

Modeling Spectral Features in TCP Traffic

Gbenga Olowoyeye , Bo-Kyoung Kim and Kavitha Chandra
Center for Advanced Computation and Telecommunications
Department of Electrical and Computer Engineering
University of Massachusetts Lowell
Lowell, MA 01854
Email: golowoye@cs.uml.edu, Kavitha_Chandra@uml.edu
Tel: 978-934-3356 Fax: 978-458-8289

ABSTRACT

The analysis of traffic from wide and local area network sites show that cyclical trends are a characteristic feature in transport control protocol (TCP) traffic. The pattern of the estimates of conditional expectations indicate that cyclical behavior is restricted to a set of amplitude regimes in the traffic rate process. The application of non-linear processes to model the aforementioned features is investigated. A non-linear time series model captures the traffic features with a reduced number of parameters relative to a linear model and is more accurate in its reconstruction of the autocorrelation estimates. The control mechanism in TCP and stochastic properties of the ack round-trip-times can play a fundamental role in generating cyclical trends in TCP traffic. A simulation based analysis of TCP is presented that demonstrates how the control mechanism in TCP can induce periodic components into a random application layer traffic stream.

1. INTRODUCTION

Traffic measurement studies ^{[1] [2] [3]} have amply demonstrated that data traffic is characterized by complex traffic patterns that cannot be explained on the basis of simple stochastic models and independent random variables. The existing demand for provision of quality of service guarantees ^[4] for Internet connections, motivates looking into methodologies for modeling and forecasting traffic on the Internet. The temporal correlation exhibited by data traffic is a primary feature that influences the performance of queueing systems ^{[5] [6]}. The causality of traffic correlation and methods for incorporating this feature into a traffic model are therefore important considerations.

Previous studies on traffic measurements have considered the characterization of aggregate traffic as measured at an outgoing or incoming link of a router connected to the Internet. Leland et. al. [3] analyzed aggregate Ethernet traffic and showed that the traffic exhibits correlation over time scales ranging from seconds to hours. This absence of reduction of traffic structure under smoothing was found to be characteristic of features represented by a self-similar process. Paxson and Floyd [2] in their analysis of wide-area network traffic observed that a few file transfer protocol (FTP) connections characterized by the

largest bursts were the dominant byte rate contributors to the aggregate traffic. The reasons for persistence of correlation over long time scales and the influence of a subset of connections on the aggregate process may be better understood if the traffic is first analyzed in terms of its individual components.

In this paper, we examine the aforementioned issues by considering the characteristics of component processes that make up the aggregate traffic. The protocol type and packet size ranges are found to be useful traffic features for decomposition of the aggregate process. We find that the traffic correlation is primarily due to TCP traffic which represents about 48% of the traffic mix. The traffic measurements considered are from two network sites, one representing traffic from a wide-area network and the other being the Bellcore Ethernet traffic analyzed in [3]. Section 2.0 discusses the traffic measurements and methods used to disaggregate the traffic into separate processes. The spectral and time domain characteristics of TCP traffic are discussed and a hypothesis is proposed for the observations. Section 3.0 describes the fitting and validation of a threshold autoregressive process as a traffic model. In Section 4.0 a hypothesis for TCP traffic patterns is proposed. Section 5.0 concludes the paper.

2. TRAFFIC MEASUREMENTS AND ANALYSIS METHODS

We consider traffic measurements made on two different networks. The first set referred to as *WAN-I* represents traffic collected from the FDDI interface at FIX-West (Federal Internet Exchange) and MAE-West interconnection facility located at NASA-Ames Research Center in California on November 20, 1997. These facilities supports the exchange of traffic between national, regional, commercial and federal service providers. The traffic was collected at an OC-3 link between two gigaswitches on an FDDI LAN, that connect MAE-West and FIX-West facilities. The traffic collection was made using the OC3MON utility [<http://www.nlanr.net/NA/Oc3mon>]. This system uses an optical splitter to connect from an OC3 link to an ATM network interface card residing on a PC. The OC3MON software extracts headers from packet traces and applies timestamps. This data is obtained from the National Laboratory for Applied Networking Research (NLANR) public-domain archive at <http://moat.nlanr.net/Traces> . From the packet header traces of *WAN-I* , we are able to determine the time-stamp, the packet size, protocol type, source and destination addresses and port numbers.

The second set referred to as *LAN-I* represents Bellcore Ethernet traffic measurements made on the BellCore Lan in 1989. In this case only packet time stamps and packet sizes are available for analysis. We consider this trace primarily to show that the traffic measurements made in 1997 show remarkable

similarity in traffic features to the measurements made in 1989. In this time period what has changed significantly is the base of Internet users whereas the technology supporting network protocols, routing and switching operations have remained essentially unchanged. This suggests that the protocols may have a significant influence in shaping the patterns in Internet traffic.

We first consider the analysis of the more recent *WAN-1* data. A contiguous trace of four million packets is considered, which corresponds to a time duration of about 352 seconds. The average and peak arrival rates of this data set are approximately 37 and 96 Mbps respectively. Based on the protocol field information, it is determined that TCP, UDP (User Datagram) and IP (IP in IP encapsulation) protocols constitute 98% of the traffic mix. The inter-arrival times and packet sizes of these individual protocols show distinct characteristics. In particular, TCP traffic is found to be the dominant contributor to the byte rate of the arrival process. In Fig. 1(a), we depict the autocorrelation estimates (acf) for the aggregate process which is to be compared to the acf estimates for the TCP, UDP, IP/IP traffic and remaining traffic in Figs. 1(b-e). It is clear that at this time scale, TCP packets exhibit the strongest correlation pattern, and this is inherited by the aggregate process. Based on these observations we restrict further analysis to TCP based traffic.

2.1 Spectral and Amplitude Characteristics of TCP traffic

By considering the analysis of TCP packets, a further decomposition of the aggregate TCP process may be undertaken by considering additional traffic features such as source and destination addresses and port identifiers. However, the large number of connections, more than 28,000 in this data set alone can complicate our understanding of the traffic dynamics. Rather, we note that TCP flows may be characterized into basically two types of connections. These are connections where the source is primarily sending data segments (Ex. http, ftp servers) or connections where the source is primarily acknowledging data received (clients downloading from the Web or ftp servers). In the first case, one may expect the packet size distribution of individual connections to correspond to TCP segment sizes which can range from a minimum segment size of 552 bytes to the maximum frame size supported by the LAN on which the application resides, such as 1500 bytes for Ethernet networks. In the second case, the ack packets are typically 40 bytes in length. Fig. 2 shows the distribution of TCP packet sizes. The clustering of packets around 40, 552, 576 and 1500 bytes result from the aggregation of TCP flows from servers and clients.

A second level of traffic decomposition is proposed based on our hypothesis of flow types being

characterized by the packet size ranges. We partition the aggregate traffic into three classes representing the range of packet sizes [0:200], [200:2000] and [2000:maximum] bytes. These are referred to as LB_1 , MB_1 and HB_1 respectively. This level of coarseness in partition was found sufficient to distinguish between dynamics in the different ranges. One may expect that the characteristic time-scales of data and ack streams to be influenced by different phenomenon. For a server transmitting data, the rate of transmission is dictated by the ack round-trip-times (RTTs) for that connection and the corresponding behavior of the TCP congestion window. The ack streams would be similarly influenced, but by the transit delay time-scales over a separate link. These features are determined by examining the spectral characteristics in the individual processes. To gauge the spectral behavior at the highest resolution the sampling rate is chosen to be of the order of the average inter-arrival time, which is approximately one millisecond. Fig. 3 shows the estimates of the Fourier amplitudes of HB_1 in decibels relative to the dc amplitude. In Fig. 3, the horizontal axis represents the frequency in radians. The traffic corresponding to HB_1 shows the presence of a fundamental frequency of 11.93 hz and its higher harmonics. A characteristic time period of 0.08 seconds can therefore be attributed to this traffic class. This parameter determines the sampling or aggregation time scale that must be considered for casting the traffic in time-series form. We consider a time scale of 0.01 seconds that will be sufficient to resolve the dominant spectral amplitudes corresponding to the fundamental, the first and third harmonics. The spectral estimates for MB_1 although of weaker magnitudes also show the presence of low frequency harmonics with time periods ranging from 0.5 seconds to 42 seconds.

Based on the choice of $\delta t = 0.01$ as the traffic aggregation time scale for all three traffic classes, the time domain features of the three traffic classes are considered next. The autocorrelation estimates of LB_1 , MB_1 and HB_1 are shown in Figs. 4(a-c). The dominant correlation structure of the aggregate TCP process may now be attributed to a class of packet sizes belonging to HB_1 , which comprises only about 2% of the entire set of TCP packets. Based on the observations of the structure in the acf estimates, linear time-series models may be determined for each of these traffic classes. Time-series models for network traffic using autoregressive moving average (ARMA) processes have been applied in Basu et. al. [7]. First the verification of linearity assumptions in the traffic amplitudes must be conducted. The sample acfs represent linear dependence in the time-series data. To test for nonlinear behavior, conditional statistics as given by sample regression functions [8] will be considered. The lag j regression function is defined as

$$r_j = E[X_n | X_{n-j} = x] \quad (1)$$

When the random variables X are Gaussian distributed, this regression function is always linear and is determined by the first two moments. Examples of estimates of r_j for various values of the lag parameter for the LB_1 , MB_1 and HB_1 are depicted in Fig. 5(a-c). The horizontal axis represents a partition of the amplitude range of the time series into a finite set of disjoint sets. The vertical axis represents the function r_j determined by calculating the average of all samples that satisfy the regression constraint. For the LB_1 and MB_1 data sequence, the linear trends are dominant in most of amplitude regime. For HB_1 in Fig. 5(a) clear departures from linearity are observed for lags 1, 7 and 8, which are representative of the lags at which the acf estimates are dominant. The pattern of the regression functions indicate that non-linear models must be considered to incorporate the amplitude dependent features in the traffic. The non-linearity in the HB_1 data may be modeled using piece-wise linear models over appropriately selected amplitude regimes. This can be carried out using the threshold autoregressive process (AR) proposed by Tong,^[9] as the basic model. Since this model incorporates linear autoregressive models within specified amplitude regimes, the traffic models for the three classes LB_1 , MB_1 and HB_1 may be integrated into a single traffic model. A discussion of modeling procedure is considered next.

3. THRESHOLD AUTOREGRESSIVE MODELS

The threshold autoregressive (TAR) model proposed by Tong [9], is a nonlinear model comprised of linear AR models which are valid in disjoint subregions in amplitude. At a given time the subregion selected will depend on the amplitudes observed over lagged time values. The TAR models and its variants have been successfully applied for modeling time-series data exhibiting cyclical properties and long-range-dependence properties^[10].

The modeling will be carried out on the aggregate TCP traffic with basic structure being dictated by the knowledge of the features in the LB_1 , MB_1 and HB_1 data classes. The model considered here will incorporate two amplitude ranges, one representing the classes LB_1 and MB_1 and another representing the class HB_1 . They will be denoted as low (L) and high (H) amplitude states. The choice of the appropriate threshold \hat{r} between these states will be determined during the model fitting. In the low-state the times-series takes on values $L: (0, \hat{r}]$, and the high state accommodates amplitudes $H: [\hat{r}, \infty)$. In addition a delay value d will be used in the conditional switching between the low and high states. The use of one delay value and one threshold will result in two amplitude subregions, labeled as R_j where the

index j takes on values of one and two. The TAR model allows one to change the parameters of the AR process over time by virtue of the switching rule. This is done based on the amplitude of the time-series at delayed values. Within each of the subregions R_j the process evolves as a stable AR process, governed by the correlations within that region.

The current value of the byte-rate at time n will be governed by an autoregressive process of order k_j .

$$x(n) = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)} x(n-i) + e^j(n) \quad x(n-d) \in R_j \quad (2)$$

Here the term $e^j(n)$ represents samples derived from an independent identically distributed random process having zero mean and finite variance. When subregion constraints are violated, the process is switched to the subregion model that obeys the proper amplitude and delay constraints. Extended sojourns in the high byte-rate state is an important feature in performance management, whereas dwell-time in the low byte-rate state has impact on multiplexing efficiency. Therefore the thresholds and delay parameters should be carefully chosen to capture the critical elements of the observed dynamics.

3.1 TAR Model Parameter Selection

To construct the model, the optimal values of \hat{r} , and the delay d must be selected along with the coefficients of the local AR processes for each subregion. For a given delay value and threshold parameters, to estimate the local AR parameters, the data is searched for all samples $x(n)$ that satisfy the given amplitude and delay constraints, R_j . For each case R_j , the samples $x_{i_1}^j, x_{i_2}^j, \dots, x_{i_{n_j}}^j$ of the time-series represent the n_j samples that satisfy constraint R_j . These n_j samples will be denoted by the vector, $\underline{x}^j = (x_{i_1}^j, \dots, x_{i_{n_j}}^j)^T$. For this data, the k_j^{th} order linear model coefficients are evaluated.

$$\underline{x}^j = A^j \underline{a}^j + \underline{e}^j \quad (3)$$

where $\underline{a}^j: (a_0^j, a_1^j, \dots, a_{k_j}^j)^T$ and A^j is a $n_j \times k_j + 1$ matrix comprised of the k_j values that lag the elements of \underline{x}^j .

$$A^j = \begin{bmatrix} 1 & x_{i_1-1}^j & x_{i_1-2}^j & \cdot & \cdot & \cdot & x_{i_1-k_j}^j \\ 1 & x_{i_2-1}^j & x_{i_2-2}^j & \cdot & \cdot & \cdot & x_{i_2-k_j}^j \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{i_{n_j}-1}^j & x_{i_{n_j}-2}^j & \cdot & \cdot & \cdot & x_{i_{n_j}-k_j}^j \end{bmatrix} \quad (4)$$

and $\underline{e}^j = (e_{i_1}^j, \dots, e_{j_{n_j}}^j)^T$ is the residual error vector. Since $n_j \gg k_j$, the solution of the system of equations in Eq. (4) is the least-squares estimate of \underline{a}^j and is denoted as $\hat{\underline{a}}^j$. The least-squares error,

$$\hat{\underline{e}}^j = \underline{x}^j - \underline{A}^j \hat{\underline{a}}^j \quad (5)$$

The error variance $\sigma_j^2 = \|\hat{\underline{e}}^j\|^2/n_j$ represents the approximate Maximum Likelihood Estimate of the noise variance for the j^{th} subregion.

In order to obtain the complete TAR model one must determine the optimal choice of the parameters. To do so, we will use the sum of the Akaike Information Criteria (AIC) ^[11] of the sub-models as the performance measure. The process of TAR parameter estimation is that outlined in Tong [9] and will be briefly described here.

For a given delay parameter d the threshold amplitude \hat{r} is sampled uniformly between a range of amplitudes in the region of the turning points identified through the regression function estimates. For the aggregate TCP data, this region extends from 20,000 to 50,000 bytes per unit time. The maximum model order of the AR model was set equal to 40. For each subregion R_j we determined the least-squares estimates of the local AR coefficient using Eqn. (4) for each order ranging from 1 to 40. In each case the residual errors $\hat{\underline{e}}^j$ determine the best model order for the subregion using the Akaike Information Criteria (AIC). The AIC is given by

$$AIC(k) = n_j \ln \left\{ \|\hat{\underline{e}}(k)\|^2/n_j \right\} + 2(k+1) \quad (6)$$

where n_j is the sample size of the fitted data and k is the candidate model order. The optimum order k_j for the j^{th} sub-model corresponds to the value k that yields the minimum value for the AIC statistic. We denote this as $AIC(k_j)$ for the j^{th} subregion.

This process is repeated for all subregions and for each value of \hat{r} . The total AIC as a function of

the threshold and delay parameters is computed as

$$AIC_{total}(d, \hat{r}) = \sum_{j=1}^2 AIC(k_j) \quad (7)$$

The optimal threshold parameter \hat{r} and AR parameters are selected to be those which yield the minimum $AIC_{total}(d, \hat{r})$. The optimal lag value was found to be equal to 8 and the optimal threshold was determined by a local minima in the AIC estimates in the range 32500 – 32600 bytes.

In generating the TAR model simulated traffic, only the initial conditions were determined from the measurements. Subsequent values were generated using the TAR parameter estimates and additive Gaussian noise variables for each subregion. A comparison of the acfs of the data and TAR model results is shown in Fig. 6. The TAR results were compared with those obtained by fitting a linear AR process. The required model order was $60 - th$, requiring a significantly higher order AR process representation than that required by the individual processes in the nonlinear case. In addition, the autocorrelation estimates were found to be less accurately modeled, with the negative correlations seen to be of much higher magnitude than that exhibited by the data. The lack of constraint on amplitude regimes results in the linear process taking on the characteristics of the dominant lags and are less influenced by the low correlation estimates from the first subregion.

The analysis of traffic based on decomposition and application of nonlinear autoregressive models has also been applied on the LAN-1 data representing the Bellcore Ethernet traffic measurements. Details of this analysis are presented in ^[12]. Since this data consisted only of the time-stamps and packet sizes, the traffic was decomposed based on packet sizes only. Among the resulting three classes which may be denoted as *LB_2*, *MB_2* and *HB_2*, the middle *MB_2* range which consisted of packet sizes in the [80,180) range exhibited strong cyclical trends of period 0.8 seconds and nonlinear features in certain amplitude regimes. A TAR model fit of the data was shown to accurately model the marginal and queuing statistics in the traffic. Although in the case of the Bellcore data a more complex TAR process, characterized by three delay parameters was found to be required. The TAR model fit of the *MB_2* acf estimates is shown in Fig. 7. The performance of the TAR model was also evaluated with respect to losses in a finite buffer queue. These results are depicted in Fig. 8. Both marginal statistics and loss probabilities were found to be reasonably well captured by the nonlinear autoregressive model.

4. TCP Congestion Control Mechanism and Traffic Shaping Effects

The observation of deterministic cyclical trends in TCP traffic motivates the consideration of the impact that TCP control mechanisms may have in shaping traffic. Here we offer a qualitative explanation for the observed TCP patterns. TCP congestion control process maintains a time-varying window containing segments marked for transmission. This send window size is the minimum of the receiver advertised window and a congestion window size maintained by the source TCP. At each instant that TCP transmits a segment, it starts a timer and waits for an ACK. The retransmission time out (RTO) is calculated using the adaptive retransmission algorithm,

$$\begin{aligned} r[n] &= r[n-1] + g(e[n] - r[n-1]) \\ v[n] &= v[n-1] + h(|e[n] - r[n-1]| - v[n-1]) \end{aligned}$$

where $r[n]$ is the estimate of the round-trip-time (RTT), $e[n]$ is the RTT measurement, $v[n]$ is the estimate of RTT deviation, g and h are smoothing factors. The RTO is then set equal to $RTO = r[n] + Kv[n]$. In current implementations, $g = 0.125$, $h = 0.25$ and $K = 4$. The estimates $r[n]$ are essentially first-order autoregressive processes with an AR parameter equal to $1 - g$. The expected value of $r[k]$ converges to the mean value of the driving noise process $e[k]$ and is an unbiased estimator. The variance of $r[k]$ is equal to the variance of $e[k]$ scaled by a factor of $g/1 - g$. Small variability in the measurements $e[k]$, translates to accurate estimates of the ack round trip times. As a result, in the limit as the RTT deviations tend to zero, the congestion window continues to grow at a deterministic rate governed by the mean round-trip-time of the acks. This is the situation one may see when the sequence of acks during their return path progress through a congested node and experience near deterministic delays in these nodes which are running close to full buffer occupancy. This feature while increasing the magnitude of the ack delays, reduces their variability resulting in deterministic and periodic transmission rates at the source TCP.

The above hypothesis was tested by a simulation study using the TCP implementation in the MIL3 OPNET network design and modeling system. This simulation considered a single TCP source node capable of transmitting a continuous stream of data and a receive node of infinite window size. The delays experienced by the acks were controlled by varying the congestion level at an intermediate node in the return path. In the congested state, the ack round-trip-times approach a Gaussian distribution, with a

mean value governed by the size of the buffer in the congested node. The variance of this distribution decreases with increase in congestion level. The behavior of the TCP congestion window in this state is depicted in Fig. 9. It is seen to exhibit a linear increase in its size with time. The reduced variability in the estimates $e[k]$ prevents the occurrence of retransmissions and the congestion window therefore continues to grow. In this linear phase, every time an ack is received the congestion window increases by one segment. The source TCP is driven by an application layer stream with Poisson distributed inter-arrivals and packet sizes. The acf estimates of the traffic that is generated by TCP is depicted in Fig. 10. An uncorrelated source stream is seen to be modulated into cyclical patterns through the influence of the congestion control dynamics. The periodicity in the ack round-trip times is seen to be mapped into a periodic transmission rate on the part of TCP. These mechanisms may be considered in trying to identify the causality of periodic components in TCP traffic.

5. CONCLUSIONS

This paper has presented a methodology for analyzing and modeling TCP traffic. Traffic analysis was carried out based on the decomposition of aggregate traffic using protocol type and packet size as features. This process was shown to provide an understanding of the origin of deterministic cyclical trends in the aggregate traffic. Periodic signals with time periods of 0.08 and 0.8 seconds were found to be dominant in the wide-area and local-area network traffic respectively. These components were found to be present in distinct amplitude ranges of the traffic arrival rate process. As a result non-linear autoregressive models were found to be appropriate to model the local dynamics in different amplitude regimes. A brief qualitative analysis of TCP dynamics provided one explanation for the existence of cyclical patterns in TCP traffic. This feature can result due to the periodic growth of the TCP congestion window in response to ack round trip times that approach deterministic values in the presence of congestion in the return path.

6. ACKNOWLEDGEMENTS

The authors would like to thank the National Laboratory for Applied Network Research for making available the traffic traces under the NSF NLANR/MOAT Cooperative Agreement (ANI-9807479). This research was supported by a National Science Foundation CAREER grant ANI-9734585.

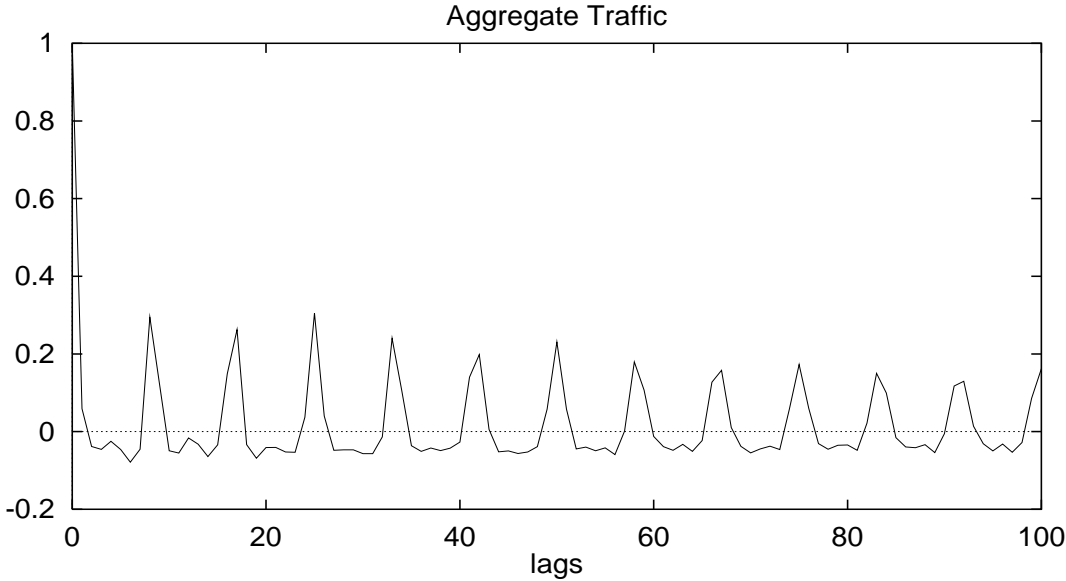


Fig. 1(a) Aggregate Traffic ACF

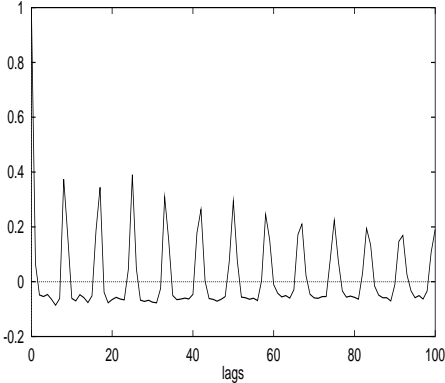


Fig. 1(b) TCP traffic

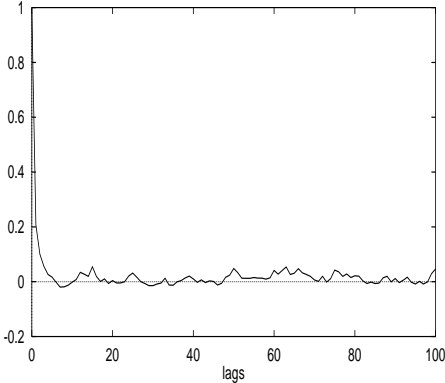


Fig. 1(d) UDP Traffic

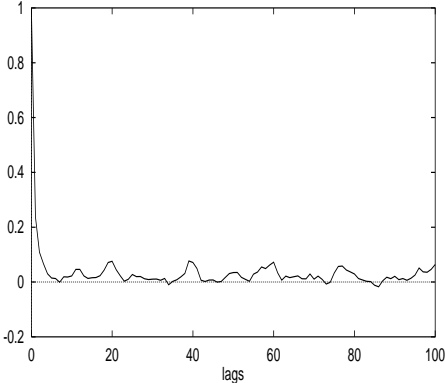


Fig. 1(c) IP/IP traffic

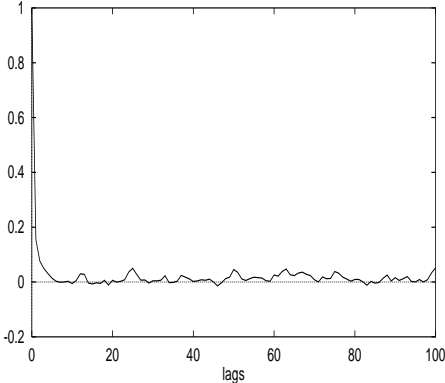


Fig. 1(e) ACF of other traffic

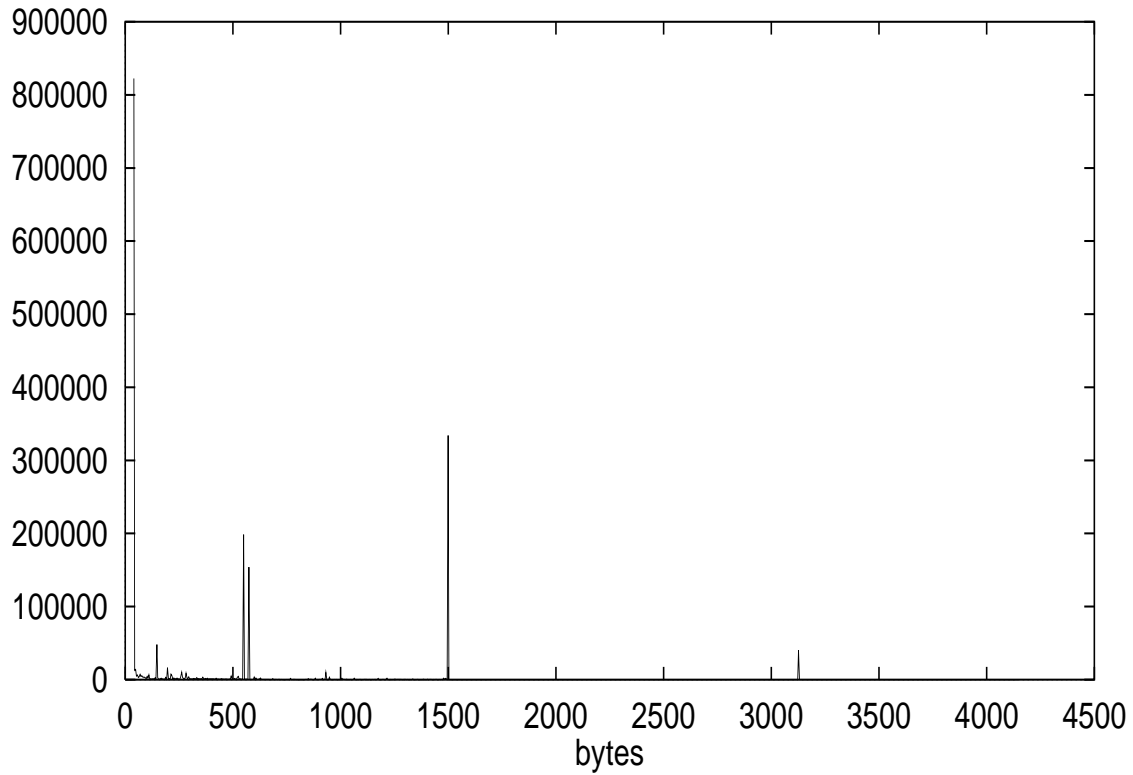


Fig. 2 Packet Size Distribution For TCP Traffic.

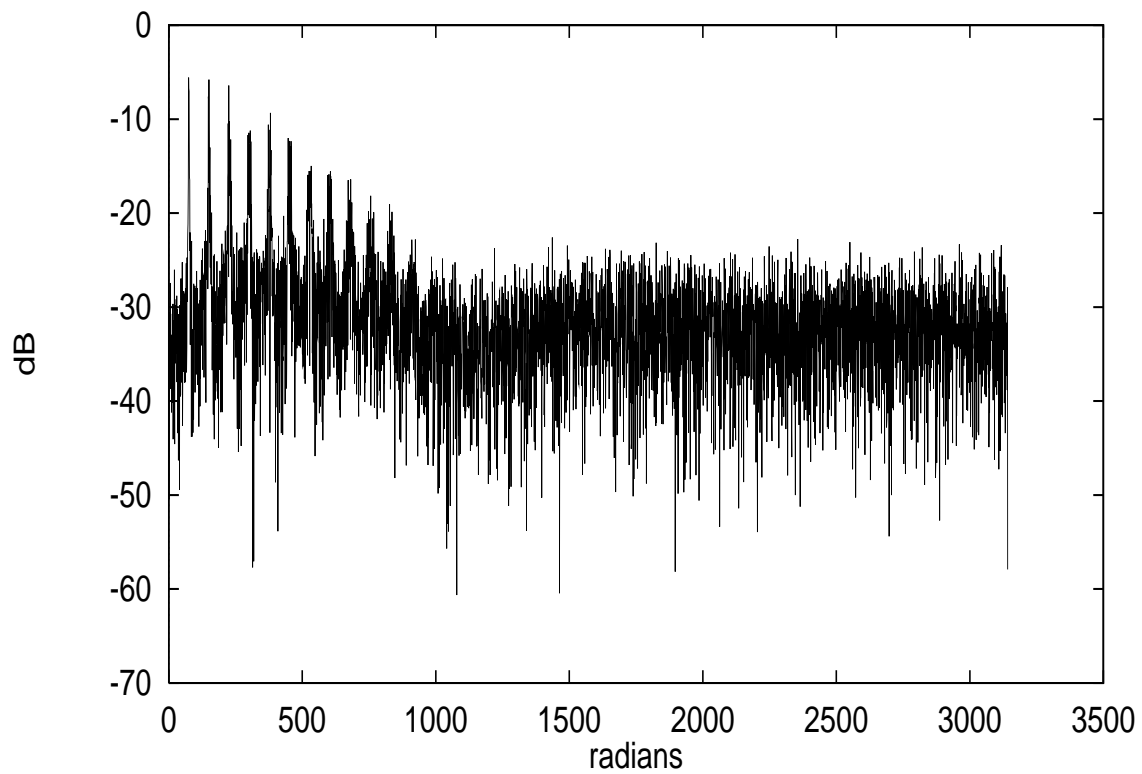


Fig. 3 Power Spectral Density (PSD) estimate for traffic class HB-1.

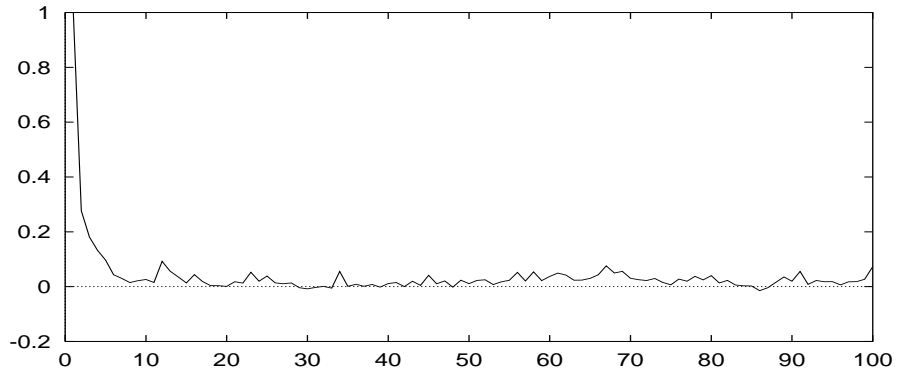


Fig. 4(a) ACF for LB-1; Packet Sizes between 0-200bytes.

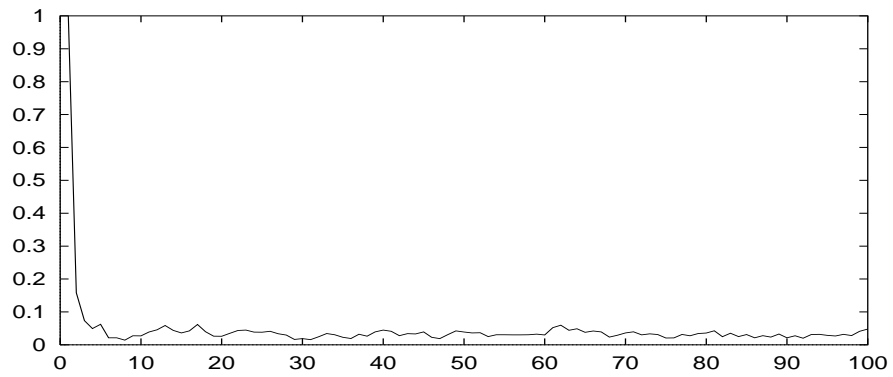


Fig. 4(b) ACF for MB-1; Packet Sizes between 200-2000bytes.

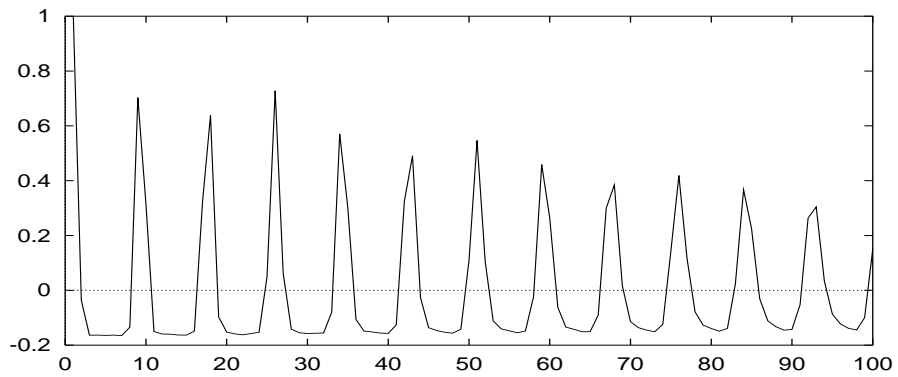


Fig. 4(c) ACF for HB-1; Packet Sizes between 2000+ bytes.

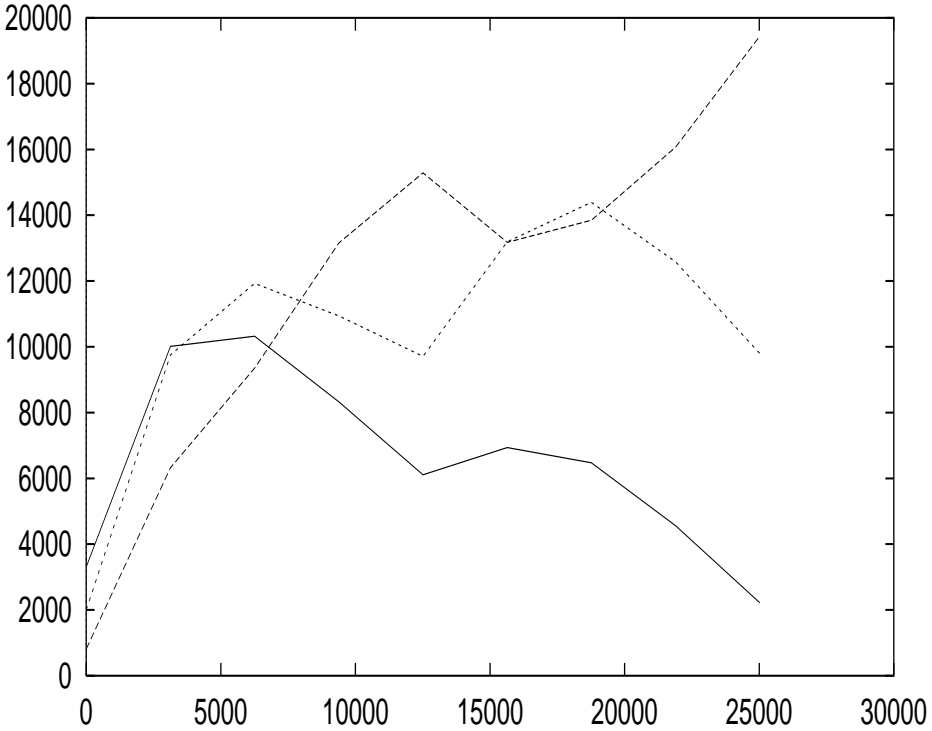


Fig. 5(a) Regression Plots for Traffic Class HB-1. Lags, 1, 7,8

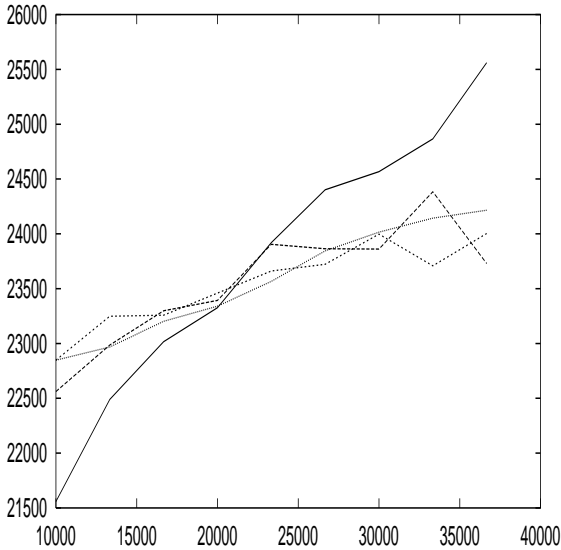


Fig. 5(b) Regression Plots for LB-1. Lags 1-4

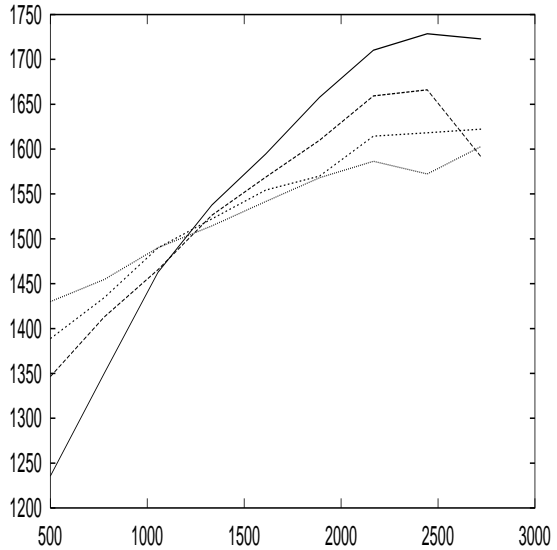


Fig. 5(c) Regression Plots for MB-1, Lags 1-4

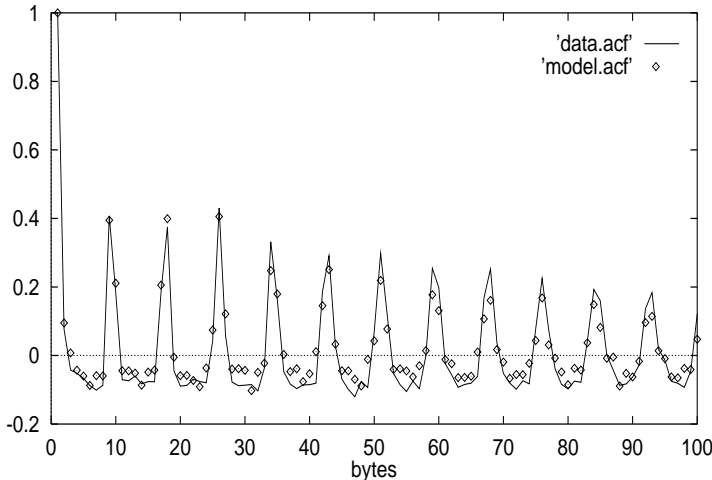


Fig. 6 Comparison of ACFs of wide area network TCP measurements and TAR model

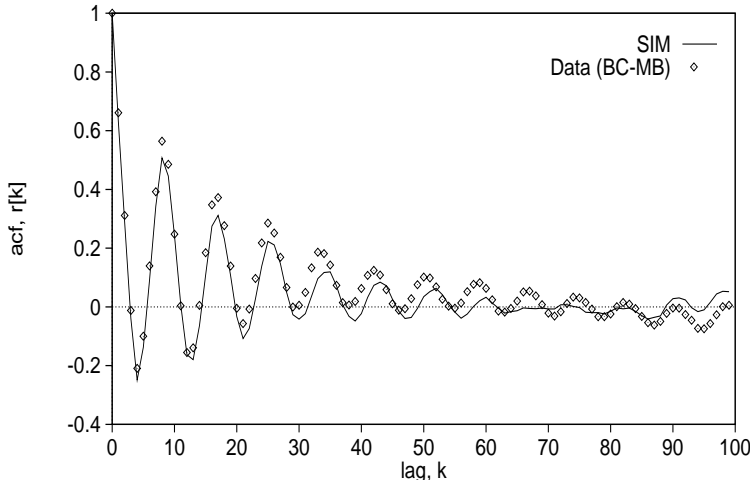


Fig. 7 Comparison of the acfs for one class of the Bellcore Ethernet traffic with traffic simulated using a TAR model.

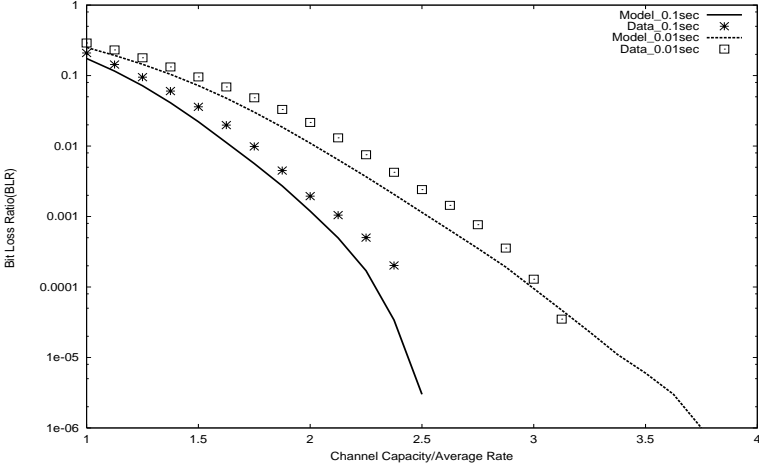


Fig. 8 Comparison of bit loss ratios for measurements and traffic simulated using the TAR model.

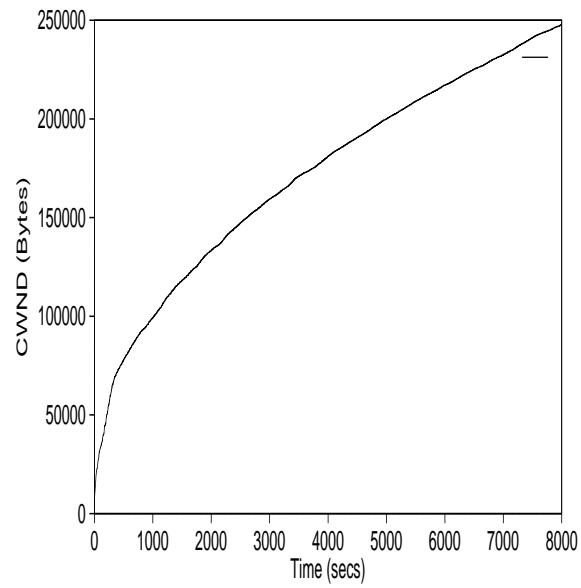


Fig. 9 TCP Congestion Window Behavior

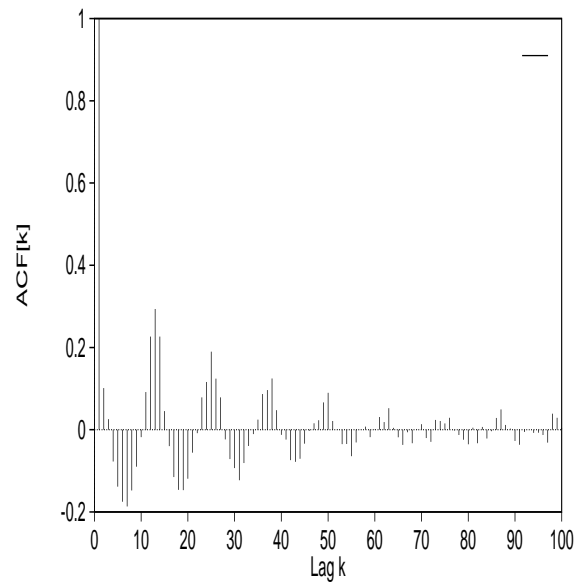


Fig. 10 TCP modulated traffic pattern.

REFERENCES

1. K. Meier-Hellstern, P.E. Wirth, Y.-L. Yan and D.A. Hoeflin, "Traffic Models for ISDN data users: Office Automation application," in *Teletraffic and Data Traffic in a Period of Change*, Eds. A. Jensen and V.B. Iversen, Proc. of ITC-13, Copenhagen, pp. 167-172, Elsevier Science Pub., Amsterdam, 1991.
2. V. Paxson and S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Trans. Networking*, **3**, (3), p226-244, June 1996.
3. W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, 'On the Self-Similar Nature of Ethernet Traffic (Extended Version),' *IEEE/ACM Trans. Networking*, p1-15, **2** (1), 1994.
4. D.D. Clark, S. Shenker and L. Zhang, "Supporting real-time applications in an integrated services packet network: Architecture and mechanism", *Proc. SIGCOMM 92*, pp. 14-26, (1992).ion,' *IEEE/ACM Trans. Networking*, p1-15, **2** (1), 1994.
5. B.K. Ryu and A. Elwalid, "The importance of long-range dependence of VBR video traffic in ATM traffic engineering: Myths and Realities," p3-14, Proc. ACM SIGCOMM'96 conf., Stanford University, CA., Aug. 1996.
6. M. Grossglauser and J.D. Bolot, "On the relevance of long-range dependence in network traffic," p15-24, Proc. ACM SIGCOMM'96 conf., Stanford University, CA., Aug. 1996
7. S. Basu, A. Mukherjee and S. Klivansky, "Time Series Models for Internet Traffic," p611-620, vol. 2, Proc. IEEE INFOCOM'96, San Francisco, CA., March, 1996.
8. H. Tong, **Threshold Models in Non-linear Time Series Analysis**, Lecture Notes in Statistics, vol. 21, Springer-Verlag, 1983
9. H. Tong, **Non-linear Time Series , A Dynamical System Approach**, Oxford Science Publications, Clarendon Press, Oxford, 1990
10. P.A.W. Lewis and B.K. Ray, "Modeling Long-Range Dependence, Nonlinearity and Periodic Phenomenon in Sea Surface Temperatures using TSMARS," *J. American Statistical Association*, **92** (439), p881-893, September 1997.
11. S.M. Kay, **Modern Spectral Estimation**, Chap. 7, Prentice Hall, Englewood Cliffs, NJ, 1988
12. K. Chandra, C. You, G. Olowoyeye and C. Thompson, "Non-Linear Time-Series Models of Ethernet Traffic", submitted to INFOCOM'99, July 1998.