

ALTA: Asynchronous Loss Tolerant Algorithms for Grid Computing

Olivier Aumage*

Jacques M. Bahi†

Sylvain Contassot-Vivier†

Raphaël Couturier†

Alexandre Denis‡

Raymond Namyst*

Guillaume Papauré‡

Christian Perez‡

Marc Sauget†

Abstract

This paper describes an environment dedicated to the building of efficient scientific applications for the Grid on top of unreliable communication networks. Nowadays, scientific computing applications are usually built on top of reliable communication protocols (such as TCP). Nevertheless, the additional cost introduced by the reliability layer is not negligible in wide area network-based grid environments. On the other hand, data loss in communications may have a dramatic impact over the performance – if not over the correctness – of classical parallel algorithms.

However, a particular class of parallel iterative algorithms happens to be tolerant to such losses. This is the class of *asynchronous iterative algorithms*, which are commonly used in large scientific applications. They are particularly prone to a good communication/computation overlap since processors are no more synchronized. In this study, we aim at proposing a new architecture suitable for the development of asynchronous iterative algorithms tolerant to message losses.

1 Introduction

The ACI Grid/ARC ALTA project involves three fields of competence. *Communication protocols*: ALTA is based on the Variable Reliability Protocol (VRP), which is built on top of UDP. VRP allows the application to control the reliability level (and therefore the cost) of its communication by specifying the percentage of data that may be lost, at the message level. *Runtime systems*: the MARCEL (multithreading) and MADELEINE (communication interface) integrated software pair which builds ALTA's core provide ALTA with the required computation/communication decoupling. *Parallel algorithms*: ALTA's targets are the members of the parallel asynchronous iterative class. Hence, this project gathers the complementary research skills of three french teams.

The following section presents the ALTA project. Then, Section 3 deals with an application. Currently, all the experiments have not been conducted yet; they will be done in the final version of the article.

2 Presentation of ALTA

ALTA relies on three different parts, corresponding to three levels. The low level is made of VRP, a protocol based on UDP which allows the percentage of loss to be specified. Relaxing reliability constraints enables us to obtain very attractive bandwidth on wide-

area networks compared to using unconditionally reliable protocols such as TCP, for instance. VRP is not really relevant on local area networks where the average loss rate at the OSI's *network* layer is usually very low and the cost of *transport* layer's reliability is not prohibitive. Its use is therefore optional.

At the medium level, ALTA relies on the MADELEINE communication interface and the MARCEL user-level multithreading library. MADELEINE is a multi-protocol communication library specifically designed to be used in multi-threaded environments. It has been extended to support the VRP protocol and to provide access to VRPs loss rate control and loss status report. MARCEL is a POSIX thread library which is designed to be efficient, portable and to provide MADELEINE with unique mechanisms for highly reactive asynchronous network event detection.

At the top level of this environment, we intend to develop asynchronous iterative algorithms in order to conduct new experiments. Indeed, those algorithms have been proved to be tolerant to message loss, but nobody has yet implemented them in such a context. Although we have already shown their interest in grid computing environments *without* message loss, we strongly think that they will be more efficient with such a dedicated programming environment. However, while the whole approach is viable from a theoretical point of view, effectively getting the most out of asynchronous iterative algorithm in the given context involves many issues to be addressed, ranging from high-level algorithmic issues such as distributed convergence detection and halting problems to implementation issues such as maximizing the overlapping of computation and communication, and even some seemingly external issues such as session deployment.

3 Application

We have developed an application which aims at solving non-linear chemical problems. This application computes the evolution of two chemical species in a two-dimensional domain. The problem involves the advection-diffusion equations and represents a classical PDE problem. To solve such an equation, the problem is first discretized using finite-difference scheme, then the implicit Euler method is used and the non-linear system is iteratively solved using the multi-splitting Newton method.

4 Conclusion

In this paper we describe our ALTA architecture allowing to efficiently implement asynchronous-iteration-based applications that can benefit from relaxed communication reliability. This original work is based on the fact that although data loss is inherently unavoidable at the network layer, enforcing strict reliability on top of it has a cost that can actually be saved for this class of applications.

*Laboratoire Bordelais de Recherche en Informatique (LABRI)

†Laboratoire d'Informatique de Franche-Comté (LIFC)

‡IRISA- INRIA Rennes