 检验结果

业务流序列	包含常数项			包含常数项和线性时间趋势项		
	MAIC 阶	ADF 统计量	p -value	MAIC 阶	ADF 统计量	p -value
finalD10s	20	-2.396 3 *	0.142 8	20	-2.438 7 *	0.359 2
finalD20s	27	-2.388 1 *	0.145 1	27	-2.312 3 *	0.426 6
finalD30s	29	-4.097 4	0.001 0	29	-4.271 8	0.003 5
traceD10s	26	-3.044 3 ***	0.031 1	26	-3.177 3 **	0.089 2
traceD20s	24	-2.260 5 *	0.185 2	24	-2.746 0 *	0.218 0
traceD30s	16	-2.252 5 *	0.188 0	16	-2.394 6 *	0.382 1
t030801D1s	10	-3.340 3 ***	0.013 4	10	-3.391 5 **	0.053 0
t030801D10s	11	-3.778 1	0.003 4	11	-3.547 4 ***	0.036 0
t030801D20s	5	-3.699 5	0.004 8	5	-3.280 9 **	0.072 7

注: *、**、*** 分别表示在 10%、5%、1% 置信水平上的显著性,以下各表记号表示类似。

从表 2 中可以看出,除了序列 finalD30 s 之外,其他各下行业务流在不同时间粒度的聚集流的 p -value 值都大于 0.01,所以不能拒绝 H_0 假设,接受原假设,故这些序列都是非平稳的;根据 ADF 检验统计量值,也可以确定这些序列是非平稳的。由 p -value 值和 ADF 检验统计量值均可确定序列 finalD30 s 是平稳的。对这些聚集业务流作一阶差分后,再进行平稳性检验发现,所有序列的 MIC 阶均为 0,差分后序列的 p -value 值都为 0(表 3),所以拒绝 H_0 假设,即各业务流的一阶差分序列不存在单位根,是平稳序列,因此笔者认为无线局域网业务流的下行聚集流都是一阶单整序列。

2.3 上行业务流的统计特性

表 4 给出了上行业务流的基本统计特性。从表 4 可以看

出,上行业务流序列也具有不为零的偏度和较大的峰度, Jarque-Bera 检验拒绝了正态分布的假设。

表 3 下行聚集业务流一阶差分序列的 ADF 检验结果

业务流序列	包含常数项		包含常数项和线性时间趋势项	
	ADF 统计量	p-value	ADF 统计量	p-value
ΔfinalD10s	-92.233 9	0.000 0	-92.222 4	0.000 1
ΔfinalD20s	-86.261 5	0.000 1	-86.251 1	0.000 1
ΔfinalD30s	-74.471 3	0.000 1	-74.462 1	0.0001
ΔtraceD10s	-65.747 5	0.000 0	-65.736 8	0.000 0
ΔtraceD20s	-50.926 1	0.000 1	-50.913 5	0.000 0
ΔtraceD30s	-34.476 1	0.000 0	-34.459 6	0.000 0
Δt030801D1s	-47.255 2	0.000 1	-47.233 0	0.000 0
Δt030801D10s	-23.484 8	0.000 0	-23.475 8	0.000 0
Δt030801D20s	-15.581 0	0.000 0	-15.617 3	0.000 0

2.4 上行业务流的单位根检验

本文用同样的方法对上行业务流的非平稳性进行检验。检验结果如表 5 所示。从表中可以看出, finalU20s、finalU30s、traceU10s 包含常数项的 ADF 检验表明这些序列是平稳的,其余各上行业务流在不同时间粒度的聚集流的 p-value 值都大于 0.01,所以不能拒绝 H_0 假设,这些序列都是非平稳的。对这些聚集业务流的一阶差分序列进行检验,检验结果如表 6 所示,它们的 MIC 阶均为 0,p-value 值均为 0,拒绝 H_0 假设,所以各业务流的上行聚集流的一阶差分序列是平稳序列,无线局域网业务流的上行聚集流也都是—阶单整序列。

表 4 上行聚集业务流序列的基本统计量

业务流序列	均值	最大值	最小值	标准差	偏度	峰度	Jarque-Bera
finalU10s	112.637 2	792	0.000 0	185.077 6	1.350 0	3.212 2	0.000 0
finalU20s	386.827 2	3 451	0.000 0	441.219 4	0.965 2	3.986 8	0.000 0
finalU30s	487.264 0	4 157	0.000 0	596.170 9	1.113 921	4.077 1	0.000 0
traceU10s	506.074 4	2 259	0.000 0	442.151 0	0.792 3	3.346 8	0.000 0
traceU20s	963.990 0	3 667	0.000 0	684.643 1	0.577 1	2.778 4	0.000 0
traceU30s	1 445.985 0	4 399	49.000 0	764.433 6	0.171 375	2.633 5	0.000 0
t030801U1s	26.293 7	146	0.000 0	24.758 8	1.124 4	4.401 8	0.000 0
t030801U10s	285.626 7	823	0.000 0	154.316 9	0.613 4	3.566 2	0.000 001
t030801U20s	569.734 0	1 428	3.000 0	282.999 2	0.311 8	2.974 6	0.217 6

表 5 上行聚集业务流序列的 ADF 检验结果

业务流序列	包含常数项			包含常数项和线性时间趋势项		
	MAIC 阶	ADF 统计量	p-value	MAIC 阶	ADF 统计量	p-value
finalU10s	29	-2.006 9 *	0.284 0	29	-2.232 3 *	0.470 9
finalU20s	31	-3.630 1	0.005 3	31	-3.874 1 ***	0.013 1
finalU30s	27	-3.483 9	0.008 5	27	-3.542 0 ***	0.035 1
traceU10s	31	-3.729 9	0.003 8	31	-4.364 8	0.002 5
traceU20s	30	3.231 16 ***	0.018 3	30	2.806 9 *	0.194 7
traceU30s	26	2.979 2 ***	0.037 0	26	2.518 2 *	0.319 1
t030801U1s	10	3.192 1 ***	0.020 8	10	3.8038 ***	0.016 8
t030801U10s	11	3.364 3 ***	0.012 9	11	3.154 7 **	0.095 4
t030801U20s	5	3.288 8 ***	0.016 8	5	3.041 8 *	0.123 9

表 6 上行聚集业务流—阶差分序列的 ADF 检验结果

业务流序列	包含常数项		包含常数项和线性时间趋势项	
	ADF 统计量	p-value	ADF 统计量	p-value
ΔfinalU10s	-105.245 9	0.000 1	-105.235 6	0.000 1
ΔfinalU20s	-94.793 5	0.000 1	-94.784 1	0.000 1
ΔfinalU30s	-92.311 1	0.000 1	-92.302 1	0.000 1
ΔtraceU10s	-133.997 4	0.000 1	-133.984 0	0.000 1
ΔtraceU20s	-113.041 7	0.000 1	-113.058 6	0.000 1
ΔtraceU30s	-528.444 9	0.000 1	-528.375 7	0.000 1
Δt030801U1s	-47.255 2	0.000 0	-44.943 0	0.000 0
Δt030801U10s	-23.484 8	0.000 0	-23.800 9	0.000 0
Δt030801U20s	-15.581 0	0.000 0	-16.551 9	0.000 0

综上所述,从单位根检验结果中可以看出,在本文分析的

样本中,不论是下行的还是上行的无线局域网聚集流,在不同显著水平上存在非平稳性,所有聚集业务流序列均是含有一个单位根的非平稳序列。

3 结束语

本文首先对真实无线局域网聚集业务流的基本统计特性进行了分析;然后采用目前普遍应用的 ADF 检验方法,对上行和下行聚集业务流的非平稳性进行了检验,研究表明,绝大部分下行和上行的聚集流在不同显著水平上存在非平稳性。另外,对这些聚集业务流的一阶差分序列的检验结果表明,差分序列都是平稳的,因此,无线局域网聚集业务流均是含有一个单位根的非平稳序列。现在又有大量的真实无线局域网 trace 公布在因特网上,需要对更多的无线局域网的聚集业务流作进一步的分析,掌握无线局域网业务流的特性,为建立更加精确的流量模型提供指导。

参考文献:

- [1] IEEE Std 802.11b-WLAN, Medium access control (MAC) and physical layer (PHY) specifications; high speed physical layer (PHY) in the 2.4 GHz band [S]. 1999.
- [2] LELAND W, TAQUU M, WILLINGER W, et al. Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements [J]. *Statistical Scitical*, 1995, 10(1): 67-85.
- [3] KARAGIANNIS T, MOLLE M, FALOUTSOS M, et al. A nonstationary poisson view of Internet traffic [C]//Proc of IEEE Conference on Computer Communications. 2004; 1558-1569.
- [4] ALBERTO D, ANTONIO P, PIERLUIGI S R, et al. Internet traffic modeling by means of hidden Markov models [J]. *Computer Networks*, 2008, 52(14): 2645-2662.
- [5] SYED A, KHAYAM, H R, SELIN A, et al. Markov and multifractal wavelet models for wireless MAC-to-MAC channels [J]. *Performance Evaluation*, 2007, 64(4): 298-314.
- [6] FENG Hui-fang, SHU Yan-tai, YANG W W. Nonlinear analysis of wireless LAN traffic [J]. *Nonlinear Analysis: Real World Applications*, 2009, 10(2): 1021-1028.
- [7] LING Qi-lian. Ad hoc wireless network traffic self-similarity and forecasting [J]. *IEEE Communications Letters*, 2002, 6(7): 297-299.
- [8] YIN Shou-yi, LIN Xiao-kang. Traffic self-similarity in mobile Ad hoc networks [C]//Proc of the 2nd IFIP International Conference on Wireless and Optical Communications Networks. 2005; 285-289.
- [9] TICKOO O, SIKDAR B. On the impact of IEEE 802.11 MAC on traffic characteristics [J]. *IEEE Journal on Selected Areas in Communications*, 2003, 21(2): 189-203.
- [10] DICKEY D A, BELL W R, MILLER R B. Unit roots in time series models: tests and implications [J]. *American Statistician*, 1986, 40(1): 12-26.
- [11] PHILLIPS P C B, PERRON P. Testing for a unit root in time series regression [J]. *Biometrika*, 1988, 75(2): 335-346.
- [12] BOX G E P, JENKINS G M, REINSEL G C. Time series analysis: forecasting and control [M]. 2nd ed. [S. l.]: Prentice-Hall, 1994.
- [13] MACKINNON J G. Numerical distribution functions for unit root and co-integration tests [J]. *Journal of Applied Econometrics*, 1996, 11(6): 601-618.
- [14] [EB/OL]. <http://mosquitonet.stanford.edu/software.html>.
- [15] [EB/OL]. <http://ramp.ucsd.edu/pawn/sigcomm-trace>.