1-1-2007

A Survey of Performance Evaluation and Control for Self-Similar Network Traffic

Karim Mohammed Rezaul

Vic Grout

Glyndwr University, v.grout@glyndwr.ac.uk

Follow this and additional works at: http://epubs.glyndwr.ac.uk/cair

Part of the Computer and Systems Architecture Commons, Digital Communications and Networking Commons, Hardware Systems Commons, and the Systems and Communications Commons

Recommended Citation


This Conference Paper is brought to you for free and open access by the Computer Science at Glyndwr University Research Online. It has been accepted for inclusion in Computing by an authorized administrator of Glyndwr University Research Online. For more information, please contact d.jepson@glyndwr.ac.uk.
A Survey of Performance Evaluation and Control for Self-Similar Network Traffic

Abstract
This paper surveys techniques for the recognition and treatment of self-similar network or internetwork traffic. Various researchers have reported traffic measurements that demonstrate considerable burstiness on a range of time scales with properties of self-similarity. Rapid technological development has widened the scope of network and Internet applications and, in turn, increased traffic volume. The exponential growth of the number of servers, as well as the number of users, causes Internet performance to be problematic as a result of the significant impact that long-range dependent traffic has on buffer requirements. Consequently, accurate and reliable measurement, analysis and control of Internet traffic are vital. The most significant techniques for performance evaluation include theoretical analysis, simulation, and empirical study based on measurement. In this research, we discuss existing and recent developments in performance evaluation and control tools used in network traffic engineering.

Keywords
Self-similarity, ACF, LRD, Heavy-tailed distribution, Hurst parameter.

Disciplines
Computer and Systems Architecture | Digital Communications and Networking | Hardware Systems | Systems and Communications

Comments
Copyright © 2007 Glyndwr University. This paper was presented at the Second International Conference on Internet Technologies and Applications (ITA 07), Wrexham, North Wales, UK, 4-7 September 2007, pp514-524. Details of the conference can be found at http://www.glyndwr.ac.uk/computing/research/ita07/index.html The full proceedings of the conference can be found at http://www.lulu.com/product/paperback/proceedings-of-the-second-international-conference-on-internet-technologies-and-applications/1333637

This conference paper is available at Glyndŵr University Research Online: http://epubs.glyndwr.ac.uk/cair/52
A SURVEY OF PERFORMANCE EVALUATION AND CONTROL FOR SELF-SIMILAR NETWORK TRAFFIC

Karim Mohammed Rezaul and Vic Grout
Centre for Applied Internet Research (CAIR), University of Wales, NEWI, Wrexham, UK
karim@cari-uk.org
vic@cair-uk.org

ABSTRACT

This paper surveys techniques for the recognition and treatment of self-similar network or internetwork traffic. Various researchers have reported traffic measurements that demonstrate considerable burstiness on a range of time scales with properties of self-similarity. Rapid technological development has widened the scope of network and Internet applications and, in turn, increased traffic volume. The exponential growth of the number of servers, as well as the number of users, causes Internet performance to be problematic as a result of the significant impact that long-range dependent traffic has on buffer requirements. Consequently, accurate and reliable measurement, analysis and control of Internet traffic are vital. The most significant techniques for performance evaluation include theoretical analysis, simulation, and empirical study based on measurement. In this research, we discuss existing and recent developments in performance evaluation and control tools used in network traffic engineering.

KEYWORDS

Self-similarity, ACF, LRD, Heavy-tailed distribution, Hurst parameter.

1. INTRODUCTION

Self-similar processes were first identified by Kolmogorov [1] in 1941. These processes were then brought to the attention of statisticians by Mandelbrot and colleagues in the late 1960’s and early 1970’s [2, 3, 4, 5]. Mandelbrot outlined the important dynamic characteristic of fractional Gaussian noises (fGn), characteristics which could be quantified with the Hurst exponent. The ordering of such series determined the dynamic measures (i.e. long-range dependence or LRD). Self-similar and LRD characteristics of Internet traffic have attracted the attention of researchers since 1994, when it was discovered that some aspects of LAN traffic exhibits self-similar, rather than Poisson, behaviour [6].

LRD is of great significance in traffic engineering problems such as measurement, queuing strategy buffer sizing and admission and congestion control. In [7] it is shown that consequences of LRD are packet and application level delays that cause a heavy-tailed distribution. Transmission Control Protocol (TCP) estimates round trip timer values from peer acknowledgements and, as a result, congestion appears more frequently when maintaining impulsive behaviour with an increase in load. The influence of LRD properties on delay at packet and application level is reported in [8]; metrics of network performance, such as throughput, packet loss, latency and buffer occupancy levels, are affected by the presence of LRD phenomenon across many types of networks. The impact of LRD on quality of service (QoS) is analysed in [9] showing that, the greater the LRD, the lower the QoS.

The LRD property of traffic fluctuations has important implications on the performance, design and dimensioning of the network [6]. A simple, direct parameter, characterizing the degree of LRD, is the Hurst parameter. The Hurst exponent (or Hurst parameter, H), which more than a half-century ago was proposed for the analysis of reservoir long-term storage capacity [10], is used today to measure the intensity of LRD in network traffic. A number of methods have been proposed to estimate the Hurst parameter. Some of the most popular include aggregated variance time (V/T) [11], Rescaled-range (R/S) [6, 10] and the Higuchi and wavelet-based methods [12, 13] although there are many others. In all these methods, H is calculated by taking the slope from a log-log plot. Over time, the wavelet-based Hurst parameter has acquired popularity in estimating LRD traffic. The intensity of
long-range dependence can be measured for file or document size [14], packet-count [15], interarrival time [16], frame size [17], connection size [18], packet length [19], byte-count [6], and bit or byte rate [20] amongst others.

A number of factors, such as a slow start phase of the congestion window, packet losses, acknowledgment of TCP traffic and multiplexing of packets at the bottleneck rate, can cause either short- or long-term burstiness in the behaviour of TCP flow [21]. The research in [22] investigates how various versions of TCP congestion control affect network performance when traffic is bursty. TCP represents the dominant transport protocol of the network (e.g. Internet), which contributes to the propagation of self-similarity. It is shown in [23] that TCP itself inherits self-similarity when it is combined with self-similar background traffic in a bottleneck buffer through the transform function of the linear system.

This paper is organised as follows. Section 2 describes the relationship between self-similarity, long-range dependence and the autocorrelation function (ACF). Section 3 discusses the performance of the estimators and their reliability. Section 4 explains heavy-tailedness in traffic patterns. Section 5 considers the issues involved in controlling network traffic.

2. THE RELATIONSHIP BETWEEN SELF-SIMILARITY, LRD AND ACF

A phenomenon that is self-similar looks or behaves the same when viewed at different degrees of magnification or different scales on a given dimension and is bursty over all time scales. Figure 1 depicts self-similarity observed in natural phenomena. The bottom left shows a higher magnification of a small segment of an image that demonstrates self-similarity. This type of self-similarity at all scales characterises the fractal nature of an image. Thus, self-similar data have a common property with fractals: when we zoom in on part of the traffic generated by the data, we observe the same structure. Many successive zooms show the same result [24] as shown in the bottom right.

Self-similarity is the property of a series of data points to retain a pattern or appearance regardless of the level of granularity used and is the result of LRD in the data series. If a self-similar process is bursty on a wide range of timescales, it may exhibit LRD. Often lagged autocorrelations are used in time series analysis for empirical stationary tests. Self-similarity manifests itself in the time series of arrivals’ LRD (i.e., long memory). The evidence of very slow, linear decay in the sample lag ACF indicates nonstationary behaviour [25]. Long-range-dependence means that all the values at any time are correlated in a positive and non-negligible way with values at all future instants.
A continuous time process $Y = \{Y(t), t \geq 0\}$ is self-similar if it satisfies the following condition:

$$Y(t) = a^{-H} Y(at), \quad \forall a > 0, \quad \text{and} \quad 0 < H < 1$$

(1)

[26] where H is the index of self-similarity (the Hurst parameter) and the equality is in the sense of finite-dimensional distributions.

The stationary process X is said to be an LRD process if its ACF is non-summable [27], that is if

$$\sum_{k=-\infty}^{\infty} \rho_k = \infty$$

Details of how the ACF decays with k are of interest because the behaviour of the tail of the ACF completely determines its summability. From [6], X is said to exhibit long-range dependence if

$$\rho_k \sim L(t) k^{-(2-2H)}, \quad \text{as } k \to \infty$$

(2)

where $\frac{1}{2} < H < 1$ and $L(.)$ slowly varies at infinity, i.e., $\lim_{t \to \infty} \frac{L(\alpha)}{L(t)} = 1$, for all $\alpha > 0$.

Equation (2) implies that the LRD is characterized by an ACF that decays hyperbolically rather than exponentially fast.
LRD processes then are characterised by a slowly decaying covariance function that is non-ssummable. Numerous studies [6, 28, 29] of different types of data networks (LANs & WANs) and services/applications (ATM, Frame Relay, WWW, etc.) have shown that aggregated traffic at packet level exhibits slowly decaying autocorrelation that leads to LRD. The periodic nature of the traffic leads to high and slowly decaying autocorrelation [30]. When network performance is affected by LRD, data are correlated over an unlimited range of time lags and this property results in a scale invariance phenomenon. Then no characteristic time scale can be identified in the process: they are all equivalent for describing its statistics - the part resembles the whole and vice versa.

In Figure 2, the top left plot illustrates typical LRD and SRD (short-range dependent) processes. The top right plot shows traffic burstiness for real data. The ACF at the bottom confirms that the traffic is LRD. Clearly, traffic persists for more than 9 seconds. These real data were taken from the UNC archive [31] with trace collections obtained from a Gigabit Ethernet link. The data set considered here are number of TCP bytes per 10 ms bin. The sample length (N) considered for the dataset is 10,000.

3. FINDING RELIABLE AND ROBUST ESTIMATORS

Several studies report problems with existing estimators, such as the unreliability of the wavelet-based Hurst parameter [32, 33, 34, 35], R/S analysis [36, 37], aggregated V/T analysis [38, 39, 40], Whittle estimator [33, 41] and Periodogram analysis [39]. The estimated Hurst parameters from the wavelet analysis, aggregated variance and Whittle method have been presented [33] for both real and synthetic data. It is observed that the Whittle and wavelet methods overestimate the degree of self-similarity (i.e., produce H>1). In [34] the advantages and limitations of the wavelet estimators are explored. Here it is found that a traffic trace with a number of deterministic shifts in the mean rate results in a steep wavelet spectrum, which leads to overestimating the Hurst parameter.
In [42] it is reported that the interpretation of the estimated Hurst parameter is problematic in practice. Numerous studies have shown that existing estimators provide an unreliable estimate of the Hurst parameter (H). Having a reliable estimator can yield a good insight into traffic behaviour and should eventually lead to improved traffic engineering. Figure 3 compares the reliability of some popular estimators. Here the 20 highest values of the Hurst parameter provided by different estimators are shown while the self-similar sequences were generated by a fractional autoregressive integrated moving average (FARIMA) process for a particular Hurst parameter, H [43].

It is possible to derive wrong conclusions and wrong models when measuring the intensity of the LRD with unreliable estimators. In [44, 45] an estimator is introduced called the Hurst Exponent from the Autocorrelation Function (HEAF) and it is shown that (and why) lag 2 in HEAF (i.e. HEAF (2)) is considered when estimating LRD of network traffic. [43] considers the robustness of HEAF(2) when estimating the Hurst parameter of data traffic (e.g. packet sequences) with outliers and also the reliability of HEAF(2). Also, based on the comparison of simulation experiments shown [43, 44, 45] for both fGn and FARIMA (0, d, 0) processes, it is evident that HEAF(2) is a stable method that quantifies the reliable degree of LRD. Through its simplicity, robustness and reliability, HEAF(2) can be used to estimate the intensity of LRD in real time network traffic.

4. HEAVY-TAILEDNESS IN TRAFFIC PATTERNS

Many researchers have discussed the effects of heavy-tailedness in network traffic patterns and shown that Internet traffic flows exhibit characteristics of self-similarity that can be explained by the heavy-tailedness of the various distributions involved. The properties of heavy-tailed distributions are qualitatively different to commonly used memoryless distributions such as the exponential, normal or Poisson distributions. [46] concludes that such exponentiality assumptions mislead exploration of the presence of heavy-tailed distributions. The condition of self-similarity is that the ACF of the time-series declines as a power-law, leading to positive correlations among widely separated observations [47]. In the Internet, heavy-tailed distributions have been observed in the context of traffic
characterization. The distributions having infinite variances are called heavy-tailed and the weight of their tails is determined by a parameter called the tail index, $\alpha < 2$ [48].

Self-similarity and heavy-tailedness are of great importance for network capacity planning purposes, in which researchers are interested in developing analytical methods for analysing traffic characteristics. The goal of traffic characterisation is to determine the nature of the traffic and develop tractable models that capture the important properties of data, which can eventually lead to accurate performance prediction. The uses of traffic characterisation include network planning, design, capacity management, performance prediction, real-time traffic management and network control. Designers of computing and telecommunication systems are increasingly interested in employing heavy-tailed distributions to generate workloads for use in simulation, although simulations employing such workloads may show unusual characteristics.

The performance of several estimators of the tail index ($\alpha$) for heavy-tailed Internet traffic have been studied in [49]. A summary of results is shown in Figure 4. In most cases, the moment estimator due to an unstable region is observed in the graph. The Hill plot, static qq plot and LLCD plot show a good level of agreement when estimating the index from graphs. The results show that there are infinite variances (i.e. $\alpha < 2$) observed in the traffic, which is indicative of the existence of heavy-tailedness in the Internet traffic.

[50] proposes some analytical models based on the Empirical Distribution Function (EDF) statistics, which can characterise web traffic. The analyses show that the Weibull (three parameters) and generalized Pareto models, with the experimental results, are the most suitable to approximate the traffic, as shown in Figure 5. In addition, the generalized Pareto model is more suitable for analysing traffic behaviour than the simple Pareto model in terms of heavy-tailedness. Hence, as an efficient analytical tool, the generalised Pareto model can be used for identifying a heavy-tailed nature based on samples from web traffic.

![Hill plot, static qq plot and LLCD plot](image)

Figure 4. Estimation of tail index by Hill plot ($\alpha = 0.76$), static-qq plot ($\alpha = 0.74$) and LLCD plot ($\alpha = 0.80$) [49]
5. CONTROLLING NETWORK TRAFFIC

Most research in this area concentrates on modelling network traffic rather than controlling it. However, controlling network traffic should not be overlooked as it can help reduce the network load and lead to the improvement of QoS in future network performance. [51] introduces a novel algorithm, called CoLoRaDe, to control the intensity of LRD traffic. Experimental results show that CoLoRaDe is capable of reducing the LRD of packet sequences received at the router buffer before they are transmitted to the core network (i.e. the Internet). As the main function of the CoLoRaDe algorithm is to reduce the LRD of packet traffic, it can contribute in reducing the network load, leading to an improvement in QoS for the Internet of the future.

A number of factors, such as a slow start phase of the congestion window, packet losses, ack-compression of TCP traffic and multiplexing of packets at the bottleneck rate, can cause either short- or long-term burstiness in TCP flow [52]. [53] investigates how various versions of TCP congestion control affect network performance when traffic is bursty. A significant adverse impact on network performance is shown, attributable to traffic self-similarity and, while throughput declines gradually as self-similarity increases, queueing delay increases more drastically. TCP represents the dominant transport protocol of the Internet, which contributes to the propagation of self-similarity [23].

Various researchers have reported that traffic measurements demonstrate considerable burstiness on several time scales, with properties of self-similarity. Bursty traffic can affect the QoS for all traffic on the network by introducing inconsistent latency. It is easier to manage the workloads under less bursty (smoother) conditions. One of the major drawbacks of TCP/IP is the lack of true QoS functionality. QoS in networks, in simple terms, is the ability to guarantee and limit bandwidth appropriately for certain services and users. Traffic shaping is an attempt to control network traffic in order to optimize, attempt to optimize or guarantee performance, low-latency or bandwidth, and deals with concepts of classification, queue disciplines, enforcing policies, congestion management. QoS and fairness. [54] introduces a novel algorithm, BPTraSha, to control the bursty nature of network traffic. Experimental results show that the BPTraSha algorithm is capable of smoothing out the bursty nature of traffic packets received at the router buffer before they are transmitted to the core network. Figure 6 shows how the bursty nature of traffic are smoothed out by BPTraSha.
6. CONCLUSION

Network performance evaluation is important for assessing the effectiveness of traffic methods and for monitoring and verifying compliance with network performance goals. Results from performance evaluation can be used to identify existing problems, guide network re-optimization and aid in the prediction of potential future problems. Research towards finding and improving suitable tools which may help to characterise various types of network traffic is consequently, and obviously, vital. It is particularly important to understand the link between the self-similarity and long-range dependence of network traffic and the performance of the networks because such characterization can be potentially applied for essential control purposes such as traffic shaping, load balancing and other strategies of the future.

REFERENCES


