Predicting Self-Similar Internet Traffic

Majid Ghaderi
ghaderi@bbcr.uwaterloo.ca

July 2002

Abstract

Self-similarity is one of the most important characteristics of the Internet traffic. In this paper, we look at the problem of traffic prediction in the presence of self-similarity and the research in this area. We first explain the concept of self-similarity and introduce short-range and long-range dependencies. Next, we briefly describe a number of short-memory and long-memory stochastic models. Finally, we discuss the problem of modeling the Internet traffic and explain our future study.

1 Introduction

One of the key issues in measurement-based network control is to predict the traffic in the next control time interval based on the online measurements of traffic characteristics. The goal is to forecast future traffic variations as precisely as possible, based on the measured traffic history.

Traffic prediction requires accurate traffic models which can capture the statistical characteristics of actual traffic. If these models do not accurately represent actual traffic, one may overestimate or underestimate network traffic.

Recently, there has been a significant change in the understanding of network traffic. It has been demonstrated in numerous studies that traffic in high-speed networks exhibits self-similarity [1], [2, 3], [4] that can not be captured by previous models, hence self-similar models have been developed.

In this study we investigate self-similarity concept and its relation with long-range and short-range dependence. We review statistical models developed to capture characteristics of the self-similar traffic. This work aims to adapt a suitable model that can capture traffic self-similarity while it has the following properties:

1. Simple: because we are talking about the Internet, we should keep every thing as simple as possible. This model should take as few assumptions about the traffic characteristics as possible,

2. On-line: most of traffic modelling have been done for off-line data. In reality, e.g., network control, we want to use on-line measurements to forecast future. We do not know any thing about the underlying traffic model instead we should estimate the model on-line.

3. Adaptive: this model should adapt to the changing traffic. As we go farther in time, more samples are available. Therefore we will get more knowledge about the traffic characteristics. This model should use new information to improve itself.

Rest of this paper is organized as follows. In section 2 we discuss the self-similarity concept and introduce short-range and long-range dependencies. Section 3 introduce mostly used short-memory and long-memory stochastic models. Finally, section 4 is the conclusion of this study.

2 Self-similarity

There is evidence that traffic is self-similar and fractal in nature. This can be explained by assuming that network workloads are described by power laws. For example, there is considerable evidence that file sizes and web object sizes are described by distributions which decay according to a power law. In other words, if X is the size of the object in bytes, then $P[X > x] = cx^{-a}$.

Let $\{X_t\}, t = 0, 1, 2, \ldots$ be a wide-sense stationary process with mean $E[X_t] = \mu$ and autocorrelation function ρ_k at lag k. For each m, let $\{X_j^{(m)}\}, m = 1, 2, \ldots$ denotes a new time series obtained by averaging the original series $\{X_t\}$ over non-overlapping blocks of size m, i.e.,

$$X_j^{(m)} = \frac{1}{m}(\sum_{l=0}^{m-1} X_{jm+l}) \tag{1}$$

The processes $\{X_j^{(m)}\}$ are also wide sense stationary with mean μ and autocorrelation $\rho_k^{(m)}$. The process $\{X_t\}$ is said to be exactly self similar if $\rho_k^{(m)} = \rho_k$ for all m. In the other word, autocorrelation structure is preserved across different time scales. The process $\{X_t\}$ is said to be asymptotically self similar if $\rho_k^{(m)} \rightarrow \rho_k$, when $m \rightarrow \infty$. Fractional Gaussian noise is an example of an exactly self-similar process and Fractional ARIMA is an example of an asymptotically self-similar process.

Stochastic self-similar processes retain the same statistics over a range of scales, and they satisfy the relation $X_{at} \simeq a^H X_t$ for all $(a > 0)$, where $\simeq$ denotes equality in distribution. This is a very strict one of self-similarity and called *self-similarity with stationary increments*. Process X_t as defined above, is a H-sssi process. Fractional Brownian motion is an example of such processes.

2.1 Short-Range and Long-Range Dependence

Long-range dependence (LRD) can be considered as a phenomenon that current observations are significantly correlated to the observations that are farther away in time. This phenomenon is of particular interest to traffic modelling, since it has been discovered that the Internet traffic posses long-range dependence.

Let ν_m denote the variance of $\{X_j^{(m)}\}$. For large m equation 1 can be approximated by:

$$\nu_m = \nu[2\sum_{k=1}^{m} \rho_k]m^{-1} \tag{2}$$

X_t is said to have a short-range dependence [5], if $\Sigma_k \rho_k < \infty$. Equivalently, ν_m decays to zero proportional to m^{-1}. According to equation 2 this requires that autocorrelation function of X_t decays exponentially to zero. That is, $\rho_k \sim C^k(-1 < C < 1)$.

The process X_t is said to have a long-range dependence [5], if $\Sigma_k \rho_k \to \infty$. Equivalently, ν_m decays at a slower rate than m^{-1}. For example processes in which $\rho_k \sim k^{-(2-2H)}$ for large k. $H(0 < H < 1)$ is the so-called *Hurst* parameter, which is an important quantity used to characterize the LRD. An interesting characteristic of the correlation structure of a long-range dependent process is that ρ_k obeys the well-known *power-law* distribution.

3 Stochastic Traffic Models

Traditional traffic models including Markov and Regression models can only capture short-range dependencies in traffic. We refer to these models as short-memory models. Long-memory models are nontraditional models which are capable to capture long-range dependencies.

3.1 Short-memory Models

We briefly discuss two classes of models: (1) Markov models, and (2) Regression models.

In Markov modelling, the activities of a system can be modelled by a finite number of states. In general, increasing the number of states results in a more accurate model at the expense of increased computational complexity. Markovian property is the common characteristic of these models; *the next state of the system depends only on the current state*. Markov-type models often result in a complicated structure and many parameters when used to model long-range dependent or a mixed process [6].

Regression models define explicitly the next random variable in the sequence by previous ones within a specified time window and a moving average of a white noise.

Define the lag operator B as $BX_t = X_{t-1}$, where $B^s X_t = X_{t-s}$. Also assume that Δ denotes the differencing operator, i.e., $\Delta X_t = X_t - X_{t-1}$, equivalently $\Delta^d = (1-B)^d$ can be expressed using the binomial expansion

$$(1-B)^d = \sum_{k=0}^{d} \binom{d}{k} (-1)^k B^k$$

where,

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

We also define polynomials $\phi(B)$ and $\theta(B)$ as follows:

$$\phi(B) = (1 - \phi_1 B - \ldots - \phi_p B^p)$$

$$\theta(B) = (1 - \theta_1 B - \ldots - \theta_q B^q).$$

1. **Autoregressive Model**:
   The autoregressive model [7] of order p, denoted as AR(p), has the form $\phi(B)X_t = \varepsilon_t$, where ε_t is white noise. In this model variable X_t is regressed on previous values of itself. AR models can be used to model stationary time series (time series that have a constant mean) and if all the roots of $\phi(B)$ lie outside the unit circle, then it is invertible (can be written in the form $X_t = \phi^{-1}(B)\varepsilon_t$).

2. **Autoregressive Moving Average Model**:
   An ARMA(p, q) [7] has the form $\phi(B)X_t = \theta(B)\varepsilon_t$. Note that $\theta(B)\varepsilon_t$ is the moving average part of this model. These models have a great flexibility in modeling time series. In practice, it is frequently true that adequate representation of actual time series can be obtained with models, in which p and q are not greater than 2 and often less than 2.

3. **Autoregressive Integrated Moving Average Model**:
   ARIMA(p, d, q) [7] is an extension to ARMA(p, q). It is obtain by allowing the polynomial $\phi(B)$ to have d roots equal to unity. The rest of the roots lie outside the unit circle. ARIMA(p, d, q) has the form $\phi(B)\Delta^d X(t) = \theta(B)\varepsilon_t$. ARIMA is used to model nonstationary processes. Note that $\Delta^d X_t = (1-B)^d X_t = \phi^{-1}(B)\theta(B)\varepsilon_t$ and accordingly $X_t = (1-B)^{-d}\phi^{-1}(B)\theta(B)\varepsilon_t$.

   In this expansion X_t is regressed to sum (integration) of infinite noise variables. In some cases it is possible that the original series X_t is not stationary but its increments $X_t - X_{t-d} = (1-B)^d X_t$ exhibit stationary characteristics. This is the philosophy behind the inclusion of differencing operator Δ in this model.

3.2 Long-memory Models

1. **Fractional Brownian Motion (fBm)**:
   Brownian motion [5] is a stochastic process, denoted by $\mathrm{Bm}_t, t \geq 0$. It is characterized by the property that increments $\mathrm{Bm}_{t+\tau} - \mathrm{Bm}_t$ are normally distributed with mean 0 and variance $\sigma^2 t$. The fractional Brownian motion fBm_t is a self-similar process with $1/2 < H < 1$. Fractional Brownian motion differs from the Brownian motion by having increments with variance $\sigma^2 t^{2H}$.

2. **Fractional Gaussian Noise (fGn)**:
   Although fBm is useful for theoretical analysis, its increments process (for finite increment τ),
   $$\mathrm{fGn}(t) = \mathrm{fBm}(t\tau) - \mathrm{fBm}((t-1)\tau),$$
   known as fractional Gaussian noise, is often more useful in practice. While fBm is not stationary, fGn is stationary. The autocorrelation function of this process is
   $$\rho_k = 1/2[(k+1)^{2H} - 2k^{2H} + (k-1)^2 H].$$

3. **Fractional ARIMA Model (FARIMA)**:
   Fractional ARIMA proposed by Hosking [8] in 1980 is the natural extension of the ARIMA process when we allow real values for parameter d. X_t has a stationary invertible FARIMA(p, d, q) process if:
   $$\phi(B)\Delta^d X_t = \theta(B)\varepsilon_t$$

where d is a real number ($-1/2 < d < 1/2$), and where $\phi(B)$ and $\theta(B)$ are stationary AR and invertible MA polynomials. Thus, X_t is a long-memory process if ($0 < d < 1/2$) and a short-memory process if $d = 0$. This model has been extensively used in network traffic modeling [9].

4. **Generalized ARMA Model (GARMA)**:
GARMA models [10] are the generalization of all the regression models. They can be used to model both short-range and long-range dependence in a time series. In addition, they can be used to model the cyclical patterns of a time series with fewer parameters than ARMA models. The GARMA(p,q) model of a process X_t is defined as

$$\phi(B)(1 - 2\eta B + B^2)^d X_t = \theta(B)\varepsilon_t$$

where $-1/2 < d < 1/2$ and $(-1 < \eta < 1)$. The term $(1 - 2\eta B + B^2)^d$ is the Gegenbauer polynomial which can be expanded using the power series expansion.

4 Conclusion

In this survey, we investigated the self-similarity of the Internet traffic. We discussed the concept of short-range and long-range dependencies intuitively and formally. Then stochastic models which are used for traffic modeling, including both short-memory and long-memory models, introduced. While there is certainly much more in the area of stochastic traffic modeling than what we have presented in this paper, we have focused on providing a comprehensive overview of self-similarity and related traffic models.

The most important criterion in choosing a predictor is the accuracy. Analysis of real traffic traces indicates that the Hurst parameter H rarely exceeds 0.85 [1], [3], [9], [10]. Although models such as fGn, FARIMA and GARMA can suitably fit to these traces, but they are computationally very complex. Particularly there is not any known method for on-line modeling based on these models. For the future work, we will implement predictors based on mean square error and fractional models such as fGn, FARIMA, and GARMA to see how exactly they can fit to the real traffic traces and compare their accuracy. Mean square error predictors are very easy to implement and if they achieve a good accuracy in comparison with fractional models then they will be more relevant for practical applications involving traffic prediction.

References

[1] W. Leland *et al.*, "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Trans. Networking*, February 1994.

[2] J. Beran *et al.*, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEEE Trans. Commun.*, February 1995.

[3] M. Crovella and A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes," *IEEE/ACM Trans. Networking*, December 1997.

[4] M. Grossglauser and J.-C. Bolot, "On the Relevance of Long-Range Dependence in Network Traffic," *IEEE/ACM Trans. Networking*, October 1999.

[5] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. Springer-Verlag, 2nd ed., 1991.

[6] A. T. Andersen and B. F. Nielson, "An Application of Superpositions of Two State Markovian Sources to the Modeling of Self-similar behavior," in *Proc. IEEE INFOCOM'1997*, 1997.

[7] G. E. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 2nd ed., 1976.

[8] J. R. M. Hosking, "Fractional Differencing," *Biometrika*, April 1981.

[9] Y. Shu *et al.*, "Prediction-Based Admission Control Using FARIMA Models," in *Proc. IEEE ICC'2000*, 2000.

[10] R. Ramachandran and V. R. Bhethanabotla, "Generalized Autoregressive Moving Average Modeling of the Bellcore Data," in *Proc. IEEE LCN'2000*, 2000.