

The master–slave paradigm on heterogeneous systems: A dynamic programming approach for the optimal mapping ☆

Francisco Almeida, Daniel González, Luz Marina Moreno *

Dpto. Estadística, I.O. y Computación, La Laguna University, La Laguna, Tenerife 38271, Spain

Received 30 June 2004; accepted 20 October 2004
Available online 23 March 2005

Abstract

We study the master–slave paradigm over heterogeneous systems. According to an analytical model, we develop a dynamic programming algorithm that allows to solve the optimal mapping for such paradigm. Our proposal considers heterogeneity due both to computation and also to communication. The optimization strategy used allows to obtain the set of processors for an optimal computation. The computational results show that considering heterogeneity also on the communication increases the performance of the parallel algorithm.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Master–slave; Heterogeneous system; Optimal mapping; Dynamic programming

1. Introduction

Cluster Systems are playing an important role in High Performance Computing due to the low cost and availability of commodity PCs networks. A typical resource pool comprises heterogeneous commodity PCs and complex servers. The parallel standard libraries allow an easy portability of the

parallel programs to the cluster contributing also to the increasingly consideration of these architectural solutions. However, a disappointing handicap appears with the behavior of the real running time of the applications. Most of the programs have been developed under the assumption of an homogeneous target architecture and this assumption has been also followed by most of the analytical and performance models [1,2]. This models cannot be directly applied to the heterogeneous environments.

We revisit in this paper the mapping problem for master–slave algorithms under this new heterogeneous context. Some authors [3] broach the same problem but consider heterogeneity due only

☆ This work has been partially supported by the EC (FEDER) and the Spanish MCyT (Plan Nacional de I+D+I, TIC2002-04498-C05-05 and TIC2002-04400-C03-03).

* Corresponding author. Fax: +34 922 31 92 02.

E-mail addresses: falmeida@ull.es (F. Almeida), dgonmor@ull.es (D. González), lmmoreno@ull.es (L.M. Moreno).

to the difference in processing speeds of workstations. However, it is a stated fact that the heterogeneity in the speed of workstations can have a significant impact on the communication send/recv overhead [4], even when the communication capabilities of the processors are the same. Our proposals consider heterogeneity due both to computation and communication, what provides a very general overview of the system and a very wide range of applicability.

The master–slave paradigm has been extensively studied and modeled in homogeneous architectures [5–8]. However the modelization of this paradigm on heterogeneous platforms is still an open question. A very good work that studies the scheduling of parallel loops at compile time in heterogeneous systems was presented in [9]. They cover both homogeneous and heterogeneous systems and regular and irregular computations, giving optimal or near optimal solutions. An analytical model and a strategy for an optimal distribution of the tasks on a master–slave paradigm has been presented in [10]. However the accuracy of the predictions with their model is still to be proved. In [3] the authors provide effective models in the framework of heterogeneous computing resources for the master–slave paradigm. They formulate the problem and approach different variants but they consider heterogeneity only due to differences in the processing speed of the workstations. In [11] a very elegant structural model has been presented to predict the performance of master–slave algorithms on heterogeneous systems. The model is quite general and covers most of the situations appearing in practical problems. The structural model can be viewed as an skeleton where the models for master–slave applications with different implementations or environments must be instantiated. The main drawback of the approach is that the instantiation of the model for the particular cases is not a trivial task. It does not find optimal distribution of works among the processors. In [12] we presented an analytical model that predicts the execution time of master–slave algorithms on heterogeneous systems. The model includes heterogeneity due both to computation and to communication. Our proposal covers some of the variants studied in [3,11]. The compu-

tational results prove the efficiency of our scheme and the accuracy of the predictions.

The scheduling problem for master–slave computations has been extensively studied in many of the aforementioned works, however, most of these works consider the problem of the distribution of work over a set of fixed processors. We claim that to solve the optimal scheduling, the set of processors must be considered also as a parameter in the minimization problem. The predictive ability of the model that we propose in [12] constitutes the basis for further studies for the master–slave paradigm. The main contribution of this work is the development of an optimal mapping strategy for master–slave computations based in the analytical model introduced in [12]. In particular, this analytical model allows to formulate the problem as a resource optimization problem that has been studied in depth [13]. According to this formulation we derive strategies for the optimal assignment of tasks to processors. Our proposal improves the classical approach of assigning tasks of size proportional to the processing power of each processor and includes the elements to obtain the set of processors for an optimal computation.

The paper is organized as follows. Section 2 introduces the heterogeneous network model and some notation used along the paper. Section 3 specifies the problem and describes an inductive analytical model that predicts the running time of a master–slave program in an heterogeneous cluster of PCs for a given distribution of work. This analytical model has been validated in Section 4. The problem of finding the optimal distribution of work among the processors (the mapping problem) is studied in Section 5. Based on the analytical model presented in Section 3 we propose a general approach that formulates the mapping problem as a particular instance of the Resource Allocation Problem (RAP). We solve this optimization problem using a dynamic programming algorithm. The dynamic programming approach allows also to solve the problem of finding the optimal number of processors. The computational results presented in Section 6 prove that considering heterogeneity on the communication capabilities increases the quality of the predictions.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات