Non-Linear Time-Series Models of Ethernet Traffic

Kavitha Chandra, Chun You, Gbenga Olowoyeye and Charles Thompson

Center for Advanced Computation and Telecommunication
Department of Electrical Engineering
University of Massachusetts Lowell
Lowell, MA 01854

ABSTRACT

In this paper non-linear threshold autoregressive models are examined for use in modeling the temporal variation in the byte-rate in Ethernet traffic. The model is comprised of a number of autoregressive processes each of which is to be used in a specified range of amplitude of the byte-rate. The local dynamics within each threshold range are captured by an autoregressive process. The switching between each submodel is conditioned on the amplitude of a lagged value of the time-series. To develop the model the Bellcore Ethernet LAN data is used. It is shown that non-linear threshold autoregressive processes can be used to capture the dynamics of Ethernet LAN traffic. This model also provides for both short and long-term prediction capability and allows us to quantitatively identify the sources of long-range-dependence features in the traffic. When the aggregate traffic is partitioned into classes based on packet sizes, certain classes of traffic follow deterministic cyclical patterns. These periodic components arise from the process switching between different amplitude regimes. Superposed on this fundamental period are longer cycles that can be localized either below or above the mean byte-rate. By constructing amplitude thresholds associated with a finite set of delay parameters, the dynamics within each threshold are captured by locally linear autoregressive processes. The aggregate process is globally nonlinear. This model is shown to provide good agreement with the marginal distributions and the correlation functions derived from the Ethernet traffic data. In addition, simulation experiments demonstrate that the loss statistics observed in finite buffer queues agree favorably with those generated by the measurements.

1. INTRODUCTION

The demand for the introduction of applications with real-time constraints on to data networks has created a need for the development of accurate forecasting models for existing traffic on these networks. Recent studies [1] [2] [3] of network traffic measurements have shown that data traffic exhibits significantly higher variability than Poisson processes. This becomes evident when one observes in this traffic larger magnitudes for the coefficient of variation for the packet inter-arrival times and monotonically increasing values of the index of dispersion of packet counts. Data traffic has also been shown to possess temporal correlation that persists over time-scales that range from milliseconds to over hundreds of seconds. When

the autocorrelation function (acf) decays hyperbolically for large lags, the traffic is said to be characterized by long-range dependence (LRD). These features prove problematic for standard traffic models.

Leland et. al. [3] analyzed Ethernet traffic on a LAN connecting workstations, file servers and personal computers. They showed that the traffic exhibits same degree of correlation when aggregated using window sizes increasing from seconds to hours. To capture this feature of persistence in correlation over several time scales, Leland et. al. proposed a self-similar process as a possible traffic model. For the wide-area networks, Paxson and Floyd, [2] analyzed connection and packet arrivals in wide-area TCP traffic and concluded that wide-area traffic, particularly at the packet level was much more bursty than that predicted by Poisson models. Paxson and Floyd also observed that between 40% and 60% of all of the data bytes measured was contributed by a small fraction of the largest bursts carrying FTP data bytes. These previous studies have amply demonstrated the presence of high variability and strong correlation patterns in traffic measurements. The basic questions resulting from these empirical studies concern: (i) the causality of the correlation and long-range dependence, (ii) the impact of LRD on the design and performance of network software and hardware elements and (iii) the choice of an appropriate model for performance studies. The causality of correlation and long-range dependence has been addressed in Willinger et.al. [4] where the superposition of individual on-off sources with heavy-tailed distributions is shown to be a possible explanation for the observed self-similar like features in the aggregate data. Crovella and Bestavros [5] carried out an analysis of World Wide Web traffic and attributed the heavy tailed distribution of document lengths to the self-similar features in the traffic. As to the impact of LRD on the queue and multiplexer performance, recent studies [6] [7] of network performance over a range of cell losses, buffer sizes and network operating parameters have shown that the network operating conditions such as utilization and buffer size determine the number of correlation time-scales that must be modeled. Long-term correlations beyond a outside time-scale do not significantly affect network performance.

Traffic models based on second-order stationary self-similar processes have been proposed in [1]. Stationary time-series models using autoregressive moving average (ARMA) processes have been applied in Basu et. al. [8]. Here differencing of the traffic is found necessary to account for drifts and trends in the mean level of the data. This approach will require frequent updates from the measurements for bounding the predicting errors.

Non-stationary behavior can arise from shifting mean levels, changing parameters of a basic

structural model or both. If one considers the sequence of byte-rate dependent transformations typically undergone by a source traffic stream from the application to the network level, non-linear time-series model may be applicable. Transport control protocols such as TCP impose constraints on the rate of data flow from the source. As a result the source responses change as a function of the instantaneous load on the transit networks and the change is observed after a finite time delay in feedback information. Both the load on the transit networks and the feedback delay are unobserved variables for the traffic modeler. File transfer protocols such as FTP, HTTP, which constitute the dominant fraction of the total traffic apply control actions in response to these unobserved variables resulting in a traffic flow that is modulated by exogenous influences. In addition, certain amount of traffic on local area networks may arise from network management protocols operating in deterministic patterns. The superposition of such disparate traffic patterns without first examining individual traffic types can lead to increased variability and complex features in the aggregate flows.

In this paper, we approach the problem of modeling dependence features in the traffic by first partitioning the aggregate traffic into a set of classes which are derived from a feature space associated with each packet arrival. This decomposition allows us to explicitly identify the components that contribute to the high variability in the aggregate traffic. Based on these features, a non-linear time-series model is proposed as a traffic model. The paper is organized as follows. In Section 2.0 we present characteristic features of Ethernet data and some results of statistical analysis that justify the use of a non-linear model. Section 3.1 describes the fitting of threshold autoregressive processes, a class of non-linear time-series models to the data. Section 3.2 presents evaluation of the model with respect to marginal distributions, autocorrelation functions and the losses in finite buffer queues. Section 4.0 presents some conclusions.

2. CHARACTERISTIC FEATURES OF ETHERNET TRAFFIC

Our approach is motivated by the temporal characteristics observed in the BellCore Ethernet traffic measurements described and analyzed by Leland et. al. in [1]. This data set is referred to as *BCAug89* and is available in the public domain at the Internet Traffic Archive site *http://ita.ee.lbl.gov/html/contrib/BC.html*. This data was also analyzed in [8] using a stationary time-series approach. The trace consists of packet arrival time stamps and packet data sizes of a million packets arriving in a time duration of about 3000 seconds. The measurements were made beginning at 11:25 am on a weekday and corresponds to a normal hour of business.

There are two salient features in the *BCAug89* data that motivate our inquiry. These are nonstationarity in the byte-rate and the temporal correlation in the packet sizes. Let us first consider the long-term non-stationarity in the arrival rates. The time series for the byte-rate is shown in Figs. 1(a-c). The time series is obtained by aggregating the packet data size over successive non-overlapping windows having a fixed time duration. The byte-rate time series is shown for time intervals of 1, 10 and 100 seconds in duration. While significant variability is retained at the 1 and 10 second time-scales, Fig. 1(c) at the 100 second time scale clearly depicts a basic structural pattern consisting of a fast transition to a peak or high arrival-rate state followed by relatively slower rate of transition to an off-peak or low state. We can identify three such cycles in Fig. 1(c), with a cycle time period of approximately 800 seconds. We may attribute this feature to the sources adapting to a time-scale typically governed by exogenous and unobserved variables. For example, increased delays on the transit networks for the sources or even the presence of one bottleneck node that interconnects the LAN to the Internet can trigger the exponential back-off mechanisms in TCP control protocols leading to a smooth transitions to the low states resulting in reduced source rates on the local network. The long-term trend shown in Fig. 1(c) causes the mean levels to shift in time.

The second feature we address is the temporal correlation in the packet sizes. The traffic measurements consist of two random variables, the packet inter-arrival times and the packet sizes. Often in data analysis packets are aggregated over fixed time intervals to yield a time-series for analysis. This process neglects to take into account any dependence that may exist between the packet sizes and inter-arrival times. Since packet sizes in data traffic depend to some extent on the application generating the packets, one may expect differing range of packet sizes to generate different traffic dynamics. The probability distribution of packet data sizes in *BCAug89* is depicted in Fig. 2. It can be seen that the data sizes are concentrated around values of 64, $160-170$ and $1090-1518$ bytes. Scatter diagrams of two and three lag dependence in the packet sizes also depict dominant clusters in these range of packet sizes, indicating temporal correlation within these packet size ranges.

Based on these observations, we will partition the data into three time-series. These series will be formed from packet sizes in the low, mid and high ranges. The low, mid and high ranges are taken to be characterized by packet sizes in the $[0-80), [80, 180)$, and $[180, 1518)$ bytes respectively. This is a departure from previous time-series studies that only considered the aggregate of data bytes over fixed

time intervals without consideration as to the packet size. A time-interval of 0.1 seconds is used to aggregate the individual packets. In such a case, the low range constitutes about 15% of the aggregate data, the mid range represents 54% of the data, with the remaining 31% being made up by packets carrying [180, 1518) bytes. Note this range as depicted in Fig. 2 primarily consists of 1090 and 1518 byte packets. The three time-series are referred to as *BC-LB*, *BC-MB*, and *BC-HB* respectively.

2.1 Statistical Features

The time-series defined in the previous section are individually analyzed here. The histograms of bytes per time interval and the autocorrelation functions (acfs) of the individual series are depicted in Figs. 3(a-c) and Figs. 4 (a-c) respectively. Under the decomposition of the aggregate traffic, the average arrival rate for the three processes are 306, 2509, and 10, 345 bytes per 0.1 second time interval respectively. The histogram of *BC-LB* is composed primarily of integer multiples of 64 byte packets , ranging from a minimum of one to twenty two packets per time slot. The histograms of *BC-MB* and *BC-HB* time-series clearly depict modes in the distribution. Observation of the acfs in Figs. (4) show that under the decomposition, the long-range dependent structure of the aggregate traffic has been relegated primarily to the *BC-HB* data. The low and mid range packet sizes exhibit relatively short-term dependence. In particular, the acf of *BC-MB* indicates the presence of strong cyclical behavior with a fundamental period of 8 lags corresponding to 0.8 seconds and a e-fold time of approximately five seconds. Since over half of the traffic data is represented by *BC_MB*, the temporal dynamics of this time-series will be first addressed.

Although the acf of *BC-MB* suggests the use of a linear ARMA model to model the short-term cyclical behavior, such a model will fail to produce the degree of variability that exists in the data sequence. The residuals of such a model will exhibit white noise properties, but the pdf is non-Gaussian. Careful examination of the data reveals that in addition to the fundamental period of *0.8* seconds, the process can exhibit longer cycles during which it sojourns in a high or a low state. An example of these features is shown in Fig. (5) which depicts a 200 second section of the *BC-MB* time series where the traffic resides for longer than its expected duration of 4 lags above the local mean value and exhibits similar features in the low state as well. We note that similar features are possibly the cause of the variability in the *BC-LB* and *BC-HB* time series, but since the surrounding envelope is less deterministic they are harder to detect and characterize. However, insight gained by modeling the dynamics of the more structured *BC-MB* series will help in this regard. Since the observed features cannot generally be captured by linear

processes we consider the application of non-linear models that can incorporate the aforementioned dynamics.

2.2 Non-Linear Features

The sample acfs represent linear dependence in the time-series data. To test for nonlinear behavior, conditional statistics as given by sample regression functions [9] will be considered. The lag j regression function is defined as $r_j = E[X_n | X_{n-j} = x]$. The estimates of r_j for $j = 1, 2, 3, 4$ derived from the three time-series are depicted in Figs. 6(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 linear processes that are normally distributed, the regression functions exhibit a linear trend. For the *BC-LB* data in Fig. 6(a), the sample regression functions are seen to be approximately linear for $j = 1$ and $j = 2$, but show weak turning points in the $300 - 800$ amplitude range of the data. In comparison, the lag $2, 3$ and 4 regression functions for *BC-MB* in Fig. 6(b) show the strongest departure from linearity. Here the cyclical properties of the data with period 8 creates the strongest deviation from linearity at a lag corresponding to half of the cycle period 4. For higher lags, the regression functions move back in the reverse direction towards the lag one function. The structure of these functions for *BC-HB* series shown in Fig. 6(c) also show departures from linearity, however the high variance in this time-series makes these features less pronounced in comparison to those given for Fig. 6(b). The turning points in the regression functions indicate the presence of amplitude regimes characterized by both positive and negative dependence. As a first departure from linearity, we consider modeling the structure in the data using piece-wise linear models at appropriate amplitude regimes. The threshold autoregressive process (TAR), proposed by Tong, [10] which affords the framework for constructing piece-wise linear models will be considered in the next section.

3. THRESHOLD AUTOREGRESSIVE MODELS

The threshold autoregressive (TAR) model proposed by Tong [10], 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 LRD features [11]. Particular examples that appear to exhibit the features seen in the traffic data are the classical

Wolf Sunspot numbers [12] and Canadian Lynx data [10] , both of which have been shown to be more accurately modeled using non-linear time series.

The model considered here will incorporate two amplitude ranges. They will be denoted as low (L) and high (H) amplitude states. In the low-state the times-series takes on values L: $(0, \hat{r}]$, where $\hat{r}$ is the threshold value. The high state accommodates amplitudes H: $[\hat{r}, \infty)$. In addition three delay values d_1, d_2, d_3 will be used in the conditional switching. This is done to capture the observed excursions about the amplitude $\hat{r}$. For such a case, the AR model selected will be determined by the amplitude of the time series at three previous points in time. The total number of amplitude conditions that occur is equal to 2^3. Each case will be labeled R_j where the index j takes on integral values between 1 and 8. Fig. 8 depicts a schematic of the location of TAR model parameters for one subregion. In this figure the amplitude of $x(n)$ at three delay values (relative to $\hat{r}$) classify the time-series model to be used. In Table I, the rest of the cases are shown. The first column denotes the threshold condition R_j , whereas the next three columns show in which of the two amplitude ranges $x(n-d_j)$ resides. 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. In each of the cases R_j the process evolves as a stable AR process, governed by the correlations within that region.

More delay parameters may be accommodated by increasing the number of subregions to be used to characterize the data. In general, the amplitude condition may be based on the lagged variables { $x(n-d_1), x(n-d_2), \cdots, x(n-d_L)$}. Each case will be denoted by R_j, where $j = 1, 2, \cdots 2^L$. The current value of the byte-rate at time n will be governed by an autoregressive process of order k_j.

$$x(n) = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^{(j)}\, x(n-i) + e^j(n) \qquad \{x(n-d_1), x(n-d_2), \cdots x(n-d_L)\} \in R_j \qquad (1)$$

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. The delay parameters afford the flexibility of capturing persistence phenomenon at the required amplitudes. This is an important feature when the process continues to reside above the mean value. This feature can be captured by including delay values beyond half of the fundamental period of oscillation. Extended sojourns in the high byte-rate state is an important feature in delay management, whereas dwell-time in the low

byte-rate state has impact on multiplexing efficiency. Therefore the thresholds R_j 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 $d_i, i = 1, 2, \cdots$ must be selected along with the coefficients of the local AR processes for each subregion. We will describe the parameter selection process used in the analysis of the *BC-MB* data. This data set exhibits the most interesting non-linear features of the three. Inspection of the sample regression functions for this data given in Fig. 5(b) suggests that the threshold $\hat{r}$ delimiting the low and high states be in the amplitude range $2000 - 4000$. This is necessary if one is to model the change in slope of the regression function in this region of amplitudes. This region coincides with the location of the local mean value. We constrain the problem by fixing the number of delay parameters to three. The maximum delay value captures the dependence in the data. For the *BC-MB* data set, the three delay values $\{ d_1, d_2, d_3 \} : \{ 1, 4, 7 \}$ were found to be adequate. Using these lag parameters, the threshold value $\hat{r}$ and the local AR coefficients $\underline{a}^j$, and the variance of the residual σ_j^2 are determined using least-squares estimators.

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 $x_{i_1}^j, x_{i_2}^j, \cdots, x_{i_{n_j}}^j$ of the time-series represent the n_j samples that satisfy constraint R_j. These n_j samples will be denoted by the vector, $\underline{x}^j = (x_{i_1}^j, \ldots, x_{i_{n_j}}^j)^T$. For this data, the k_j^{th} order linear model coefficients are evaluated.

$$\underline{x}^j = A^j \, \underline{a}^j + \underline{e}^j \tag{3}$$

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

$$A^j = \begin{bmatrix} 1 & x_{i_1-1}^j & x_{i_1-2}^j & . & . & . & x_{i_1-k_j}^j \\ 1 & x_{i_2-1}^j & x_{i_2-2}^j & . & . & . & x_{i_2-k_j}^j \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 1 & x_{i_{n_j}-1}^j & x_{i_{n_j}-2}^j & . & . & . & x_{i_{n_j}-k_j}^j \end{bmatrix} \tag{4}$$

and $\underline{e}^j = (e_{i_1}^j, \ldots, e_{j_{n_j}}^j)^T$ 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 $\underline{\hat{a}}^j$. The least-squares error,

$$\underline{\hat{e}}^j = \underline{x}^j - \underline{A}^j \, \underline{\hat{a}}^j \tag{5}$$

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

In order the 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) [13] of the sub-models as the performance measure. The process of TAR parameter estimation is that outlined in Tong[9] and and will be briefly described here.

For the fixed set of delay parameters $d_1, d_2, .. d_3$ the threshold amplitude $\hat{r}$ is sampled uniformly between 2400 and 3600 in 200 byte-per second intervals. The maximum model order of the AR model was set equal to 30. 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 30. In each case the residual errors $\underline{\hat{e}}^j$ determine the best model order for the subregion using the Akaike Information Criteria (AIC). The AIC is given by

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

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

$$AIC_{total}(\underline{d}, \hat{r}) = \sum_{j=1}^{8} AIC(k_j) \tag{7}$$

where $\underline{d} = \{d_1, d_2, d_3\}$. The optimal threshold parameter $\hat{r}$ and AR parameters is one which yields the minimum $AIC_{total}(\underline{d}, \hat{r})$.

The process outlined for TAR parameter estimation is carried out on the *BC-MB* set of data. Since the arrival process is non-stationary in the mean, we consider the TAR model fitting over segments of the data where the arrival rate as calculated over 100 second intervals does not exhibit significant deviation.

Incorporating up to seven lagged variables in the delay parameters $\underline{d} = (1, 4, 7)$ in the constraints, allows the isolation of cases where the process can cycle for twice its expected duration in the selected subregions. The threshold $\hat{r}$ that results from the minimum AIC criteria generally falls in the $2800 - 3000$ byte range. An example of the minimum $AIC_{total}(\underline{d}, \hat{r})$ for various values of $\hat{r}$ is shown in Table 2, for three different ensembles of length 300 in the *BC-MB* data. These results are consistent as the ensemble sizes are increased, although the magnitude of the AIC_{total} values also increase, due to an increase in the range of the residual error. In most all cases considered the threshold that yields the minimum AIC_{total} value is around the mean value of the data sequence considered. The optimal AR orders for the different regimes however vary up to $k_j = 25$ for some regimes. An example of a segment of the $BC - MB$ data considered in the model and results of the TAR simulation are depicted in Figs 7(a) and (b). In generating the TAR model simulation, only the initial conditions were determined from the measurements. Subsequent values were generated using the TAR parameter estimates and additive noise variables for each subregion. The noise process was derived from the empirical distributions of the error residuals obtained during the fitting of the measurements to the model. The acfs of the residual errors were found to pass the white noise test for all of the eight subregions. In the simulation, any negative values that resulted were set to be equal to zero.

The structural patterns in the *BC-MB* time-series can be discerned to a large extent due to a reasonably distinct partition between the pdfs of the low and high states as seen in the histogram of Fig. 3(b). Such a feature is less evident for the *BC-HB* time series. Although the acf for this data, as shown in Fig. 4(c) depicts weak cyclical structure at about the same period as in *BC-MB*. Based on the insight obtained from analyzing the data in *BC-MB*, a similar TAR model was considered for the *BC-HB* data as well. Since this data exhibits a longer range of dependence than that in *BC-MB* , the number of delay variables was required to be increased from 3 to 4. In particular, the set of delay parameters found appropriate for *BC-HB* time-series were , $\underline{d}: (1, 3, 7, 10)$, accounting for dependence up to 10 prior lags. The results that evaluate the proposed models for *BC-MB* and *BC-HB* are presented in the next section.

3.2 Model Evaluation

In this section we present several statistical features that validate the TAR model against measured data. Results are shown in Figs. 9 and 10 for the *BC-MB* and *BC-HB* data. First agreement between marginal distributions, autocorrelation functions and counting statistics are considered. The QQ

(quantile-quantile) plots [14] provide a visual assessment of the agreement of the distributions of the byte amplitudes in the time series. Figs 9(a) and 10(a) depict the QQ plots for the simulated and measured data. Figs. 9(b) and 10(b) depict the match between the autocorrelations functions from the measurements and simulation. For both traces, the model captures the dependence in the data upto about 60 lags. Figs. 9(c) and 10(c) depict the counting statistics of N_t, representing the total number of bytes arriving in a time interval t. This feature is captured in terms of the expected value $E[N_t]$, the variance $Var[N_t]$, and the index of dispersion $ID[N_t] = Var[N_t]/E[N_t]$. The index of dispersion represents the degree of variation of this traffic, relative to a Poisson process for which the index of dispersion converges to a constant value of unity. This second order descriptor has been used to capture the burstiness properties of arrival processes. The TAR model is seen to do reasonably well in capturing the traffic variability as measured by these statistics.

Next the performance of the model is evaluated in a queue consisting of a fixed buffer size and a constant service rate. The buffer size is expressed in units of seconds. This parameter represents the maximum delay experienced by the arrivals. For given source average (r_{avg}) and peak rates (r_p), the buffer drain rate C is incremented from r_{avg} to r_p. For each value of C the losses are expressed in terms of the bit loss ratio,

$$BLR = \textit{Number of bits lost / Total number in Data Sequence}$$

Figs. 11(a-b) depict the bit loss ratios obtained by driving queue simulations using TAR model generated data. These results are compared to the losses experienced by the measurements. The BLRs are depicted for buffer sizes of 0.01 and 0.1 seconds for the *BC-MB* data. Since the dependence range in this data is relatively short-term, for larger buffer sizes, no losses are observed. However, losses are experienced for larger buffer sizes for the *BC-HB* time-series. For this case the results are shown upto a buffer size of 0.5 second. In calculation of the BLRs, twenty ensembles of the TAR simulated data, each $10,000$ samples in length were considered. The average of the BLRs experienced over this set is depicted in Figs. 10. The results indicate that the TAR model captures the relevant long-range features in the data.

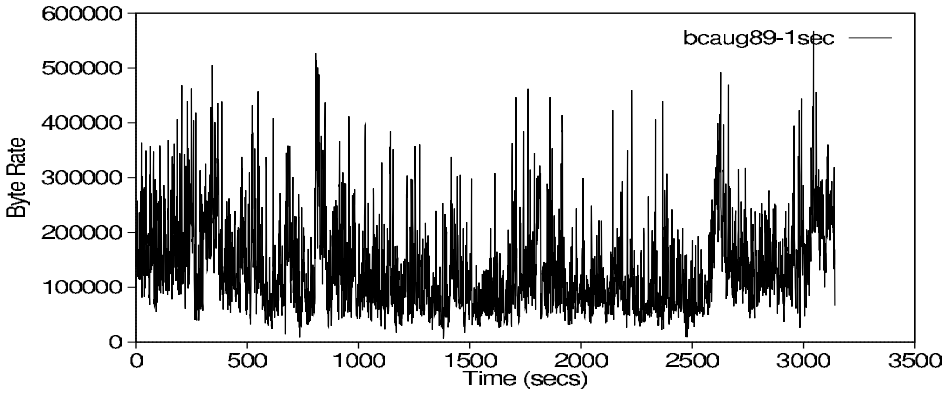
4. CONCLUSIONS

This paper has presented a traffic model for Ethernet LAN data using non-linear threshold autoregressive models. The traffic analysis carried out on one of the Bellcore Ethernet traces is based on

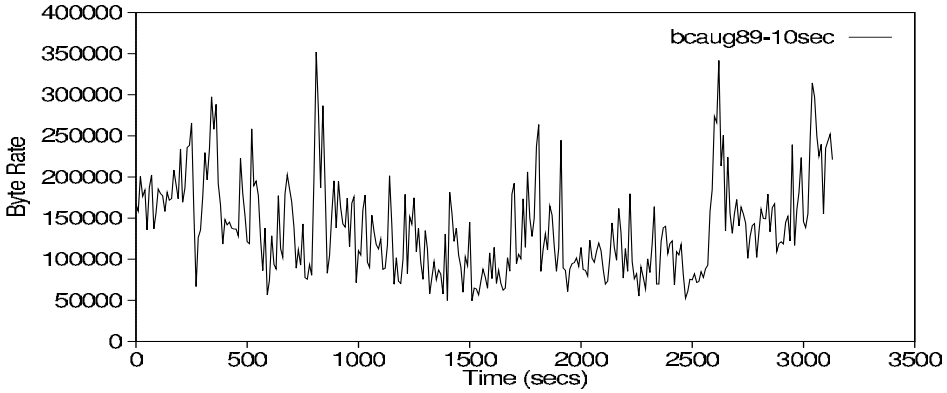
decomposition of the aggregate traffic into three components based on packet sizes. This process highlights a salient feature in the data that is captured in the mid-range of packet sizes from $[\ 80-180\ ]$ bytes. This set comprises over half of the measured data and exhibits deterministic cyclical patterns of period 0.8 seconds. Superposed on this pattern are randomly occurring longer cycles where the process sojourns either above or below the mean byte rate. These features cause the long-range dependence characteristics. Sample regression functions for this data show clear departures from linear behavior in certain amplitude regimes. A threshold autoregressive model with two amplitude thresholds and upto three delay parameters is shown to capture these features in the data. The time-series formed from the mid-range packet sizes exhibit clearly discernible trends due to the reduced variance in their distribution. Based on this insight, it was determined that the same structural patterns cause the LRD features in the high-range of packet sizes that comprise about 35% of the data in the tail of the distribution, but the high variance in this range reduces the detection capacity of these patterns. By increasing the delay parameters to four lags, the TAR model was shown to adequately model the data for the high-byte range time-series as well. The TAR model simulated traces were compared with 500 second segments of the Ethernet data and shown to agree favorably both in terms of the marginal distributions, autocorrelation functions and counting statistics. The time-scales considered in the TAR model were also shown to be adequate to match the losses in finite-buffer queues with buffer sizes ranging from 0.01 to 0.5 seconds.

5. ACKNOWLEDGEMENTS

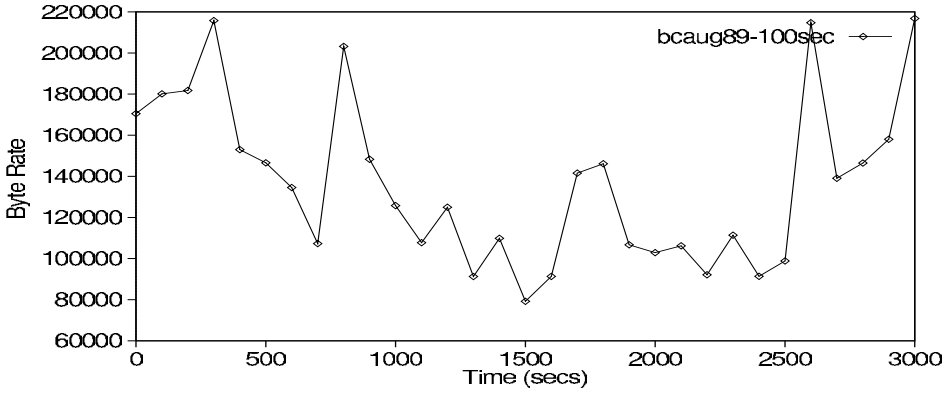
The authors would like to thank W. Leland and the Internet Traffic Archive sponsors for making the Ethernet traces available. This research was supported by the National Science Foundation CAREER grant NCR-9734585 and grant NCR-9729082.



(a)



(b)



(c)

Fig. 1(a-c) Time-series of arrival rate (bytes/sec) aggregated over (a) 1 second, (b) 10 seconds and (c) 100 second intervals. *(BCAug89)*

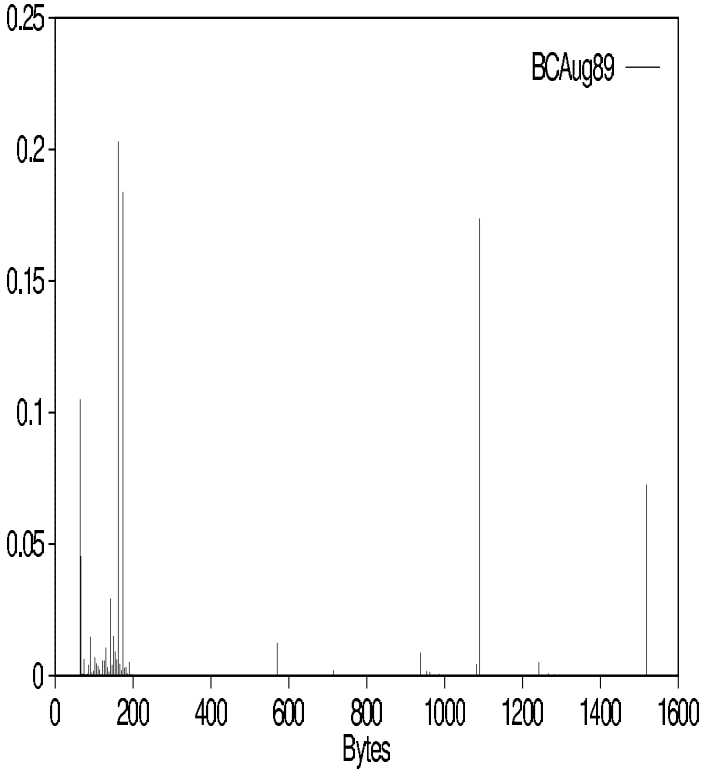
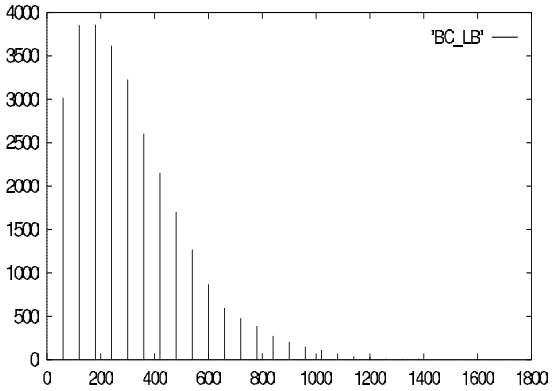
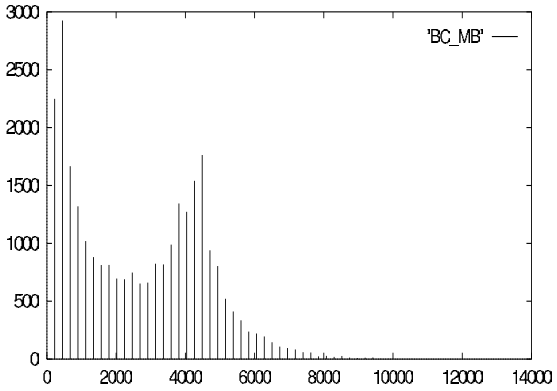


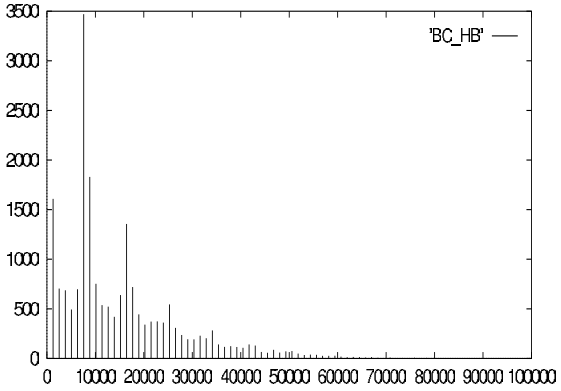
Fig. 2(a) Probability distribution of packet sizes in *BCAug89* **.**



(a)

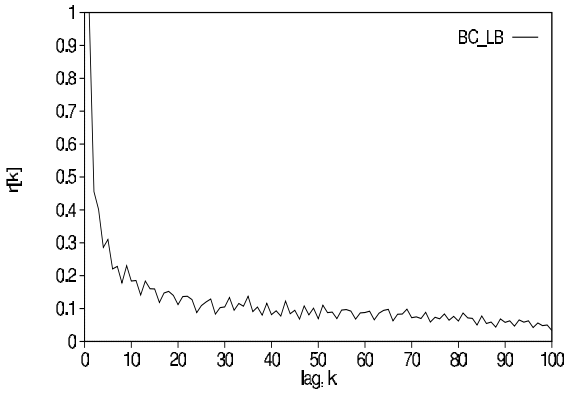


(b)

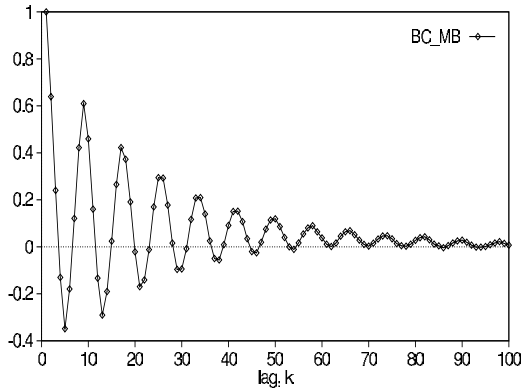


(c)

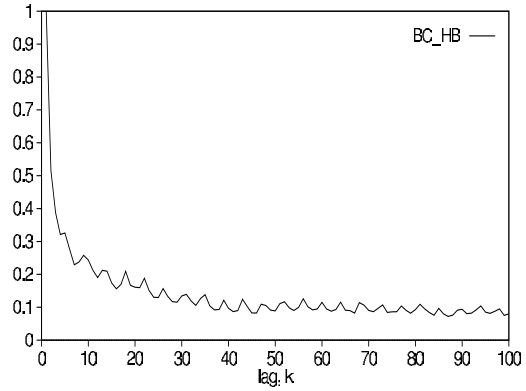
Fig. 3 Histogram of bytes per unit time (0.1s).



(a)



(b)



(c)

Fig. 4 Autocorrelation functions for low, medium and high byte range time series.

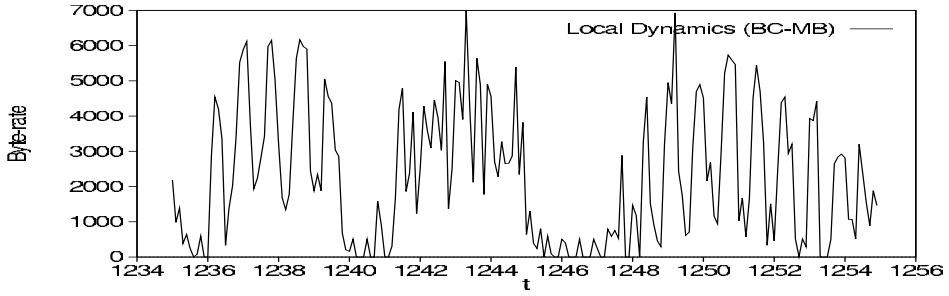


Fig. 5 Example of local dynamics in *BC-MB* data.

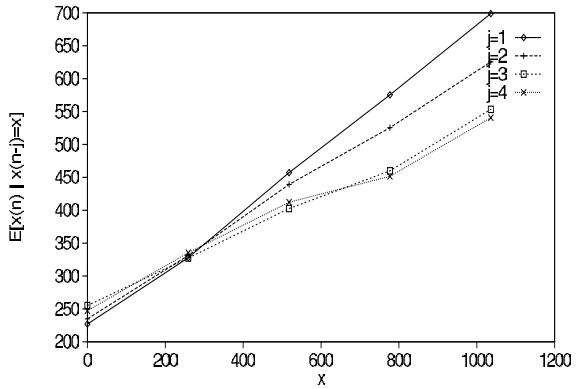


Fig. 6(a) Lag 1-4 regression function estimates for BC-LB

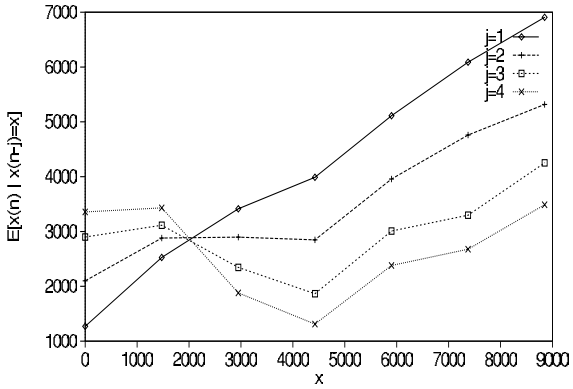


Fig. 6(b) Lag 1-4 regression function estimates for BC-MB

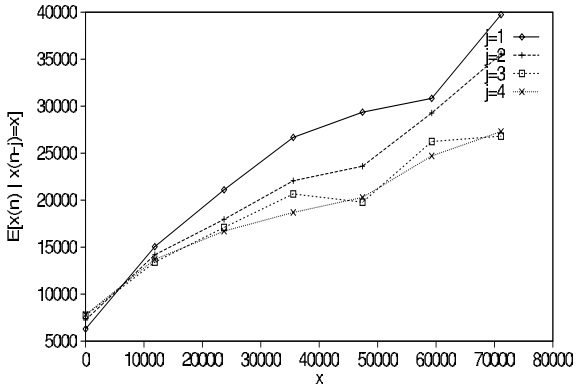


Fig. 6(c) Lag 1-4 regression function estimates for BC-HB

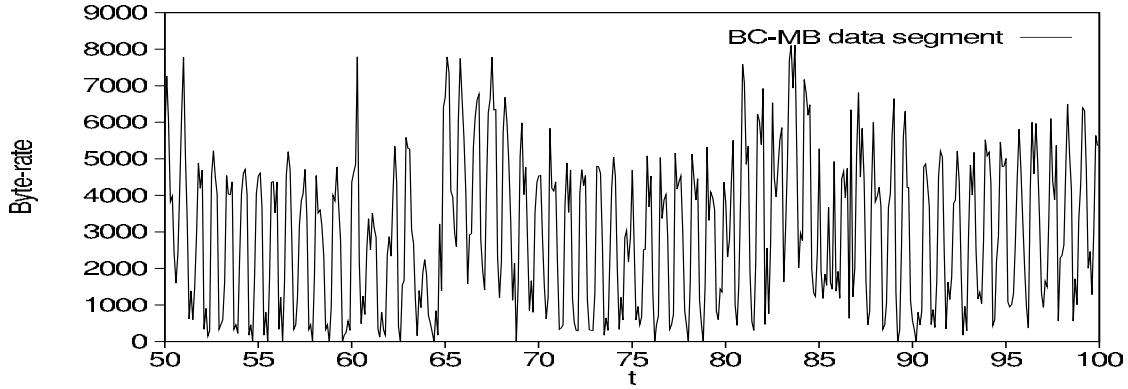


Fig. 7(a): A segment of *BC-MB* trace.

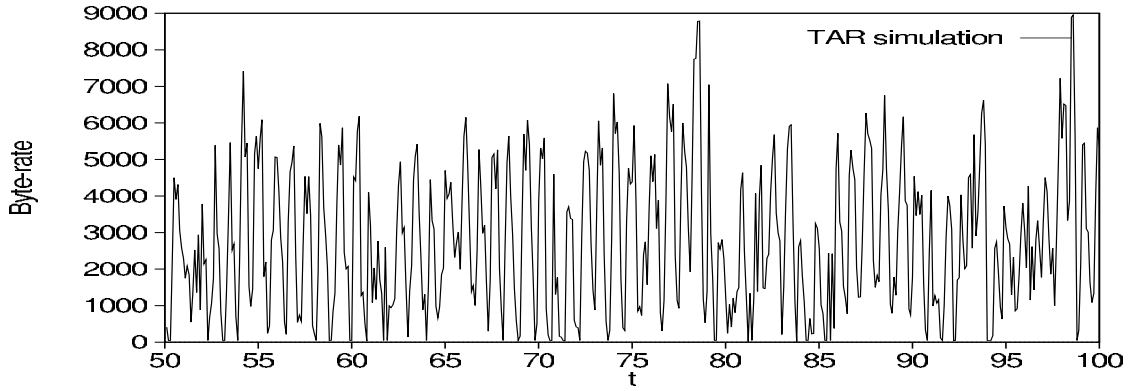


Fig. 7(b): TAR model simulated trace.

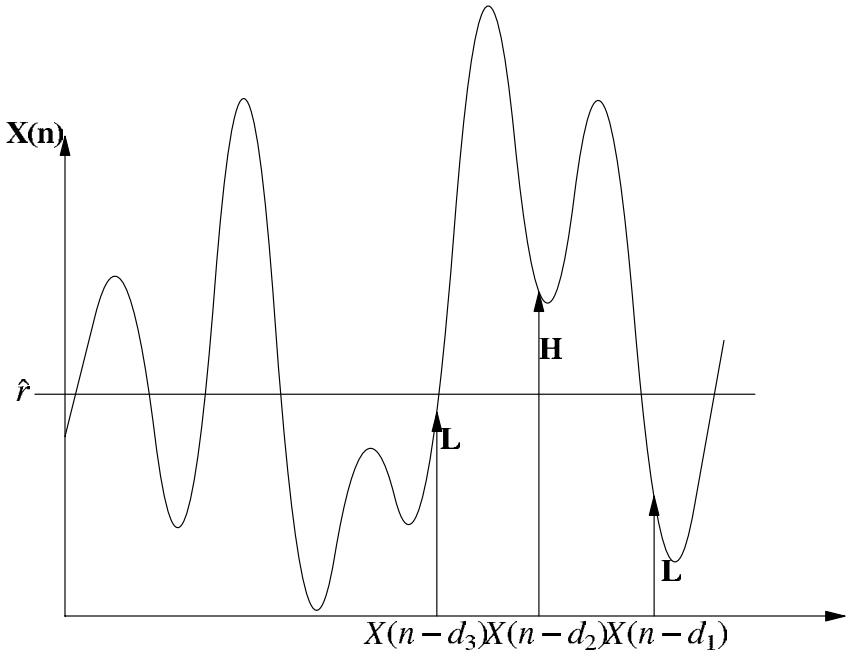
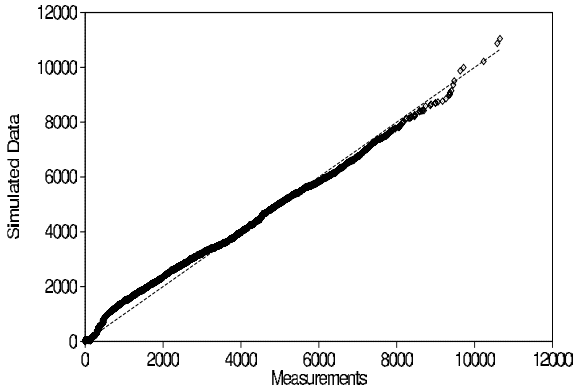
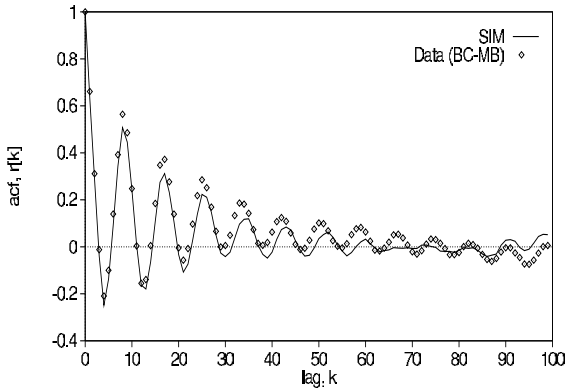


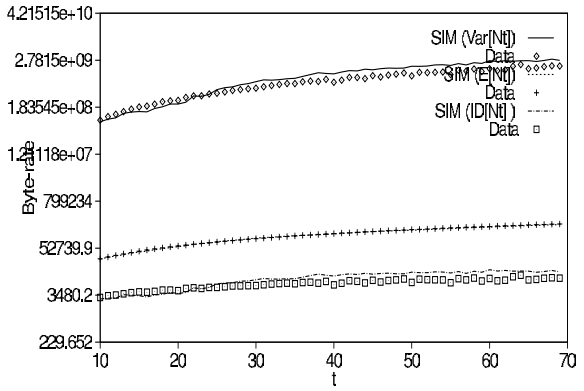
Fig. 8 A schematic of the TAR model parameters for one subregion.



(a) QQ plots

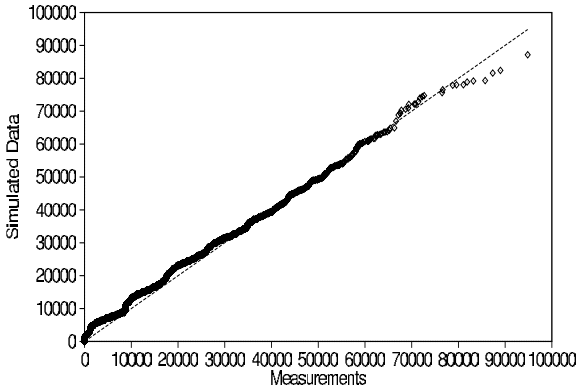


(b) ACFs.

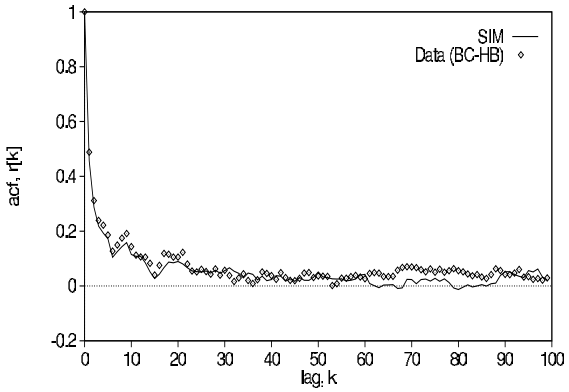


(c) Counting statistics.

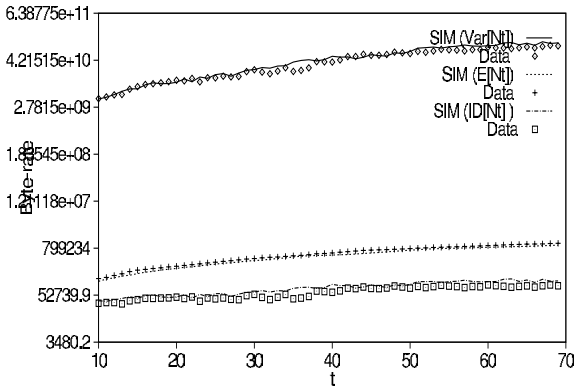
Fig. 9 Comparison between measurements (500 secs) and TAR model simulation for *BC-MB*



(a) QQ plots

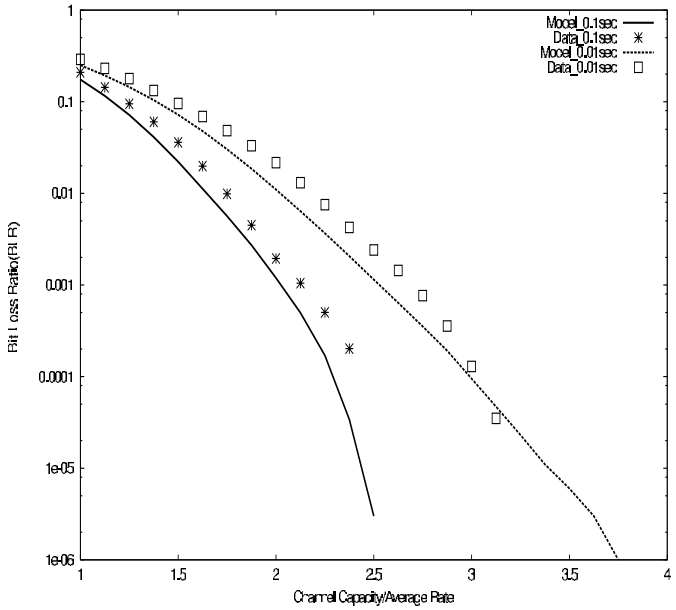


(b) ACFs

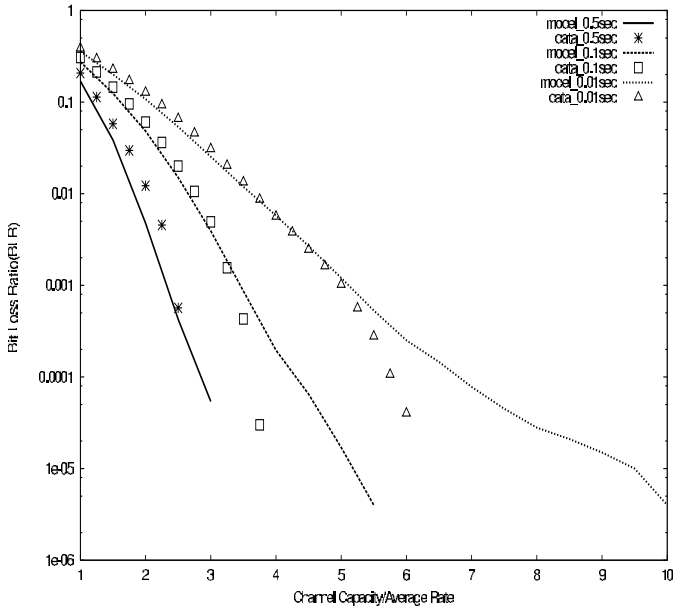


(c) Counting statistics.

Fig. 10 Comparison between measurements (500 secs) and TAR model simulation for *BC-HB*



(a) BLR comparison for *BC-MB* .



(b) BLR comparison for *BC-HB* .

Fig. 11 Bit loss ratios for buffer sizes ranging from 0.01 to 0.5 seconds. Solid lines represent TAR model simulations and symbols depict results from measurements.

Regime R_j	$x(n-d_1)$	$x(n-d_2)$	$x(n-d_3)$
1	L	L	L
2	L	L	H
3	L	H	L
4	L	H	H
5	H	L	L
6	H	L	H
7	H	H	L
8	H	H	H

Table 1: The subregion classification for the *BC-MB* data. L and H denote a low and high states. Ex. Row 3 denotes the constraint $\{x(n-d_1) \in L\ , x(n-d_2) \in H$ and $x(n-d_3) \in L\ \}$.

Threshold	AIC_{total}			
$\hat{r}$	ensemble 1	ensemble 2	ensemble 3	ensemble 4
2200.000	96.36282	75.07628	94.09926	76.19842
2400.000	89.06292	88.52559	80.53676	78.62704
2600.000	89.71453	83.96621	84.24383	94.15655
2800.000	85.07215	**83.84380**	86.25938	90.77125
3000.000	82.40987	100.1935	**81.61975**	82.95047
3200.000	**82.19209**	96.75581	98.33761	**70.82937**
3400.000	89.11894	103.1935	95.15688	84.69061
3600.000	87.23269	101.6886	105.6746	78.82215

Table 2 : $AIC_{total}(\hat{r}, \underline{d})$ estimates for TAR model fitting for three different ensembles of length 300 samples from the *BC-MB* data.

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. W. Willinger, M.S. Taqqu, R. Sherman and D.V. Wilson, "Self-Similarity through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level," in Proc. SIGCOMM'95, p100-113, Cambridge, MA., 1995.

5. M.E. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic, Evidence and Possible Causes," IEEE/ACM Transactions on Networking, **5** , p835-846, December 1997.

6. 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.

7. 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

8. 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.

9. H. Tong, **Threshold Models in Non-linear Time Series Analysis,** Lecture Notes in Statistics, vol. 21, Springer-Verlag, 1983

10. H. Tong, **Non-linear Time Series , A Dynamical System Approach,** Oxford Science Publications, Clarendon Press, Oxford, 1990

11. 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.

12. M.B. Priestley, **Non-linear and Non-Stationary Time Series Analysis,** Academic Press, 1988

13. S.M. Kay, **Modern Spectral Estimation,** Chap. 7, Prentice Hall, Englewood Cliffs, NJ, 1988

14. J.M. Chambers, W.S. Clevelan, B. Kleiner and P.A. Tukey, **Graphical Methods for Data Analysis," The Wadsworth Statistics/Probability Series, Duxbury Press, Boston, 1983**