Effective Prediction of Job Processing Times in a Large-Scale Grid Environment

Menno Dobber
Vrije Universiteit,
De Boelelaan 1081a,
1081HV Amsterdam,
The Netherlands
E-mail: amdobber@few.vu.nl

Rob van der Mei
CWI and Vrije Universiteit,
c/o Kruislaan 413,
1098SJ Amsterdam,
The Netherlands
E-mail: mei@few.vu.nl

Ger Koole
Vrije Universiteit,
De Boelelaan 1081a,
1081HV Amsterdam,
The Netherlands
E-mail: koole@few.vu.nl

Abstract

Grid applications that use a considerable number of processors for their computations need effective predictions of the expected computation times on the different nodes. Currently, there are no effective prediction methods available that satisfactorily cope with those ever-changing dynamics of computation times in a grid environment. Motivated by this, in this paper we develop the Dynamic Exponential Smoothing (DES) method to predict job processing times in a grid environment. To compare predictions of DES to those of the existing prediction methods, we have performed extensive experiments in a real large-scale grid environment. The results illustrate a strong and consistent improvement of DES in comparison with the existing prediction methods.

1. Introduction

In [4] an extensive set of statistical properties of job processing times has been investigated. The results of the statistical data analysis demonstrate that: (1) the characteristics of the datasets very often differ completely, (2) the datasets show on average more long- as short-term fluctuations and the proportion differs per dataset, (3) the average job-time fluctuate continuously during the run, with differences in the amount of fluctuations between the different nodes, (4) the standard deviations fluctuate considerably during the run, with differences in the amount of fluctuations between the different nodes, and (5) the datasets contain a small amount of peaks that have a substantial influence on the standard deviation, the total amount of fluctuations, and the variance. This has raised the need for the development of effective prediction methods of job processing times that are able to deal with those characteristics. To this end, these observations form the basis for both an investigation of the applicability of existing predictors for predicting job processing times and the development of a new prediction method.

In this paper, we develop a new prediction method for the running times of successive jobs on nodes in a grid environment. To this end, in Section 2 we combine the results of [4] (outlined above), and a theoretical analysis, to identify the strong and weak points for an extensive set of predictors. In Section 3, these findings form the basis for the development of the prediction method DES. Subsequently, in Section 4 the accuracy of the predictions based on DES are compared to that of the other prediction methods. The results show that DES strongly outperforms the existing methods in the vast majority of the datasets.

2. Analysis of Existing Prediction Methods

In this section, we analyze a variety of existing methods that can be used to predict job times on shared processors, with the results from [4]. The most commonly used predictors in grid studies are Exponential Smoothing (ES) [3], the Network Wheather Service (NWS) [10] and Autoregressive (AR) [2] . ES defines y_t as the measured value at time t, $\hat{y}_t$ as the prediction for y_t, α as a chosen parameter between 0 and 1. Next, the prediction for y_t is defined as:

$$\hat{y}_t = \alpha y_{t-1} + (1-\alpha)\hat{y}_{t-1}, \text{ where } 0 \leq \alpha \leq 1. \quad (1)$$

For brevity, we do not discuss the details of the other methods. For predicting job processing times there are many other useful predictors applicable from other areas (e.g. the economics area). We selected the Adaptive Exponential Smoothing Predictors (AESP) and Smooth Transition Exponential Smoothing (STES) predictors as potential methods that can give accurate predictions in grids. From the AESP we analyze Trigg and Leach [8], Whybark [9], Mentzer [5], Pantazopoulos and Pappis [6], and from the STES predictors [7] we discuss STES $|e|$ with $\gamma < 0$, STES $|e|$, STES e^2, and STES Whybark.

The results of the prediction-methods analysis show that the main reasons for inaccuracy of these existing methods

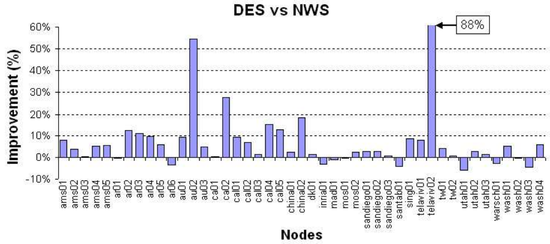

Figure 1. DES-predictor improvements compared to the NWS prediction method

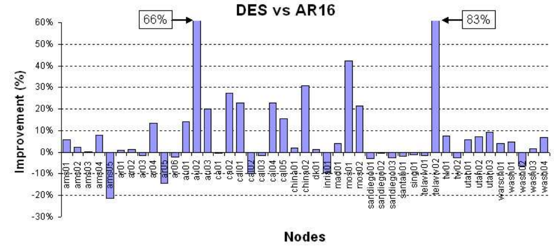

Figure 2. DES-predictor improvements compared to the AR16 predictor

are (1) delayed reaction to "level switches", and (2) overreaction to sudden peaks in the job processing times. This raises the need for the development of a new prediction method particularly suited for the specifics of a grid environment.

3. DES prediction method

The main idea behind DES is that it uses ES with a factor α that switches between multiple dynamically adapting values. Those values represent the α-parameters that would be optimal for the previous measurements that are of the same class as the most recent measurement. The DES-predictor classifies different types of fluctuations: 3 huge fluctuations classes ($H1$, $H2$ and $H3$), 1 medium fluctuations class (M), and 1 base class (B).

4. Experimental Results

In this section we compare the performance of the DES prediction method with those of the predictors described in Section 2. For the comparison we use 45 datasets of 18 different nodes of Planetlab [1], a global-scale grid testbed environment. For brevity, we only show the results of the comparison with the NWS and the AR methods.

Figure 1 illustrates the improvement percentage in the root of the mean squared error of the DES prediction method in comparison with the NWS prediction method. The figure shows that DES outperforms the NWS prediction method on average with 8%.

Figure 2 illustrates the performance improvement of the DES prediction method in comparison with the AR(16) predictor. On average, the DES prediction method is 9% more accurate as the AR(16).

5. Conclusions

Extensive theoretical and practical analysis, based on the results of [4] (outlined in the Introduction), of a variety of prediction methods have been performed. The results show that none of these methods is well-suited for dealing with the specifics of job processing times in a grid environment, including the presence of level switches and sudden peaks. Due to the absence of an effective prediction method for running times of jobs, we have developed a new prediction method, called DES, that overcomes the shortcomings of the existing methods by properly reacting to level switches and peaks. The power of DES lays in the fact that it is able to deal with on the one hand with peaks and on the other hand with level switches.

Extensive comparisons with a large number of datasets show that DES is a highly effective method for predicting running times of jobs on shared processors. DES consistently outperforms the existing prediction methods.

References

[1] http://www.planet-lab.org.

[2] P. A. Dinda and D. R. O'Hallaron. Host load prediction using linear models. *Cluster Computing*, 3(4):265–280, 2000.

[3] A. M. Dobber, G. M. Koole, and R. D. van der Mei. Dynamic load balancing experiments in a grid. In *Proceedings of the 5th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2005)*, pages 123–130. IEEE Press, May 2005.

[4] A. M. Dobber, R. D. van der Mei, and G. M. Koole. Statistical properties of task running times in a global-scale grid environment. 2005. Submitted for publication.

[5] J. T. Mentzer. Forecasting with adaptive extended exponential smoothing. *Journal of the Academy of Marketing Science*, 16:62–70, 1988.

[6] S. N. Pantazopoulos and C. P. Pappis. A new adaptive method for extrapolative forecasting algorithms. *European Journal of Operational Research*, 94:106–111, 1996.

[7] J. W. Taylor. Smooth transition exponential smoothing. *Journal of Forecasting*, 23:385–404, 2004.

[8] D. W. Trigg and A. G. Leach. Exponential smoothing with an adaptive response rate. *Operations Research Quarterly*, 18:53–59, 1967.

[9] D. C. Whybark. Comparison of adaptive forecasting techniques. *Logistics Transportation Review*, 8:13–26, 1973.

[10] R. Wolski. Experiences with predicting resource performance on-line in computational grid settings. *SIGMETRICS Performance Evaluation Review*, 30(4):41–49, 2003.