

Research Article

Stationarity Testing of Accumulated Ethernet Traffic

Zhiping Lu,¹ Ming Li,² and Wei Zhao³

¹ School of Finance and Statistics, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China

² School of Information Science and Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, China

³ Department of Computer and Information Science, University of Macau, Av. Padre Tomas Pereira, Taipa, Macau

Correspondence should be addressed to Zhiping Lu; camping9@gmail.com

Received 26 January 2013; Revised 12 March 2013; Accepted 13 March 2013

Academic Editor: Shengyong Chen

Copyright © 2013 Zhiping Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigate the stationarity property of the accumulated Ethernet traffic series. We applied several widely used stationarity and unit root tests, such as Dickey-Fuller test and its augmented version, Phillips-Perron test, as well as the Kwiatkowski-Phillips-Schmidt-Shin test and some of its generalizations, to the assessment of the stationarity of the traffic traces at the different time scales. The quantitative results in this research provide evidence that when the time scale increases, the accumulated traffic series are more stationary.

1. Introduction

Stationarity testing is essential for time series. According to the meaning of nonstationarity discussed in [1, 2], investigating time-varying spectra turns to be a natural way of testing nonstationarity of time series [3–7]. However, one may encounter difficulties in nonstationarity testing of long-range dependent (LRD) network traffic (traffic for short) because the LRD property implies $1/f$ noise, which is divergent at the zero frequency [8]. In addition to time-frequency distributions, there are many approaches in this regard, such as bootstrap testing [9] and so on [10–12]. However, conventional approaches may not be properly used for testing the nonstationarity of LRD traffic as can be seen from Abry and Veitch [13], Grossglauser and Bolot [14]. That difficulty may also refer to a paper by Mandelbrot in 1976 [15]. Abry and Veitch [13] proposed a test method for LRD traffic by investigating the time invariability of the Hurst parameter, H . However, in terms of engineering, time-varying H may not always imply nonstationarity but multifractal in general, see [16, 17] for details. Therefore, further research of nonstationarity of LRD traffic is desired.

The paper aims at providing the following contributions.

- (i) Different scales are taken into account for investigating the stationarity of the accumulated traffic data through unit root tests.

- (ii) The quantitative description of the large scales for the Ethernet traffic to be stationary is given.

The remainder of this paper is organized as follows. In Section 2, the dataset is described. In Section 3, several widely used unit root tests and stationarity tests are presented. In Section 4, the numerical results of the statistical tests are included and the discussion is followed. Finally, Section 5 presents some concluding remarks.

2. Datasets

This research utilizes four real Ethernet traffic traces, listed in Table 1, which were measured on an Ethernet at the Bellcore Morristown Research and Engineering facility [18] in 1989. We may refer to [19] for more details about the datasets.

Denote by $x(t(i))$ a traffic time series, where $t(i)$ ($i = 0, 1, \dots$) is the time-stamp series, which indicates the time stamp of the i th packet. Note that $x(t(i))$ represents the packet size of the i th packet at time $t(i)$. This research uses $x(i)$ to represent the packet size of the i th packet on a packet-by-packet basis. Further, on an interval-by-interval basis, we consider the accumulated traffic, denoted by $y(n)$. It is given by

$$y(n) = \sum_{i=nT}^{(n+1)T} x(i), \quad (1)$$

TABLE 1: Four Ethernet traffic traces.

| Series name | Starting time | Duration | Series length |
|-------------|---------------------|------------|---------------|
| pAug.TL | 11:25 AM, 29 Aug 89 | 52 minutes | 1 million |
| pOct.TL | 11:00 AM, 05 Oct 89 | 29 minutes | 1 million |
| OctExt.TL | 11:46 PM, 03 Oct 89 | 34.111 h | 1 million |
| OctExt4.TL | 2:37 PM, 10 Oct 89 | 21.095 h | 1 million |

where T is the interval width. It is in fact the accumulation scale. Without causing confusion, we still call it time scale (or scale for short) throughout the paper. Thus, $y(n)$ stands for the accumulated bytes of arrival traffic in the n th interval with the time scale T . We shall exhibit an interesting phenomenon that the stationarity of $y(n)$ of the Ethernet traffic traces considerably relies on the time scale T . More precisely, qualitatively speaking, $y(n)$ is nonstationary if T is small while it is stationary when T is large. The key point of this research is to exhibit the quantitative descriptions of T at which $y(n)$ changes from the non-stationary case to the stationary one.

3. Statistical Tools

Computer scientists concern about the stationarity property of traffic, see for example, [13]. Nevertheless, the consensus of that property of traffic may not be achieved due to different attributes of traffic. For instance, our research investigates the stationarity property of traffic at different time scales using the attribute of accumulation scale, which is rarely reported, to the best of our knowledge. For that reason, we explain our results in what follows.

On one hand, from Figures 1, 2, 3, and 4, we could observe the trajectories of the accumulated $y(n)$ under different time scales. Visually, it seems that there exists some nonstationarity in all the four series, especially under the large time scales. However, we need some more objective methods to verify the reliability of this opinion. On the other hand, we present in Figures 5, 6, 7, and 8 the spectrum of the accumulated $y(n)$ under different time scales, which clearly indicates the existence of long range dependence in the series.

It is important to learn about the statistical property of the data before quantitatively using the accurate models. In statistics, one of the most investigated branches is that of unit root and stationarity testing. Suppose a discrete time stochastic process can be written as an autoregressive process. If 1 is a root of the characteristic equation, the stochastic process is said to have a unit root or, alternatively, is integrated of order one, denoted $I(1)$ and is non-stationary. If the other roots of the characteristic equation lie inside the unit circle—that is, have a modulus (absolute value) less than one—then the first difference of the process will be stationary. This is the original idea of the unit root testing proposed in 1979, and hereafter many extensions and modifications have been developed. In the following we concentrate on the existing stationarity testing as well as some of their variants and generalizations to identify the existence of unit root in the series.

The standard Dickey-Fuller test (DF) [20] is based on i.i.d. errors and has as the null hypothesis the unit root. The DF test is valid if the time series is well characterized by an Auto-Regressive (AR(1)) process with white noise errors. Many time series, however, have a more complicated dynamic structure than is captured by a simple AR(1) model. Said and Dickey (1984) [21] augment the basic autoregressive unit root test to accommodate Auto-Regressive and Moving-Average (ARMA(p, q)) models with unknown orders and their test is referred to as the augmented Dickey-Fuller (ADF) test. On the other hand, the Phillips-Perron test (PP) [22] is non-parametric and allows for some heterogeneity and serial correlation in the innovations, which is different from the ADF tests. Some class of so-called efficient unit root tests was proposed by Elliott et al. (1996) (hereafter ERS) [23] whose test statistics come very close to the power envelope for a wide range of alternatives, and they can have substantially higher power than the ADF or PP unit root tests. Unlike AR(MA) unit root tests, stationarity tests are far less numerous and have as null hypothesis the stationarity assumption and as alternative the unit root. Among the most well known stationarity tests, the most extensively used one is the Kwiatkowski et al. (1992) test, also known as KPSS [24], which is intended to complement unit root tests and can be used to distinguish short memory and long memory stationary processes.

There exist many other unit root and stationarity tests as well as generalizations and combinations of the ones mentioned above. For supplementary material on unit root and stationarity testing, see [10, 25–29], and so forth, for more details.

However, one fundamental problem is that there is still no consensus on the optimal choice of the stationarity test. Therefore, to be more objective, we take advantage of several common tests to make a comprehensive study of the Ethernet traffic data in the following empirical study.

4. Discussion

In this section, we present the statistical testing results of the four series, that is, pAug.TL, pOct.TL, OctExt.TL, and OctExt4.TL data. Our interest lies in the identification of the existence of the unit root in the series. We finally adopt the most widely used ADF, PP, ERS, and KPSS tests in the empirical application. Based on the software R, we mainly utilize the functions of the packages “urca” and “fUnitRoots” to realize the statistical tests. The null hypothesis is that the series is $I(0)$ stationary, while the alternative is that it is $I(1)$. The P value of each test under the time scales $T = 2^n$, $n = 9, \dots, 17$ are presented in Tables 2, 3, 4, and 5. It is worth mentioning that we have also tried to utilize the Robinson (1994) test [30] to test the stationarity of the series due to their identified long range dependence features. However, according to the result of Ferrara et al. (2010) [31], this test is appropriate to use only for series whose sample size is at least 3000, which could never be satisfied in our case.

As seen in Tables 2–5 as time scale increases, the accumulated traffic series $y(n)$ is more likely to be stationary,

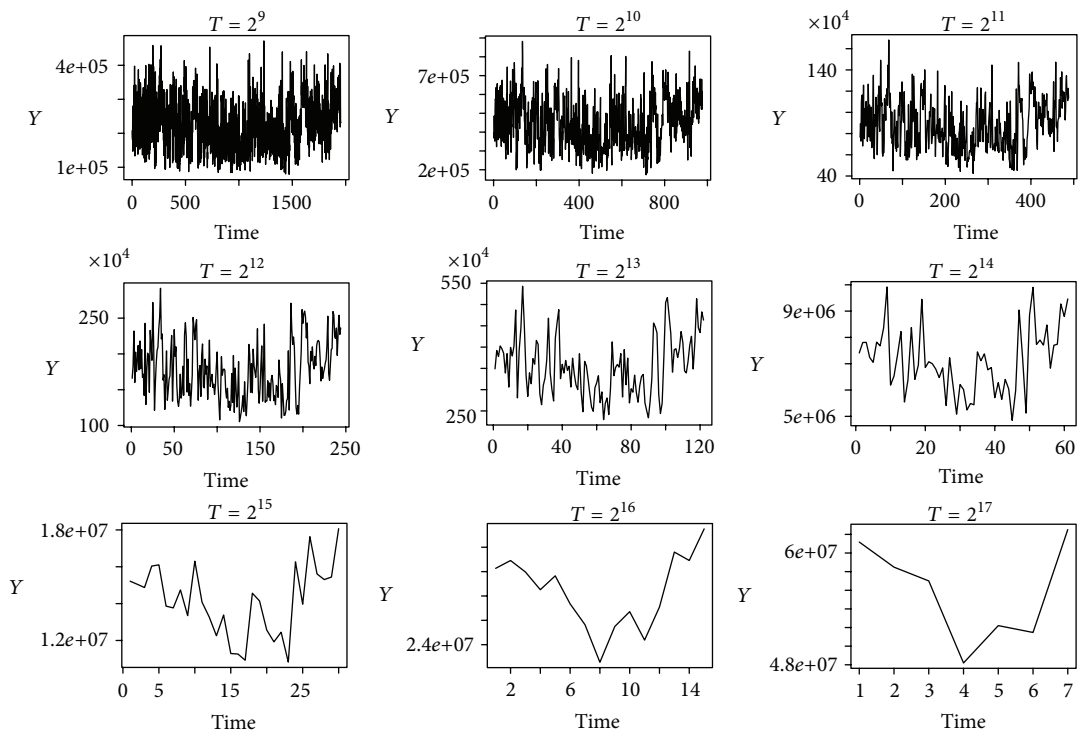


FIGURE 1: Trajectory of the accumulated pAug.TL data under different time scales.

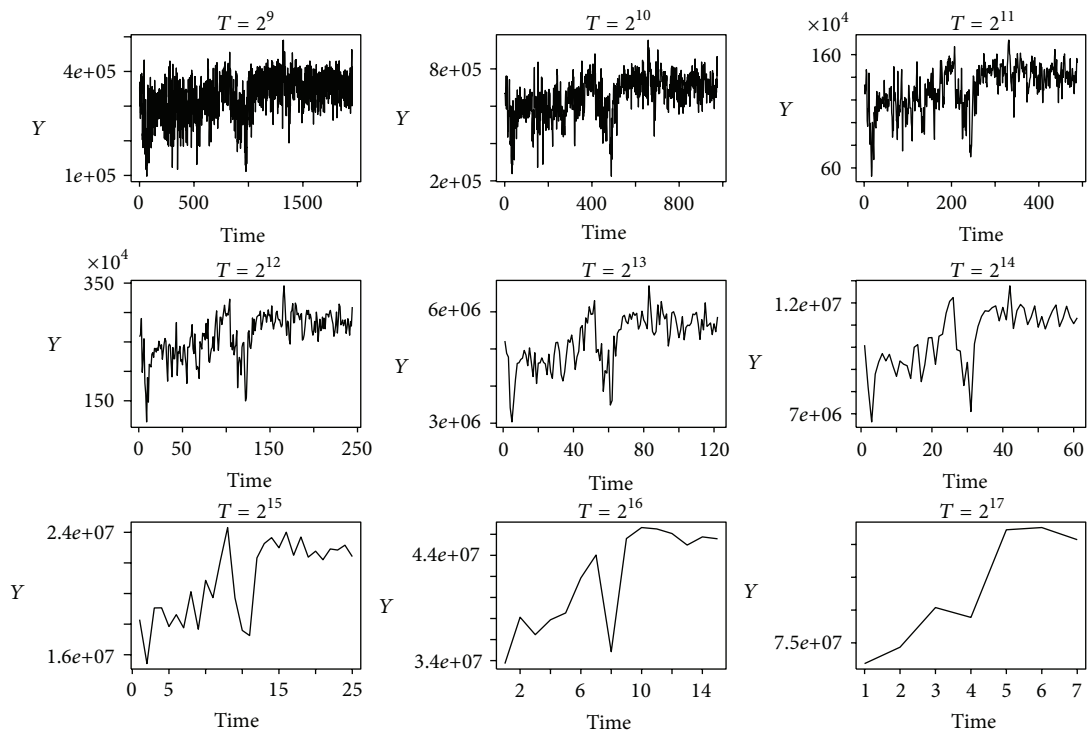


FIGURE 2: Trajectory of the accumulated pOct.TL data under different time scales.

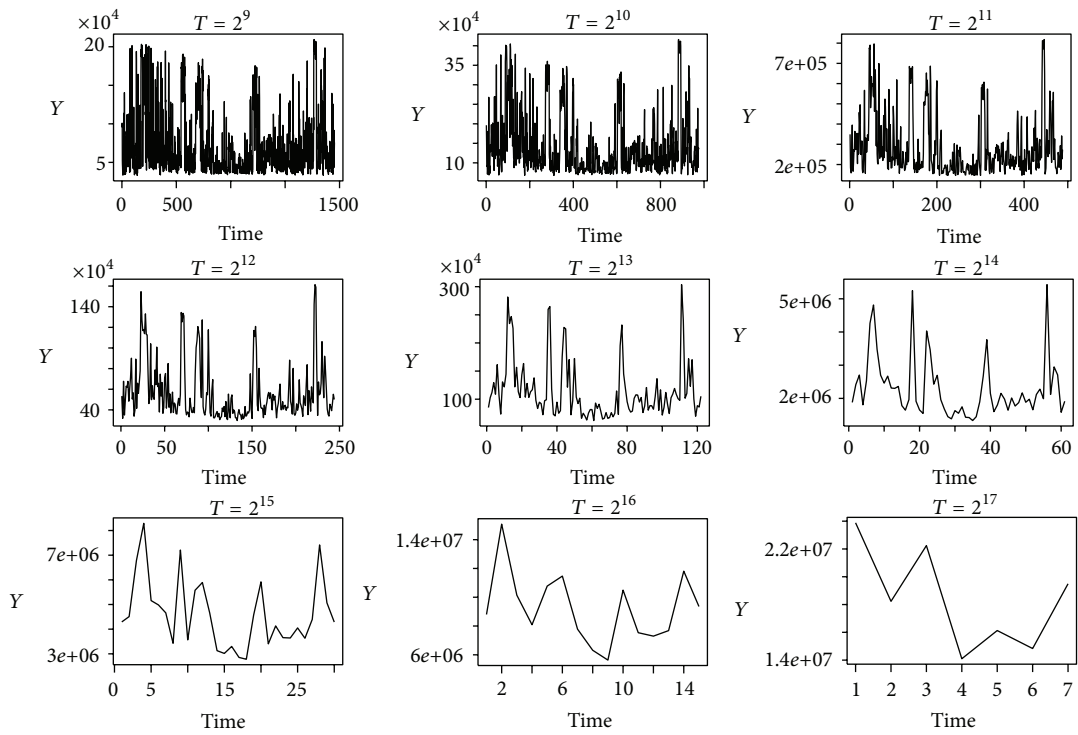


FIGURE 3: Trajectory of the accumulated Octxt.TL data under different time scales.

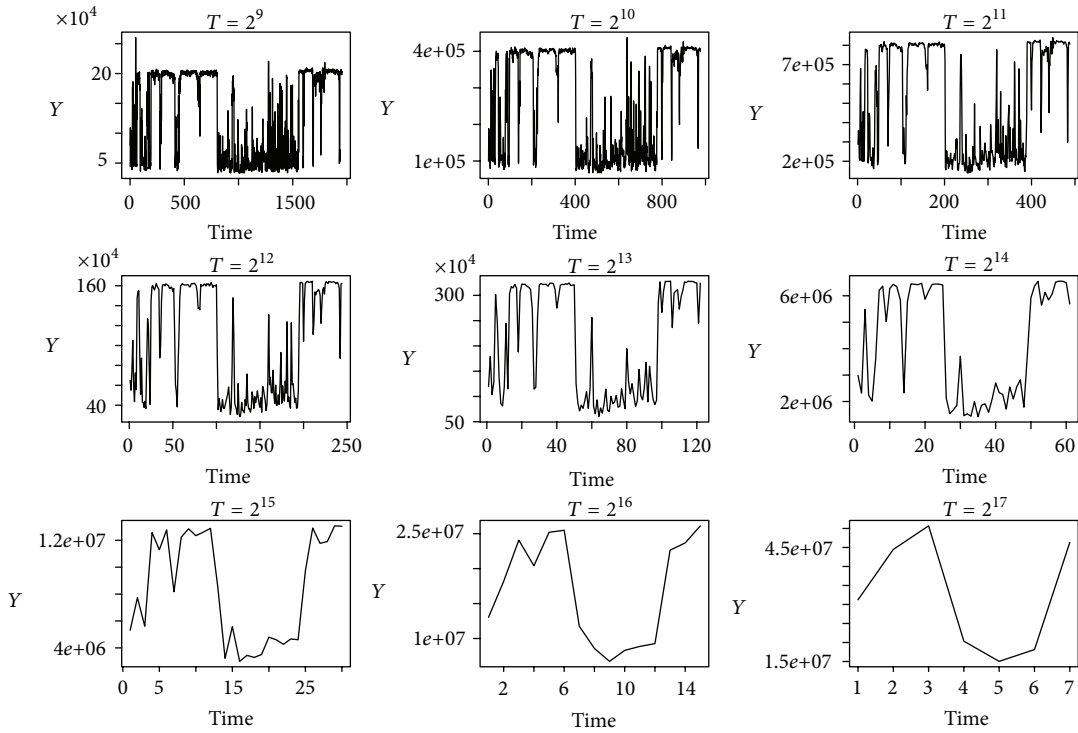


FIGURE 4: Trajectory of the accumulated OctExt4.TL data under different time scales.

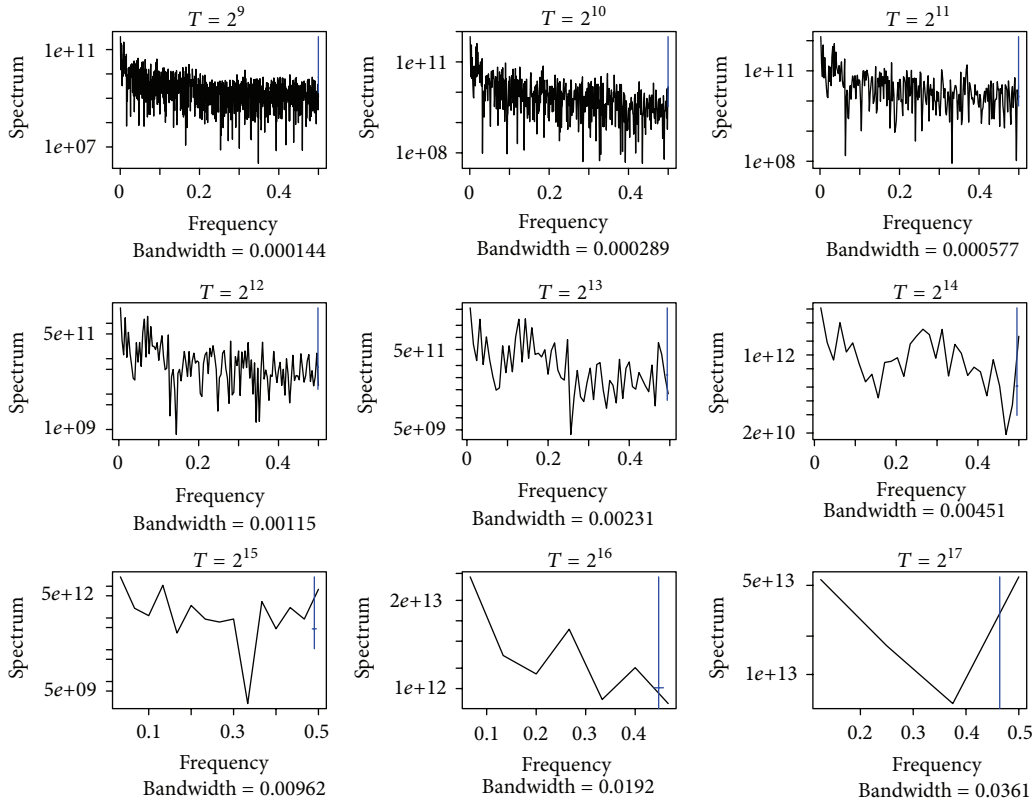


FIGURE 5: Spectrum of the accumulated pAug.TL data under different time scales.

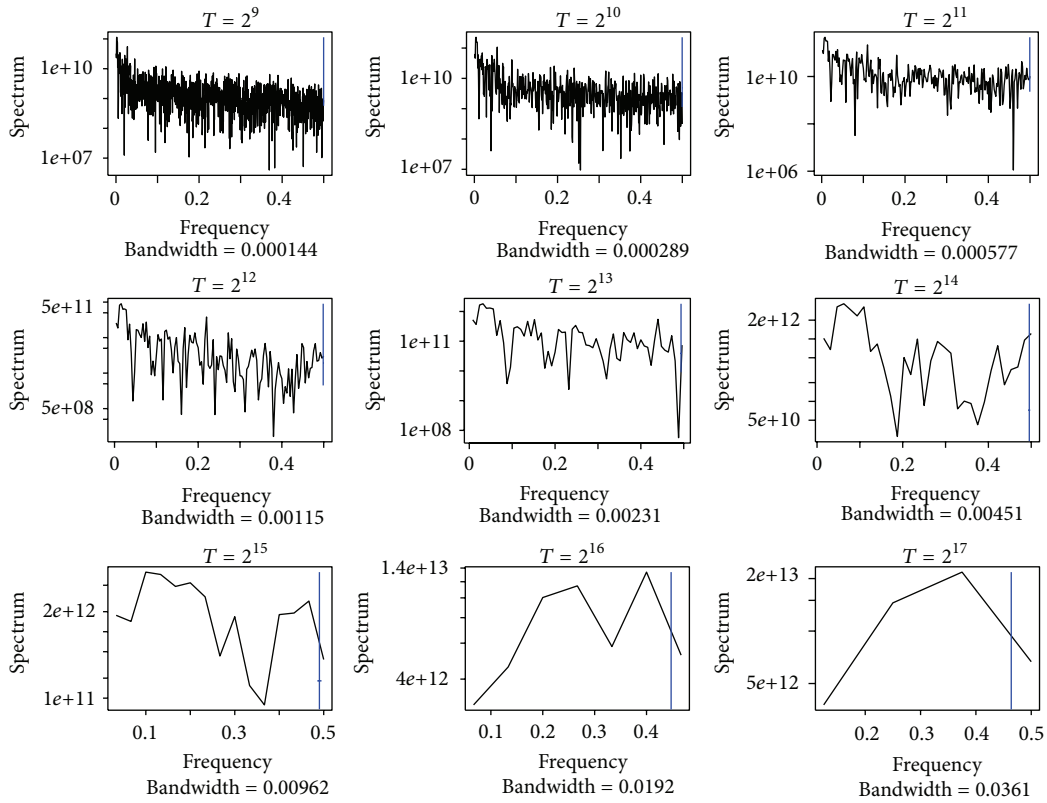


FIGURE 6: Spectrum of the accumulated pOct.TL data under different time scales.

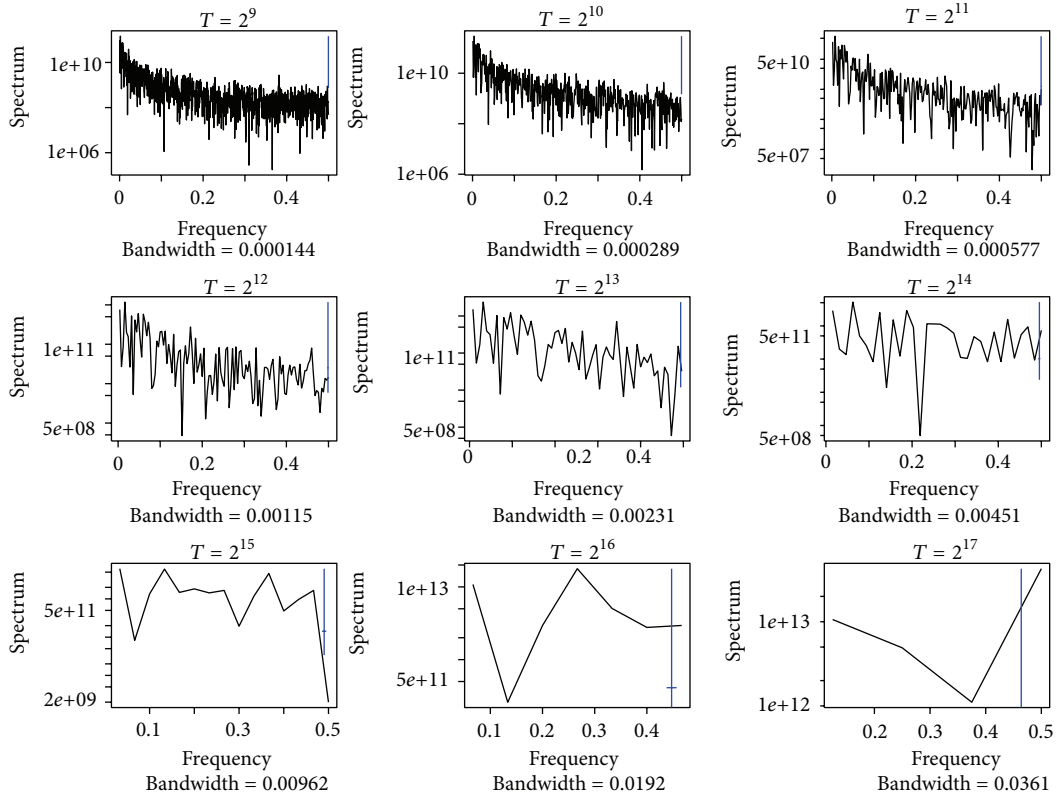


FIGURE 7: Spectrum of the accumulated Octxt.TL data under different time scales.

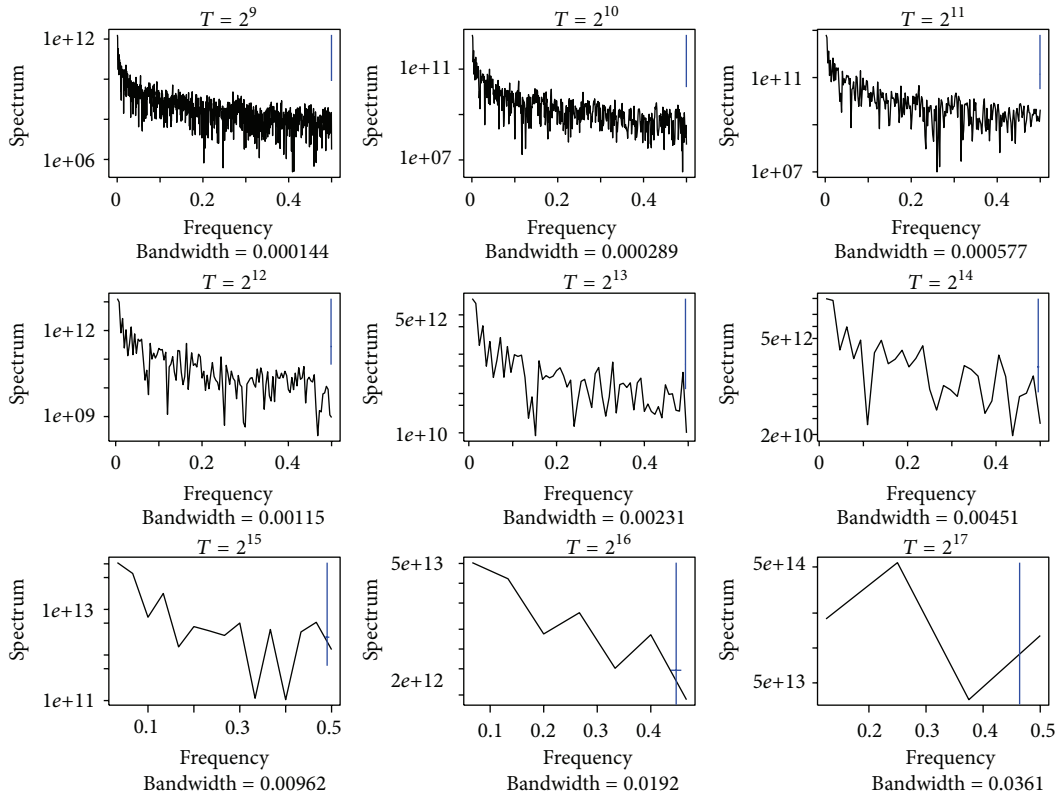


FIGURE 8: Spectrum of the accumulated OctExt4.TL data under different time scales.

TABLE 2: Test results for pAug.TL series.

| Test | KPSS | ADF | PP | ERS |
|--------------|---------|---------------|---------------|---------------|
| $T = 512$ | <0.01 | $<2.2e - 16$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 1024$ | <0.01 | $<2.2e - 16$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 2048$ | 0.01531 | $6.622e - 14$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 4096$ | 0.02589 | $8.969e - 08$ | $<2.2e - 16$ | $1.736e - 13$ |
| $T = 8192$ | 0.05159 | 0.1222 | $1.076e - 11$ | $6.255e - 07$ |
| $T = 16384$ | 0.07265 | 0.03636 | 0.003181 | $1.563e - 05$ |
| $T = 32768$ | >0.1 | 0.02883 | 0.0157 | 0.1148 |
| $T = 65536$ | >0.1 | 0.9794 | 0.007272 | 0.3533 |
| $T = 131072$ | >0.1 | 0.9433 | 0.8024 | NA |

TABLE 3: Test results for pOct.TL series.

| Test | KPSS | ADF | PP | ERS |
|--------------|---------|---------------|---------------|---------------|
| $T = 512$ | <0.01 | $<2.2e - 16$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 1024$ | <0.01 | $<2.2e - 16$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 2048$ | <0.01 | $7.589e - 15$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 4096$ | <0.01 | 0.0002329 | $<2.2e - 16$ | $2.749e - 09$ |
| $T = 8192$ | <0.01 | 0.005249 | $<2.2e - 16$ | $1.107e - 05$ |
| $T = 16384$ | <0.01 | 0.07512 | $1.609e - 10$ | 0.01228 |
| $T = 32768$ | <0.01 | 0.294 | $8.265e - 05$ | 0.1865 |
| $T = 65536$ | <0.01 | 0.197 | 0.03797 | 0.7795 |
| $T = 131072$ | 0.01835 | 0.4773 | 0.07518 | NA |

TABLE 4: Test results for OctExt.TL series.

| Test | KPSS | ADF | PP | ERS |
|--------------|---------|---------------|---------------|---------------|
| $T = 512$ | <0.01 | $<2.2e - 16$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 1024$ | 0.01112 | $3.262e - 11$ | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 2048$ | 0.02846 | $3.615e - 05$ | $<2.2e - 16$ | $3.718e - 13$ |
| $T = 4096$ | 0.0487 | 0.000314 | $<2.2e - 16$ | $1.216e - 08$ |
| $T = 8192$ | 0.08652 | 0.04858 | $5.691e - 11$ | $1.65e - 05$ |
| $T = 16384$ | >0.1 | 0.01177 | 0.01622 | 0.0006824 |
| $T = 32768$ | >0.1 | 0.1303 | 0.1326 | 0.0247 |
| $T = 65536$ | >0.1 | 0.2777 | 0.6618 | 0.1359 |
| $T = 131072$ | >0.1 | 0.2846 | 0.8794 | NA |

TABLE 5: Test results for OctExt4.TL series.

| Test | KPSS | ADF | PP | ERS |
|--------------|---------|-----------|---------------|---------------|
| $T = 512$ | <0.01 | 0.0001743 | $<2.2e - 16$ | $<2.2e - 16$ |
| $T = 1024$ | <0.01 | 0.001538 | $<2.2e - 16$ | $6.507e - 11$ |
| $T = 2048$ | <0.01 | 0.05514 | $<2.2e - 16$ | $3.832e - 05$ |
| $T = 4096$ | 0.01045 | 0.1265 | $<2.2e - 16$ | 0.000242 |
| $T = 8192$ | 0.03833 | 0.1103 | $<2.2e - 16$ | 0.04042 |
| $T = 16384$ | 0.07791 | 0.3414 | $1.6e - 14$ | 0.4146 |
| $T = 32768$ | >0.1 | 0.5884 | $1.004e - 06$ | 0.3668 |
| $T = 65536$ | >0.1 | 0.885 | 0.008174 | 0.314 |
| $T = 131072$ | >0.1 | 0.8827 | 0.6979 | NA |

which is on the contrast with the intuitive observation from the trajectory figures. Moreover, given the significance level at 1%, when the time scale is small, the accumulated traffic

could be judged as non-stationary and possesses a unit root. Specifically, we have the following findings.

- (i) The accumulated traffic of pOct.TL series could be regarded as stationary until the time scale reaches 17, while the other three series cannot be rejected to be stationary once the time scale is greater than 15. Thus, among the four series, the pOct.TL series seems to possess the least possibility to be stationary under each time scale.
- (ii) Considering the series of pAug.TL, pOct.TL, OctExt.TL, and OctExt4.TL, it seems that their stationarity behaviors are quite similar and hard to distinguish from each other under the corresponding time scales.
- (iii) For all these four series, when the time scale is small, the accumulated traffic has a unit root and exhibits at the same time some long range dependence behavior, which indicates after first order difference we could obtain a stationary process.

The previous discussions are for the Ethernet traffic but the methods may also be a reference for other types of time series, for instance, those in [32–35]. Our future research will work on the general description of network traffic rather than the Ethernet one alone.

5. Conclusions

In this paper, we have carried out several widely used tests of stationarity in order to study the stationary property of the accumulated Ethernet series at different time scales. The quantitative results reveal that when the time scale increases, the investigated accumulated Ethernet traffic is more likely to be stationary, which coincides with the normality investigation results for the same series and provide a useful empirical evidence for the traffic data modeling under large time scale.

Acknowledgments

This work was in part supported by the 973 Plan under Project Grant no. 2011CB302800, the National Natural Science Foundation of China under Project Grant nos. 11101158, 61272402, 61070214, and 60873264, and “the Fundamental Research Funds for the Central Universities”. The authors acknowledge W. Willinger, W. Leland, and D. Wilson with Bellcore, Morristown, who provided the data in this research.

References

- [1] R. M. Loynes, “On the concept of the spectrum for non-stationary processes,” *Journal of the Royal Statistical Society Series B*, vol. 30, pp. 1–30, 1968.
- [2] M. B. Priestley, “Evolutionary spectra and non-stationary processes,” *Journal of the Royal Statistical Society Series B*, vol. 27, pp. 204–237, 1965.
- [3] Z. Guo, L. G. Durand, and H. C. Lee, “Time-frequency distributions of nonstationary signals based on a Bessel kernel,” *IEEE*

- Transactions on Signal Processing*, vol. 42, no. 7, pp. 1700–1706, 1994.
- [4] N. E. Huang, Z. Shen, S. R. Long et al., “The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis,” *Proceedings of the Royal Society A*, vol. 454, no. 1971, pp. 903–995, 1998.
 - [5] A. R. Messina and V. Vittal, “Nonlinear, non-stationary analysis of interarea oscillations via Hilbert spectral analysis,” *IEEE Transactions on Power Systems*, vol. 21, no. 3, pp. 1234–1241, 2006.
 - [6] C. R. Pinnegar and L. Mansinha, “Time-local spectral analysis for nonstationary time series: the s-transform for noisy signals,” *Fluctuation and Noise Letters*, vol. 3, no. 3, pp. L357–L364, 2003.
 - [7] J. Xiao and P. Flandrin, “Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement,” *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2851–2860, 2007.
 - [8] M. Li and W. Zhao, “On 1/f noise,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 673648, 23 pages, 2012.
 - [9] Z. Psaradakis, “Blockwise bootstrap testing for stationarity,” *Statistics and Probability Letters*, vol. 76, no. 6, pp. 562–570, 2006.
 - [10] S. Ling, “Estimation and testing stationarity for double-autoregressive models,” *Journal of the Royal Statistical Society Series B*, vol. 66, no. 1, pp. 63–78, 2004.
 - [11] P. M. M. Rodrigues and A. Rubia, “A note on testing for non-stationarity in autoregressive processes with level dependent conditional heteroskedasticity,” *Statistical Papers*, vol. 49, no. 3, pp. 581–593, 2008.
 - [12] R. Von Sachs and M. H. Neumann, “A wavelet-based test for stationarity,” *Journal of Time Series Analysis*, vol. 21, no. 5, pp. 597–613, 2000.
 - [13] P. Abry and D. Veitch, “Wavelet analysis of long-range-dependent traffic patrice abry and darryl veitch,” *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 2–15, 1998.
 - [14] M. Grossglauser and J. C. Bolot, “On the relevance of long-range dependence in network traffic,” *IEEE/ACM Transactions on Networking*, vol. 7, no. 5, pp. 629–640, 1999.
 - [15] B. B. Mandelbrot, “Note on the definition and the stationarity of fractional Gaussian noise,” *Journal of Hydrology*, vol. 30, no. 4, pp. 407–409, 1976.
 - [16] M. Li and W. Zhao, “Quantitatively investigating locally weak stationarity of modified multifractional Gaussian noise,” *Physica A*, vol. 391, no. 24, pp. 6268–6278, 2012.
 - [17] B. B. Mandelbrot, *Multifractals and 1/f Noise*, Springer, New York, NY, USA, 1999.
 - [18] <http://ita.ee.lbl.gov/html/contrib/BC.html>.
 - [19] Z. Lu, M. Li, and W. Zhao, “Normality of Ethernet traffic at large time scales,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 471963, 2013.
 - [20] D. A. Dickey and W. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 427–431, 1979.
 - [21] S. E. Said and D. A. Dickey, “Testing for unit roots in autoregressive-moving average models of unknown order,” *Biometrika*, vol. 71, no. 3, pp. 599–607, 1984.
 - [22] P. C. B. Phillips and P. Perron, “Testing for a unit root in time series regression,” *Biometrika*, vol. 75, no. 2, pp. 335–346, 1988.
 - [23] G. Elliott, T. J. Rothenberg, and J. H. Stock, “Efficient tests for an autoregressive unit root,” *Econometrica*, vol. 64, no. 4, pp. 813–836, 1996.
 - [24] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, “Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?” *Journal of Econometrics*, vol. 54, no. 1–3, pp. 159–178, 1992.
 - [25] P. J. Brockwell and R. A. Davis, “Nonstationary and seasonal time series models,” in *Introduction to Time Series and Forecasting*, pp. 179–196, Springer, New York, NY, USA, 2002.
 - [26] C. B. Foster, *Economics and Business Forecasting—Lectures 4, 5 and 6*, Lecture notes-Ec 178, University of California, San Diego, Calif, USA, 2005.
 - [27] M. Lanne and P. Saikkonen, “Reducing size distortions of parametric stationarity tests,” *Journal of Time Series Analysis*, vol. 24, no. 4, pp. 423–439, 2003.
 - [28] R. Refinetti, “Non-stationary time series and the robustness of circadian rhythms,” *Journal of Theoretical Biology*, vol. 227, no. 4, pp. 571–581, 2004.
 - [29] E. Zivot, “Economic Theory II: Time Series Analysis—Unit Root Tests,” Lecture Notes-Econ 584. University of Washington, 2005. <http://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf>.
 - [30] P. M. Robinson, “Efficient tests of nonstationary hypotheses,” *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1420–1437, 1994.
 - [31] L. Ferrara, D. Guegan, and Z. Lu, “Testing fractional order of long memory processes: a monte carlo study,” *Communications in Statistics*, vol. 39, no. 4, pp. 795–806, 2010.
 - [32] E. G. Bakhoun and C. Toma, “Specific mathematical aspects of dynamics generated by coherence functions,” *Mathematical Problems in Engineering*, vol. 2011, Article ID 436198, 10 pages, 2011.
 - [33] C. Cattani, “On the existence of wavelet symmetries in archaea DNA,” *Computational and Mathematical Methods in Medicine*, vol. 2012, Article ID 673934, 21 pages, 2012.
 - [34] C. Cattani, G. Pierro, and G. Altieri, “Entropy and multifractality for the myeloma multiple TET 2 gene,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 193761, 14 pages, 2012.
 - [35] C. Toma, “Advanced signal processing and command synthesis for memory-limited complex systems,” *Mathematical Problems in Engineering*, vol. 2012, Article ID 927821, 13 pages, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

