

"Modelos de predicción de performance en sistemas paralelos"

Laura De Giusti ¹, Marcelo Naiouf ², Armando De Giusti³

*LIDI - Laboratorio de Investigación y Desarrollo en Informática⁴.
Facultad de Informática. Universidad Nacional de La Plata.*

Resumen

Esta línea de investigación (dentro del proyecto de Procesamiento Concurrente, Paralelo y Distribuido del LIDI) se coordina con la Universidad Autónoma de Barcelona estudiando los modelos de predicción de performance en sistemas paralelos y analizando algunas modificaciones a los mismos según el tipo de arquitectura y paradigma de paralelismo utilizado.

En el estudio se analiza la respuesta de los modelos para "clases" de arquitecturas reconocidas (Memoria compartida, Memoria Distribuida, Memoria compartida Distribuida) con diferentes esquemas de comunicación: punto a punto, bus, múltiple bus. Asimismo se considera en cada caso la adaptación de diferentes paradigmas de programación paralela al modelo de arquitectura utilizado.

El aporte de la línea de investigación tiende a la especificación de un modelo unificador para la predicción de performance, en particular para procesos asincrónicos bajo el paradigma cliente-servidor y sobre una arquitectura tipo cluster de workstations.

Palabras Claves

Paralelismo - Concurrencia - Modelos de Computación Paralela – Sincronización – Mensajes asincrónicos – Predicción de performance

¹ Licenciada en Informática. Becario UNLP. Jefe de Trab. Práct. SD Idgiusti@lidi.info.unlp.edu.ar

² Profesor Adjunto Dedicación Exclusiva. mnaiouf@lidi.info.unlp.edu.ar

³ Investigador Principal CONICET. Profesor Titular Ded. Exclusiva. degiusti@lidi.info.unlp.edu.ar

⁴ LIDI - Facultad de Informática. UNLP - Calle 50 y 115 1er Piso, (1900) La Plata, Argentina.

TE/Fax +(54)(221)422-7707. <http://lidi.info.unlp.edu.ar>

Introducción

El *paralelismo*, asociado con *multiprocesamiento* en hardware y *conurrencia* en software requiere especificar e implementar procesos explotando la concurrencia implícita o explícita en el problema a resolver, y *al mismo tiempo* optimizar la adaptación del algoritmo desarrollado a la arquitectura física de soporte.

Los paradigmas de expresión de algoritmos paralelos se asocian entonces en forma directa con la arquitectura de procesamiento que soporta la ejecución, dando lugar a la noción de *sistema de procesamiento paralelo*, que incluye el software propio de la aplicación y el modelo de hardware y comunicaciones elegido.

Cuatro paradigmas básicos relacionados con los sistemas de procesamiento paralelo son los modelos *master-worker*, *pipeline*, *single processor multiple data* y *divide and conquer*. Cada uno de estos paradigmas (con sus variantes particulares) se adapta mejor a *clases* de problemas en programación paralela.

Avanzando en el concepto anterior, existen *patrones* de solución a problemas de procesamiento paralelo utilizando cada uno de estos cuatro paradigmas básicos, y dentro de cada uno de ellos hay un *modelo de prestación* o *performance* que nos puede indicar los límites alcanzables en cuanto a *speedup*, eficiencia y rendimiento de la solución paralela, en función de los parámetros propios de la aplicación y la arquitectura física de soporte.

Un área importante de investigación es la **predicción de performance** alcanzable por un *sistema paralelo* (normalmente asociado a uno de los patrones mencionado anteriormente). La especificación, transformación, optimización y evaluación de algoritmos distribuidos y paralelos, según las predicciones realizadas marca una posibilidad de optimización de soluciones en función de distintos modelos de arquitectura y el estudio de métricas de complejidad y eficiencia.

En el cómputo secuencial, el modelo RAM ha sido prácticamente aceptado como un estándar. En el procesamiento distribuido/paralelo no hay un modelo teórico unificador. Precisamente el objetivo de esta línea de investigación es profundizar el estudio de los modelos actualmente utilizados (PRAM, LOGP, BSP, CCM, BSPRAM, OBBSP) especialmente considerando arquitecturas MIMD débilmente acopladas, con procesadores homogéneos. En particular interesan los procesos asincrónicos bajo el paradigma cliente-servidor, sobre una arquitectura tipo cluster de workstations.

La *performance* obtenida en un sistema paralelo está dada por una compleja relación en la que intervienen una gran cantidad de factores (el tamaño del problema, la arquitectura de soporte, la distribución de procesos en procesadores, la existencia o no de un algoritmo de balanceo de carga, etc). Por otra parte, existen varias métricas para evaluar sistemas paralelos, tales como el tiempo de ejecución paralelo; el *speedup*; la eficiencia; la escalabilidad o la isoeficiencia, métricas que pueden tener mayor o menor importancia según la aplicación particular del sistema paralelo.

En este marco, interesa *predecir* y *evaluar performance* de sistemas paralelos, utilizando modelos de propósito general y medir la calidad de las estimaciones sobre diferentes arquitecturas reales.

Temas de Investigación

- Sistemas Paralelos. Paradigmas de Programación paralela. Estudio de patrones para los paradigmas master-worker, pipeline, single processor multiple data y divide and conquer.
- Modelos de predicción de performance clásicos: PRAM, LOGP, BSP, CCM, OBSP, BSPRAM
- Análisis del ajuste de los modelos a “clases” de sistemas paralelos. Desarrollos experimentales.
- Métricas de performance en sistemas paralelos.
- Dependencia de los modelos respecto del balance de carga en los procesadores y la homogeneidad / heterogeneidad de los mismos.
- Análisis de la dependencia de los resultados obtenidos en función de cambios en la arquitectura de soporte. En particular pasar de la red homogénea con memoria distribuida y comunicación vía bus a un esquema multiprocesador con memoria compartida distribuida.
- Análisis de la escalabilidad de las soluciones para los diferentes patrones.
- Análisis de las extensiones necesarias para tener un modelo de predicción de performance para procesos asincrónicos bajo el paradigma cliente-servidor.
- Aplicaciones paralelas de tratamiento de imágenes: Transformadas aplicables en la codificación de imágenes con pérdida, Clustering, Reconocimiento de patrones, Análisis de similitud de imágenes.

Equipamiento de experimentación

En el LIDI se dispone de un cluster de PCs con 16 equipos homogéneos. Por otra parte se puede utilizar la computadora Clementina que tiene 40 procesadores con memoria compartida distribuida.

Naturalmente también se puede experimentar con arquitecturas seudo-paralelas tal como redes heterogéneas con un soporte de comunicaciones tipo PVM o MPI.

BIBLIOGRAFIA BASICA

- [And91]** Gregory R. Andrews. "Concurrent Programing" The Benjamin/Cummings Publishing Company. 1991.
- [And99]** Gregory R. Andrews. "Foundations on Multithread and Distributed Programing" Addison Wesley. 1999.
- [Bax94]** Gregory A. Baxes "Digital Image Processing. Principles and Applications" John Wiley & Sons Inc. 1994.
- [Bri95]** Brinch Hansen, P., "Studies in computational science: Parallel Programming Paradigms", Prentice-Hall, Inc., 1995.
- [Bub97]** Bubak, Funika, Moscinski, "Performance Analysis of Parallel Applications under Message Passing Environments", www.icsr.agh.edu.pl/publications/html/perf_full/, 1997.
- [Bust88]** Bustard, Elder, Welsh, "Concurrent Program Structures", Prentice Hall, 1988.
- [Cof92] M. Coffin, "Parallel programming- A new approach", Prentice Hall, Englewood Cliffs, 1992.
- [Cha88]** Chandhi K. M., Misra J., "Parallel Program Design. A Foundation", Addison Wesley, 1988.
- [Esp 99]** Espinosa A., Parcerisa F., Margalef T., Luque E. "Relating the Execution Behaviour with the Structure of the Application" (Euro PVM/MPI) Lecture Notes in Computer Science, vol. 1697, 91-98. Springer, Alemania, 1999.
- [Esp 00a]** Espinosa A., Margalef T., Luque E. "Automatic Performance Analysis of Master/Worker PVM Applications with Kpi" (Euro PVM/MPI) Lectures Notes in Computer Science, vol. 1908, 47-55. Springer . Alemania, 2000.
- [Esp 00b]** Espinosa A., Margalef T., Luque E. "Integrating Automatic Techniques in a Performance Analysis Session" (Euro-Par 2000) Lectures Notes in Computer Science, vol. 1900, 173-177. Springer, Alemania, 2000.
- [Gon92]** Rafael C. González, Richard E. Woods, "Digital Image Processing", Addison-Wesley Publishing Comp., 1992.
- [Heer91]** D. W. Heermann, A. N. Burkitt, "Parallel Algorithms in Computational Science", Springer-Verlag, 1991.
- [Hoar85]** C. A. R. Hoare, "Communicating Sequential Processes", Pentice-Hall, 1985.
- [Hus91]** Zahid Hussain "Digital Image Processing" Ellis Horwood Limited. 1991.
- [Hwa93]** Hwang K., "Advanced Computer Architecture: Paralelism, Scalability, Programability", McGraw-Hill, 1993.
- [IEEE]** Colección de Transactions on Parallel and Distributed Processing.
- [Kum94]** Kumar V., Grama A., Gupta A., Karypis G., "Introduction to Parallel Computing. Desing and Analysis of Algorithms", Benjamin/Cummings, 1994.
- [Laws92]** H. Lawson, "Parallel processing in industrial real time applications", Prentice Hall 1992.
- [Lei92] Leighton F. T., "Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes", Morgan Kaufmann Publishers, 1992.
- [Luq95]** Luque E, Ripoll A, Cortès A, Margalef T, "A Distributed Diffusion Method for Dynamic Load Balancing on Parallel Computers", Proceedings of the EUROMICRO Workshop on Parallel and Distributed Processing, IEEE Computer Society, Jan. 1995.
- [Mor94]** Morse F., "Practical Parallel Computing", AP Professional, 1994.
- [Pri01]** Printista Marcela, "Modelos de Predicción en Computación Paralela", Tesis de Magister en Ciencias de la Computación, 2001.
- [Sim97]** Sima D, Fountain T, Kacsuk P, "Advanced Computer Architectures. A Design Space Approach", Addison Wesley Longman Limited, 1997.
- [Tin98]** Tinetti F., De Giusti A., "Procesamiento Paralelo. Conceptos de Arquitectura y Algoritmos", Editorial Exacta, 1998.
- [Zom96]** Zomaya A. (ed), "Parallel Computing. Paradigms and Applications", International Thomson Computer Press, 1996.