

A Practical Method for Weak Stationarity Test of Network Traffic with Long-Range Dependence

Ming Li¹, Yun-Yun Zhang², and Wei Zhao³

Abstract— Testing the stationarity of real traffic remains a problem worth studying. Due to the importance of traffic theory in the Internet, to find a solution to such a problem brooks no delay. This paper presents a way to do the weak stationarity test of traffic with long-range dependence (LRD) as a single history traffic series of finite length. How to apply this method to real traffic on a packet-by-packet basis is demonstrated.

Keywords— Network traffic; Long-Range Dependence; Stationarity Test; Autocorrelation function.

I. INTRODUCTION

WE start with the concept of weak stationarity (stationarity for short) in mathematics and the problem statements regarding stationarity test of traffic in the Internet. Let $\{y_l(t)\}$ ($-\infty < t < \infty$) be a random process, where $y_l(t)$ is a sample function. In practice, a sample function (or some time series of finite length from a sample function) may be thought as the observed result of a single experiment. For a positive integer L and any fixed times, $t_1, t_2, \dots, t_L, y_l(t_1), y_l(t_2), \dots, y_l(t_L)$ stands for L random variables over the index l .

Let E be the mean operator. Then,

$$\mu_y(t) = E[y_l(t)]$$

is the ensemble mean at arbitrary fixed values of t and

$$r(t_1, t_2) = E[y_l(t_1)y_l(t_2)]$$

is called autocorrelation function (autocorrelation for short).

Usually, $\mu_y(t)$ and $r(t_1, t_2)$ are dependent on time. For a weakly stationary (stationary for short below) process, $\mu_y(t)$ and $r(t_1, t_2)$ are independent of time such that

$$\mu_y(t) = \text{const.}$$

and

$$r(t_1, t_2) = E[y_l(t + \tau)y_l(t)] = r(\tau),$$

Manuscript received July 2, 2007; Revised version received October 10, 2007. This work was supported in part by the National Natural Science Foundation of China under the project grant number 60573125. Wei Zhao's work was also partially supported by the NSF (USA) under Contracts 0808419, 0324988, 0721571, and 0329181. Any opinions, findings, conclusions, and/or recommendations in this paper, either expressed or implied, are those of the authors and do not necessarily reflect the views of the agencies listed above.

¹Ming Li (corresponding author) is with the School of Information Science & Technology, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, PR. China. (Tel.: +86-21-5434 5193; fax: +86-21-5434 5119; e-mails: ming_lihk@yahoo.com, mli@ee.ecnu.edu.cn).

²Yun-Yun Zhang is with the Department of Statistics, East China Normal University, No. 500, Dong-Chuan Road, Shanghai 200241, PR. China.

³Wei Zhao is with Rensselaer Polytechnic Institute, 110 Eighth Street, 1C05 Science, Troy, NY 12180-3590, USA. (e-mail: zhaow3@rpi.edu).

where $\tau = t_1 - t_2$ is time lag [1-3].

In the field of traffic engineering, one may meet the difficulty to directly apply the definition of stationarity in mathematics to test the stationarity of traffic with long-range dependence (LRD) because traffic is such a fractal time series that its mean does not exist, see e.g. [16]. Therefore, the mean of traffic cannot be used as a quantity to test the stationarity. We use the following two statements to further interpret the difficulty of the stationarity test of traffic with LRD.

Problem statement 1 [4, Sentence 1, Paragraph 3, Section 1, pp. 629]. "Unfortunately, it is not possible to tell with certainty whether or not a realization is stationary from its observation."

Problem statement 2 [5, Sentence 1, Section 3, pp. 7]. "The testing of the stationarity hypothesis is particularly difficult in the presence of LRD, where many classical statistical approaches cease to hold."

It is noted that stationarity test of traffic can also be regarded as an issue of nonstationarity test. Thus, the above statements imply that the reality of either the stationarity or the nonstationarity of traffic remains ambiguous though there are descriptions about the stationarity or nonstationarity of traffic, see e.g. [4-6,9,21,22]. The reality of stationarity (or nonstationarity) of real traffic to be processed greatly affects the research of the network of interest as stationary processes are substantially different from nonstationary ones in analysis and data processing [1,3]. For this reason, studying a method for the stationarity test of traffic is vital to traffic theory which plays an important role in the Internet in both theory and practice [7-10]. In this regard, [5] states particular test method specifically for LRD processes by investigating the time invariability of the Hurst parameter. However, the concept of stationarity does not relate to the statistical dependence of a process [1-3, 11]. Thus, this paper presents a general method that is irrelevant of the statistical dependence of a series.

Note that the difficulty of the testing of the stationarity hypothesis is mainly caused by the fact that the mean of traffic does not exist. However, one can use the autocorrelation to do the stationarity test from a view of second order processes.

In this research, we adopt the following expression of autocorrelation [2,3]:

$$r(t, t + \tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{l=1}^N y_l(t)y_l(t + \tau). \quad (1)$$

A key point we should pay attention to is that any real-traffic series to be studied is of single history and of finite length.

Hence, we first focus on the stationarity test regarding a single history traffic series of finite length. Then, a demonstration with a real-traffic series is given as an application case. We address a method for the stationarity test of a single history traffic series with LRD for finite length in Section 2 and demonstrate case study in Section 3. Discussions are given in Section 4 and Section 5 concludes the paper.

II. METHODOLOGY

According to the concept of stationary processes, in the case that the mean of traffic does not exist, we shall do the stationarity test of traffic with LRD by checking if the autocorrelation of $\{y_l\}$ is time invaring or not. For a single history traffic series of finite length, we need completing three tasks as follows.

- 1) First, to construct a process of a single history traffic series of finite length.
- 2) Then, to describe a meaning of stationarity of a single history traffic series of finite length in practical terms.
- 3) Finally, to determine a measure to test the stationarity.

For completing the first task, we let $y(i)$ be a traffic series of N length and divide $y(i)$ into M non-overlapped intervals. Each interval is of L length. Denote

$$y_l(i) = y(i) \text{ for } i \in [(l-1)L, lL-1], l = 1, 2, \dots, M. \quad (2)$$

Then, $y_l(i)$ represents the l th sample series and $\{y_l\}$ a traffic process. In this case, the autocorrelation of $y_l(i)$ is estimated over an interval L by

$$r(k; l) = \frac{1}{L} \sum_{i=(l-1)L}^{lL-1} y_l(i)y_l(i+k). \quad (3)$$

Thus, $r(k; l)$ is a series of M length over the index l .

Usually,

$$r(k; l) \neq r(k; n) \text{ for } l \neq n.$$

However, this may not mean that $\{y_l\}$ is nonstationary because each y_l is of finite length, the number of samples is finite and there are errors in measurement or computation. In practice, therefore, we say that a single history traffic series of finite length is referred to as being stationary if $r(k; l)$ defined by (3) does not vary significantly as l changes. Here, that $r(k; l)$ does not vary significantly implies that $r(k; l)$ and $r(k; n)$ are similar in the sense of pattern matching according to a certain rule in pattern recognition for all l and n . The above completes the second task.

To consider the third task, we introduce the correlation matrix of sample autocorrelations. Without losing generality, only normalized autocorrelations are considered here and below.

As known, correlation matching is a commonly used tool in pattern matching [12,13,15]. Let

$$\text{corr}[r_l, r_n] = c_{ln}$$

be the correlation coefficient between two sample autocorrelations r_l and r_n . Then,

$$C = [c_{ln}] = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1M} \\ c_{21} & c_{22} & \dots & c_{2M} \\ \dots & \dots & \dots & \dots \\ c_{M1} & c_{M2} & \dots & c_{MM} \end{bmatrix} \quad (4)$$

is an $M \times M$ matrix. For matrix C , $|c_{ln}| = 1$ if $l = n$ and $|c_{ln}| < 1$ for $l \neq n$.

Let γ be the threshold regarding pattern similarity. Then, if $|c_{ln}| \geq \gamma$ for all l and n , we say r_l does not vary significantly as l varies in the sense of $|c_{ln}| \geq \gamma$. In practical terms, the patterns of r_l and r_n are reasonably similar in many cases for $\gamma = 0.8$ and quite satisfactorily similar for $\gamma = 0.85$ [13]. As a matter of fact, from a view of engineering, $\gamma = 0.7$ is an acceptable threshold, see e.g. [15]. This paper adopts $|c_{ln}| \geq 0.85$ as a similarity threshold for the following discussions, though, generally, from a view of engineering, it may be suitable to claim that a series is nonstationary if $\gamma < 0.7$.

Remark 1. If $|c_{ln}| \geq 0.85$ for all l and n , we say that the series is stationary in the sense of $|c_{ln}| \geq 0.85$. Otherwise, nonstationary in that sense.

Remark 2. To test whether $\{y_l\}$ is stationary or not is to check if $r(k; l)$ is time invaring according to Remark 1. This completes the third task.

III. CASE STUDY

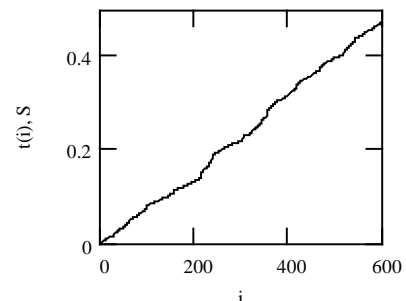
We use a real-traffic trace named DEC WRL-4, which can be obtained freely in the Internet Traffic Archive [14]. This trace contains 5.7 million packets for the time length of an hour. It was measured at Digital Equipment Corporation. The following shows two cases.

A. Interarrival Times

Let $t(i)$ be a timestamp series, indicating the timestamp of the i th packet on a packet-by-packet basis for $i = 0, 1, \dots$. Then, the timestamp increment series (also called interarrival times in communication networks) $s(i)$ is given by

$$s(i) = t(i+1) - t(i). \quad (5)$$

Fig. 1 shows $t(i)$ and $s(i)$ of that trace.



(a)

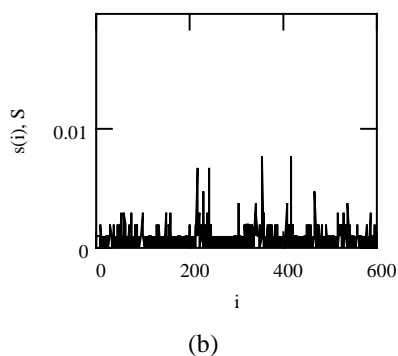


Fig. 1. Real traffic DEC WRL-4. (a) Timestamp series. (b) Timestamp increment series.

According to (3), the sample autocorrelations of $s(i)$ are given by

$$r_l(k) = \frac{1}{L} \sum_{i=(l-1)L}^{lL-1} s_l(i)s_l(i+k).$$

Ninety sample autocorrelations are computed for $s(i)$. Fig. 2 indicates the sample autocorrelations r_1 and r_{90} of $s(i)$. By computing correlation matrix, we obtain the correlation matrix for $s(i)$ as indicated in Fig. 3.

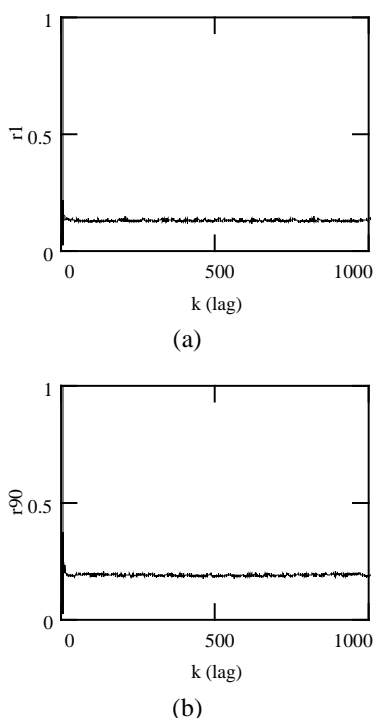


Fig. 2. Sample autocorrelations of timestamp increment series of DEC WRL-4. (a). The autocorrelation r_1 . (b). The autocorrelation r_{90} .

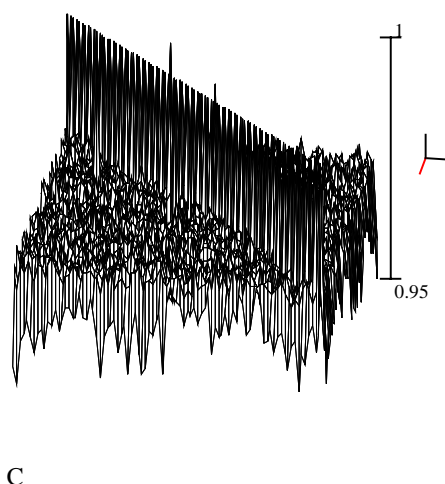


Fig. 3. Correlation matrix consisting of 90 sample autocorrelations of $s(i)$ of DEC WRL-4.

Fig. 3 indicates that the minimum value of the element in C is greater than or equal to 0.95, implying that $r_l(k)$ of $s(i)$ of DEC WRL-4 does not vary significantly as l changes. Thus, $s(i)$ of DEC WRL-4 is stationary.

It should be noted that interarrival traffic has the property of LRD [26]. By LRD, we mean that the following holds for an autocorrelation of a traffic series:

$$r(k) \sim ck^{-\beta} (k \rightarrow \infty), \beta \in (0, 1), \tag{6}$$

where $c > 0$ is a constant, \sim stands for asymptotical equivalence under the limit $k \rightarrow \infty$ [11]. Considering the autocorrelation $R(k)$ that is the mean of 90 sample autocorrelations and by using the lease-square curve fitting of the autocorrelation estimation, we attain.

$$R(k) \sim k^{-0.026} (k \rightarrow \infty). \tag{7}$$

The above expression clearly shows the LRD of that series as $R(k)$ of (7) is nonsummable.

B. Series in Packet Size

Let $x[t(i)]$ be a series, representing the packet size at time $t(i)$ on a packet-by-packet basis. Then, $x(i)$ stands for a series, indicating the packet size of the i th packet. Traffic series in packet size is of LRD, see e.g. [5-9,23,24,25].

Fig. 4 is the plot of the first 1024 points of $x(i)$ for DEC WRL-4. Fig. 5 is the correlation matrix consisting of 90 sample autocorrelations of $x(i)$. From Fig. 5, one sees that the minimum value of the element in C is greater than or equal to 0.912, meaning that $r_l(k)$ of $x(i)$ of DEC WRL-4 does not vary considerably as l changes. Thus, $x(i)$ of DEC WRL-4 is stationary.

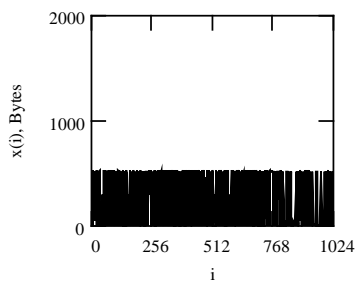
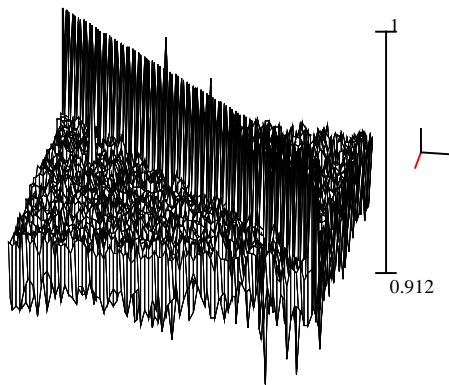


Fig. 4. Real traffic DEC WRL-4 in packet size.



C

Fig. 5. Correlation matrix consisting of 90 sample autocorrelations of $x(i)$ of DEC WRL-4.

IV. DISCUSSIONS

The research of traffic exhibits that traffic has the property of multifractal either on a point-by-point basis or on an interval-by-interval basis, see e.g. [17-20]. We note that the concept of the stationarity differs from that of multi-fractality. The former implies that $r(k; l)$ does not vary significantly with l while the latter means that the statistical parameter, such as the Hurst parameter H or fractal dimension D , is time varying. To be precise, time varying H or D of the investigated traffic series may not imply that series is nonstationary unless $r(k; l)$ changes considerably with l .

Note that the meaning of ‘considerably’ or ‘significantly’ is characterized by the correlation threshold. The selection of the correlation threshold γ , however, relies on the individual researchers, usually ranging from 0.7 to 0.9 or greater than 0.9 in engineering. Thus, high value of γ , e.g., $\gamma = 0.95$, implies strict classification. Nevertheless, over-high value of γ may be inappropriate when one takes into account measurement errors or computation errors, such as errors in autocorrelation estimation, and or uncertainties. On the other side, low value of γ , e.g., $\gamma = 0.65$, may yield rough classification. We use a traffic trace named Lbl-pkt-5 to explain this further. That trace was recorded at the Lawrence Berkeley Laboratory [14]. It contains 710614 packets.

Let $x(i)$ be a series, representing the size of the i th packet of

Lbl-pkt-5. Fig. 6 is the plot of $x(i)$. Due to relatively short sample size, we only section $x(i)$ into 10 non-overlapped sections. Thus, the correlation matrix for Lbl-pkt-5 is 10×10 , see Fig. 7.

The minimum element of C equals to 0.865. Therefore, $x(i)$ of Lbl-pkt-5 is stationary in the sense of $\gamma = 0.85$ but it may be taken as nonstationary if one sets $\gamma > 0.865$.

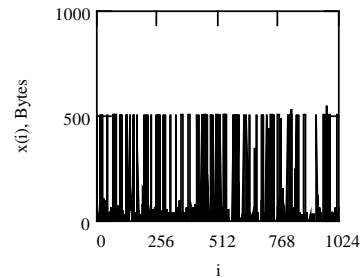
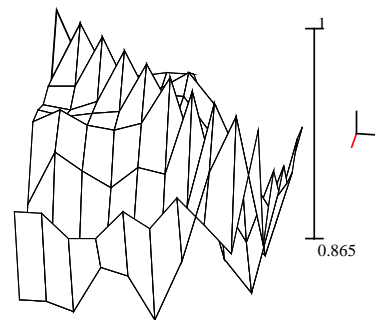


Fig. 6. Real traffic Lbl-pkt-5 in packet size.



C

Fig. 7. Correlation matrix consisting of 10 sample autocorrelations of $x(i)$ of Lbl-pkt-5.

V. CONCLUSION

We have discussed a method to do the weak stationarity test of traffic with LRD as a single history series of finite length.

REFERENCES

- [1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York, McGraw-Hill, 1984.
- [2] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, McGraw-Hill, 1961.
- [3] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedure*, 2nd Edition, John Wiley & Sons, 1991.
- [4] D. P. Heyman and T. V. Lakshman, "On the relevance of long-range dependence in network traffic," *IEEE/ACM Trans. Networking*, Vol. 7, No. 5, June, 1999, pp. 629-640.
- [5] P. Abry and D. Veitch, "Wavelet analysis of long-range dependent traffic," *IEEE Trans. Information Theory*, Vol. 44, No. 1, Jan. 1998, pp. 2-15.
- [6] W. Willinger and V. Paxson, "Where mathematics meets the Internet," *Notices of the American Mathematical Society*, Vol. 45, No. 8, Aug. 1998, pp. 961-970.
- [7] H. Michiel and K. Laevens, "Teletraffic engineering in a broad-band era," *Proc.the IEEE*, Vol. 85, No. 12, Dec., 1997, pp. 2007-2033.
- [8] B. Tsybakov and N. D. Georganas, "Self-similar processes in communications networks," *IEEE Trans. Information Theory*, Vol. 44, No. 5, Sep. 1998, pp. 1713-1725.

- [9] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. Networking*, Vol. 3, No. 3, June 1995, pp. 226-244.
- [10] S. Floyd and V. Paxson, "Difficulties in simulating the Internet," *IEEE/ACM Trans. Networking*, Vol. 9, No. 4, Aug. 2001, pp. 392-403.
- [11] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, 1994.
- [12] K. S. Fu, Editor, *Digital Pattern Recognition*, Springer, 1976.
- [13] M. Li, et al., "An on-line correction technique of random loading with a real-time signal processor for a laboratory fatigue test," *J. Testing and Evaluation*, Vol. 28, No. 5, Sep. 2000, pp. 409-414.
- [14] <http://www.acm.org/sigcomm/ITA/>
- [15] J. L. Talmaon, *Pattern Recognition of the ECG: a Structured Analysis*, Ph.D. Thesis, 1983.
- [16] W. Willinger, V. Paxson, R. H. Riedi, and M. S. Taquu, Long-Range Dependence and Data Network Traffic, in *Long-range Dependence: Theory and Applications*, P. Doukhan, G. Oppenheim and M. S. Taquu, eds., Birkhauser, 2002.
- [17] M. S. Taquu, V. Teverovsky, and W. Willinger, "Is network traffic self-similar or multifractal?" *Fractals*, Vol. 5, 1997, pp. 63-73.
- [18] M. Li, "Change trend of averaged hurst parameter of traffic under ddos flood attacks," *Computers & Security*, Vol. 25, No. 3, May 2006, pp. 213-220.
- [19] M. Li, S. C. Lim, and Huamin Feng, A novel description of multifractal phenomenon of network traffic based on generalized Cauchy process, *Springer LNCS 4489*, May 2007, pp. 1-9.
- [20] M. Li, S. C. Lim, B.-J. Hu, and Huamin Feng, "Towards describing multi-fractality of traffic using local hurst function," *Springer LNCS 4488*, May 2007, pp. 1012-1020.
- [21] C. Liu, J. Yang, and J. Yang, "A nonstationary traffic train model for fine scale inference from coarse scale counts," *IEEE J. Selected Areas in Communications*, Vol. 21, No. 6, 2003, pp. 895-907.
- [22] D. Rincón and S. Sallent, "On-line segmentation of non-stationary fractal network traffic with wavelet transforms and log-likelihood-based statistics," *Springer LNCS 3375*, 2005, pp. 110-123.
- [23] M. Li, S. C. Lim, and W. Zhao, Long-Range Dependent Network Traffic: A View from Generalized Cauchy Process, in *Progress in Applied Mathematical Modeling*, edited by F. Yang and et al., Nova Science Publishers, USA, 321-338, to be published in 2008 1st Quarter.
- [24] M. Li, and et al., "Modeling autocorrelation functions of self-similar teletraffic in communication networks based on optimal approximation in Hilbert space," *Applied Mathematical Modelling*, Vol. 27, No. 3, 2003, pp. 155-168.
- [25] M. Li, "Modeling autocorrelation functions of long-range dependent teletraffic series based on optimal approximation in Hilbert space-a further study," *Applied Mathematical Modelling*, Vol. 31, No. 3, Mar. 2007, pp. 625-631.
- [26] M. Li, W. Jia, and W. Zhao, "Correlation form of timestamp increment sequences of self-similar traffic on Ethernet," *Electronics Letters*, Vol. 36, No. 19, Sep. 2000, pp. 1168-1169.
- [27] R. L. Cruz, "A calculus for network delay, part i: network elements in isolation," *IEEE Trans. Inform. Theory*, Vol. 37, No. 1, Jan. 1991, pp. 114-131.
- [28] K. Pandit, J. Schmitt, and Ralf Steinmetz, "Network calculus meets queueing theory-a simulation based approach to bounded queues," *12th IEEE International Workshop on Quality of Service (IWQoS'04)*, June 7-9, 2004, Montreal, Canada.
- [29] M. Li and W. Zhao, "A statistical model for detecting abnormality in static-priority scheduling networks with differentiated services," *CIS 2005*, Springer LNAI 3802, Dec. 2005, 267-272.

researcher in CSSRC. From 1995-1999, he was with the Automation Department, Wuxi University of Light Industry. From 2002 to 2004, he was with the School of Computing, National University of Singapore.

In 2004, he joined East China Normal University (ECNU) as a professor in both electronics engineering and computer science. He is currently a Division Head for Communications & Information Systems at ECNU. His research areas relate to applied statistics, computer science, measurement & control. He has published over 70 papers in international journals and international conferences in those areas. Dr. Li is a member of IEEE.



Yun-Yun Zhang is a bachelor candidate from the Department of Statistics, East China Normal University.



Wei Zhao is a professor of computer science and the dean for the School of Science at Rensselaer Polytechnic Institute. His research interests include distributed computing, real-time systems, computer networks, and cyberspace security. Zhao received a PhD in computer and information sciences from the University of Massachusetts, Amherst. He is a Fellow of the IEEE. He has published over 280 papers in international journals, international conferences, and book chapters.



Ming Li was born in 1955 in Wuxi, China. He completed his undergraduate program in electronics engineering at Tsinghua University. He received the M.S. degree in ship structural mechanics from China Ship Scientific Research Center (CSSRC) and Ph.D. degree in computer science from City University of Hong Kong, respectively. From 1990 to 1995, he was a