On the Self-Similar Nature of Workstations and WWW Servers Workload

Olivier Richard and Franck Cappello

LRI, Université Paris-Sud, 91405
Orsay, France
(auguste,fci)@lri.fr.

Abstract. This paper presents a workload characterization for workstations, servers and WWW servers. Twelve data sets built from standard UNIX tools and from access.log files are analyzed with three different time scales. We demonstrate that the workload of these resources is statistically self-similar in the periods of irregular activity.

1 Introduction

The resource sharing in a network of workstations and in Internet is the aim of a large number of research projects. Fundamental to this goal is to understand the initial workload characteristics of each resource.

Most of the mechanisms (CONDOR [1] and GLUNIX [2]) for workload distribution try to detect the periods when the user does not use its workstation. Nevertheless, even when the user uses its workstation, the initial workload may stay very low. This under-exploitation provides the potential to gather unexploited resources for more agressive users. This goal requires to study carefully the initial workload of a workstation during the user activity. The workload characterization of WWW servers is more recent. In [3], the self-similarity of WWW server workload is investigated. The automatic workload distribution of the WWW requests on mirror sites is an open issue.

We use a global approach to characterize the complexity of the workload. This approach is close to the one used for network traffic analysis [4] [5].

2 Method for the Workload Study

More than a hundred of resources are connected in our network of workstations. The machines are used for numerical processing, some heavy simulations and student and researcher works (compiler, mailer, browser, etc.). The highly computing applications saturate the computer and lead to a typical flat plot for the workload. The server workload is typically much more irregular. The workstation workload is a composition of the previous ones: null during the night, sometimes irregular and sometimes saturated during the day.

We have also analyzed two WWW servers: our lab WWW server, called LRI and the Ethernet backbone of the computing research department of the

Virginia University of technology (USA). The trace for the LRI server starts at October 31, 1995 and finishes at August 4, 1997. For the second WWW server, only the external requests at *.cs.vt.edu* domain have been recorded during a 38 days period. We analyze the hits and the transfered bytes per time.

We choose to use three time scales of analysis because the event granularities are different in a workstation, a server and a WWW server. The statistical analysis, requires about 100.000 events to give results with an acceptable quality. This number corresponds to 50 days for the minute time scale (WWW server workload analysis). It represents a daily period for the 1 second time scale (workstation workload analysis). To understand the workstation and server micro-workloads we have used a 5 milliseconds time scale (100.000 events represent 8 minutes).

The workload measurement of the servers and workstations has been done with the *vmstat* UNIX command (average percentage of the CPU availability). We have used a method based on a snoopy process with low priority to measure the workload with the 5 millisecond resolution. The accesses to the WWW servers have been recorded in the *access.log* file.

3 Self-similarity

Intuitively, a self-similar signal may be observed graphically as presenting invariant features for differents time scales. As in [4], we consider the workload signal as a wide-sense stationary stochastic process (stationary up to order 2). In such process, the mean (μ) and the variance (σ^2) stay constant over time.

Let $X(t)$ be such a stochastic process. $X(t)$ has an auto-covariance function, $R(\tau)$, and an autocorrelation function $\rho(\tau)$ of the form $R(\tau) = E[(X(t) - \mu)(X(t+\tau) - \mu)]$ and $\rho(\tau) = R(\tau)/R(0) = R(\tau)/\sigma^2$. We assume that $X(t)$ is of the form $\rho(\tau) \to \tau^{-\beta} L(\tau)$, when $\tau \to \infty$ (1), where $L(\tau)$ is slowly varying at infinity. (examples of such function are: $L(\tau) = const$, $L(\tau) = log(\tau)$).

Let $X^{(m)}$ denotes a new time series obtained by averaging the original series X in non-overlapping blocks of size m. That is: $X^{(m)} = (1/m)(X_{tm-m+1} + X_{tm-m+2} + \ldots + X_{tm})$. Let $\rho^{(m)}(\tau)$ denotes the autocorrelation function of $X^{(m)}$. If the aggregated process $X^{(m)}$ has the same autocorrelation structure as X, the process X is called exactly second order self-similar with self-similarity parameter $H = 1 - \beta/2$, i.e., $\rho^{(m)}(\tau) = \rho(\tau)$. X is called asymptotical second order self-similar with self-similarity parameter $H = 1 - \beta/2$, if we assume $\rho^{(m)}(\tau) \to \rho(\tau)$ for large m and τ.

In the next section we use three methods to test and characterize the self-similarity of a stochastic process: the variance-time graphic of the $X^{(m)}$ processes, the R/S analysis [4] and Whittle estimator [6]. All methods give an estimate of H (the self-similarity parameter). H larger than 1/2 and lower that 1 suggests the self-similarity of the signal. A rigorous introduction to the self-similar phenomenon can be found in [4].

4 Results

The data sets size is about 70.000 measurements. Eight machines have been closely analyzed in the network of workstations. The table 4 gathers the estimates of H with the three methods for the observed machines in the network of workstations. For all machines, H is larger than 1/2 and lower than 1 and β is between 0 and 1. So, together the 3 methods suggest that self-similar stochastic processes may be used to represent closely these machine workloads.

Table 1. H estimates for the CPU workload signals (1 second and 5 ms time scales), variance analysis of the $X^{(m)}$ processes, R/S analysis and Whittle estimates

	1 second time scale				5 ms time scale			
	$\beta(VAR)$	$H(VAR)$	$H(R/S)$	$H(W)$	$\beta(VAR)$	$H(VAR)$	$H(R/S)$	$H(W)$
Sun1	0.41	0.79	0.69	0.92	-	-	-	-
Sun2	0.29	0.85	0.64	-	0.31	0.84	0.78	-
Sun3	0.21	0.89	0.57	0.86	0.49	0.75	0.66	0.84
Sun4	0.48	0.75	0.64	0.99	-	-	-	-
Sun5	-	-	-	-	0.68	0.65	0.89	0.91
Sun6	-	-	-	-	0.34	0.82	0.84	0.81
HP1	0.30	0.84	0.79	0.86	-	-	-	-
HP2	0.37	0.81	0.80	0.79	-	-	-	-

The results of the estimates for the H parameter for the WWW servers with the various methods are presented in table 4. The first three raws present the estimates for H and β for three different segments of the LRI trace. The raw labeled BR gives the results for the Ethernet backbone of the computer science department (Virginia University). The values of H and $beta$ for these traces indicates that both signals (hits per minute and transmitted bytes per minute) might be closely represented by self-similar processes.

Table 2. H estimates of hits per second and bytes per second for the WWW servers. The method used for each estimate is indicated between parentheses.

	Hits per second				Bytes per second			
	$\beta(VAR)$	$H(VAR)$	$H(R/S)$	$H(W)$	$\beta(VAR)$	$H(VAR)$	$H(R/S)$	$H(W)$
hit09	0.50	0.74	0.71	0.86	0.64	0.68	0.71	0.74
hit28	0.62	0.68	0.74	0.78	0.62	0.68	0.76	0.72
hit31	0.64	0.67	0.78	0.75	0.75	0.62	0.72	0.61
BR	0.43	0.78	0.88	-	0.63	0.68	0.83	-

5 Analysis of Results and Conclusion

We have shown that self-similar stochastic processes correspond to the workload of these resources (in the periods of irregular activity).

A more accurate confidence interval for the estimates of H may be obtained with the Whittle estimator applied to larger data sets or with other methods such as the wavelet methods. [7] has recently shown that some network traffics were multi-fractal instead of mono-fractal (H is varying with time).

Our result about the self-similar nature of the WWW server workload is contrary to the one given in [3]. In their data set, there are several defects : zero hit for several hours, or very punctual high volume transfers. Removing these punctual defects changes the conclusions and suggest that the self-similarity is not as rare as the authors have suggested and that high perturbations may alterate the results of the self-similarity tests.

The knowledges about self-similarity in WWW servers are getting wider. In [8], the authors propose an explanation for the nature of client site WWW traffic. The article [9] provides an interesting result: superposition of self-similar processes yields to a self-similar process with fractional Gaussian noises. [4] has presented some consequences of traffic self-similarity on communication network engineering. There are drastic implications on modeling the individual source, on the notion of "burstiness" and on the congestion management.

The methods and the results obtained in the field of communication networks may be used as the bases for further works in the workload management area. For example, the estimate of H may be used to generate synthetic traces from the fractional Gaussian noise model [10] or from the fractional ARIMA (p, d, q) processes [11]. The synthetic traces would represent the initial workload of a workstation, a server or a WWW server.

References

1. M. J. Litzkow, M. Livny, and M. W. Mutka. Condor : A hunter of idle workstations. In *8th International Conference on Distributed Computing Systems*, pages 104–111, Washington, D.C., USA, June 1988. IEEE Computer Society Press.
2. Amin M. Vahdat, Douglas P. Ghormley, and Thomas E. Anderson. Efficient, portable, and robust extension of operating system functionality. Technical Report CSD-94-842, University of California, Berkeley, December 1994.
3. Martin Arlitt and Carey L. Williamson. Web server workload characterization: The search for invariants. In *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer SYstems*, volume 24,1 of *ACM SIGMETRICS Performance Evaluation Review*, pages 126–137, New York, May23–26 1996. ACM Press.
4. Will E. Leland, Murad S. Taqqu, Walter Willinger, and Daniel V. Wilson. *On the Self-Similar Nature of Ethernet Traffic (Extended Version). IEEE/ACM Transactions on Networking*, 1994.
5. Vern Paxson and Sally Floyd. Wide-area traffic: the failure of Poisson modeling. *ACM SIGCOMM 94*, pages 257–268, August 1994.
6. Jan Beran. *Statistic for Long-Memory Processes*. Chapman and Hall, 1994.
7. R.H. Riedi and J. Lvy Vhel. Multifractal properties of tcp traffic: a numerical study. Technical Report 3129, INRIA, 1997.
8. Mark Crovella and Azer Bestavros. Self-similarity in World Wide Web traffic: Evidence and causes. In *Proceedings of the ACM SIGMETRICS International Conference on Measurement and Modeling of Computer SYstems*, volume 24,1 of *ACM SIGMETRICS Performance Evaluation Review*, pages 160–169, New York, May23–26 1996. ACM Press.
9. K. Krishnan. A new class of performance results for a fractional brownian traffic model. *Queueing System*, 22:277–285, 1996.
10. B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motion, fractional noises and applications. *SIAM Review*, 10(4):422–437, October 1968.
11. J. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. *Time Series Analysis*, 1, 1980.