

Procesamiento Paralelo en Clusters

Fernando G. Tinetti, Andrés Barbieri, Mónica Denham, Franco Chichizola, Laura De Giusti, Marcelo Naiouf, Armando De Giusti

{fernando, barbieri, mdenham, francoch, ldgiusti, mnaiouf, degiusti}@lidi.info.unlp.edu.ar

III-LIDI: Instituto de Investigación en Informática - LIDI Facultad de Informática, UNLP 50 y 115, 1900, La Plata, Argentina

Resumen

En el contexto de procesamiento paralelo en clusters, existen muchas líneas de investigación que aún están abiertas a nivel internacional. Las dos más importantes que se tienen en cuenta en esta línea de investigación son: a) rendimiento paralelo y b) creación de ambientes de desarrollo y ejecución de aplicaciones en clusters. En el contexto de rendimiento paralelo se tienen en cuenta aspectos tales como la optimización, estimación y modelización; todos temas que aún no han sido resueltos en general y en el mejor de los casos se han presentado soluciones para algunas áreas de aplicación. En el contexto de ambientes de desarrollo y ejecución de aplicaciones en clusters se tiene en cuenta la posibilidad de proveer a los usuarios una "imagen única del sistema", o Single System Image (SSI) al menos en algunos aspectos o subsistemas específicos que son de interés general.

Introducción

En el contexto del procesamiento paralelo y distribuido se ha establecido desde hace un tiempo, con mucho interés a nivel internacional, el procesamiento paralelo en clusters [3]. Si bien en principio se ha aplicado y se aplica en lo que se denomina usualmente "cómputo científico", actualmente también se está utilizando satisfactoriamente en otras áreas. Una de las razones más importantes de esta evolución ha sido la relación sumamente ventajosa entre el costo de los clusters con respecto al rendimiento obtenido, dado que las computadoras de escritorio (principalmente PCs), han aumentado significativamente el rendimiento a un costo básicamente constante [4].

A medida que se tienen mayor cantidad de usuarios y aplicaciones utilizando clusters, y también por la propia relación de costo rendimiento cada vez más ventajosa, se hace más fuerte la necesidad de adaptación de algoritmos y ambientes de desarrollo y ejecución de aplicaciones paralelas. Aunque inicialmente la utilización de clusters estuvo orientada a la optimización de recursos disponibles en las redes locales de computadoras para aumentar la productividad (throughput) [22] [7], también se están utilizando para los problemas clásicos de procesamiento numérico y también se ha extendido a otras áreas de procesamiento distribuido [4].

Desde el punto de vista de la optimización de los recursos se hace necesaria, por ejemplo, una visión única del sistema a nivel de interfaz de ejecución, monitorización y eventualmente sintonización de planificación de aplicaciones a resolver utilizando un cluster. En cierta forma, ambientes como CONDOR [24] [25] [21] han resuelto muchos de los problemas en este sentido, pero aún se tienen muchos aspectos no resueltos o por lo menos sin consenso general en cuanto a cómo resolverlos [6] [10].

Dado el interés por la utilización de clusters para procesamiento paralelo, se hace muy importante el análisis y la optimización de rendimiento paralelo de los algoritmos a utilizar. Por un lado, se tienen un conjunto de algoritmos paralelos muy grande desarrollados (principalmente en el área de

procesamiento numérico) [16] [18] [15] [32], que no necesariamente son óptimos para ser utilizados en el contexto de los clusters. Por otro lado, existen muchas aplicaciones que se necesitan resolver pero para las cuales aún no se han desarrollado los algoritmos paralelos que puedan ser utilizados en el contexto de los clusters.

Quizás en un punto medio entre la optimización de rendimiento paralelo y la optimización de parámetros como el throughput, se tienen aún muchos detalles por resolver en el contexto del aprovechamiento para cómputo paralelo de las redes locales de computadoras instaladas. Algunos de los problemas que se deben resolver involucran detalles tan específicos (y complejos de resolver) como el balance de carga en los clusters heterogéneos y los problemas de disponibilidad de los clusters no dedicados (que son mayoría en el caso de las redes locales instaladas).

Líneas de Investigación y Desarrollo

Dada la amplia gama de problemas a resolver, se podrían dividir por niveles de abstracción o áreas de aplicación. Entre los temas que se están resolviendo se pueden mencionar en niveles crecientes de abstracción:

- Resolución de problemas numéricos, específicamente de álgebra lineal, en clusters. En este caso se tienen numerosas propuestas y alternativas con las cuales hacer comparaciones directas, tales como [1] [2] [9] [11] [12]. En el contexto más específico aún de problemas simples y de mayor cantidad de usuarios potenciales se tienen [19] [13] [14].
- Optimización de rendimiento paralelo en clusters homogéneos. Involucra temas como: balance de carga, optimización/estimación de granularidad, optimización de middleware de ejecución, optimización de uso de recursos específicamente involucrados en los clusters, etc.
- Optimización de rendimiento paralelo en clusters heterogéneos. En este caso específico el problema de balance de carga es especialmente complejo, teniendo en cuenta que se combina con el problema de la granularidad de las aplicaciones.
- Transformación y optimización de algoritmos numéricos y no numéricos para su ejecución en clusters homogéneos/heterogéneos.
- Ambientes de SSI para ejecución en clusters con computadoras sin disponibilidad completa. Específicamente interesa resolver el problema de "tolerancia a fallas" en caso que alguna computadora del cluster comience a ser utilizada para sus tareas estándares fuera del procesamiento paralelo.
- Siempre es importante la estimación de rendimiento paralelo y en el caso específico de los clusters también debe ser resuelto [33]. La solución que actualmente se está investigando es la adaptación de un sistema completo de checkpoint and restart para aplicaciones paralelas en clusters [20].

Resultados Obtenidos/Esperados

Entre los resultados obtenidos se pueden destacar varios, específicamente relacionados con la optimización de rendimiento paralelo, tales como:

- Desarrollo de un conjunto de principios de paralelización específicamente orientados a la obtención de rendimiento, optimizado en redes locales interconectadas por el hardware estándar Ethernet [17] [29].
- Comprobación experimental satisfactoria de los principios de paralelización mencionados anteriormente, en varios problemas específicos de álgebra lineal, tanto básicos [31] como computacionalmente más complejos [30]
- Optimización de recursos de comunicaciones, específicamente de los recursos disponibles definidos por el estándar Ethernet [26].

Otros de los resultados obtenidos, en un mayor nivel de abstracción, se relacionan con:

- Herramientas y métodos de evaluación de las comunicaciones en clusters [27] [28].
- Herramienta de instalación automática y replicación de clusters como un único sistema [5].

Además se han desarrollado numerosos algoritmos implementados y evaluados sobre clusters con resultados satisfactorios.

En el contexto más general, se intenta desarrollar una metodología de evaluación y estimación de rendimiento paralelo, que se pueda aplicar al contexto específico de los clusters, al menos por áreas de aplicación. Por otro lado, se intenta avanzar con un ambiente de ejecución y monitorización de aplicaciones paralelas en clusters que incluya al menos la posibilidad de checkpoint and restart en el ámbito de los clusters no dedicados.

También se continúa con la paralelización de problemas específicos, tanto del área de procesamiento numérico como de otras áreas de interés. Esto incluye tareas de cooperación con universidades del exterior, intentando incluir clusters de cómputo en España y Brasil.

Formación de Recursos Humanos

Desde hace varios años se dirigen alumnos en proyectos finales de la Licenciatura en Informática en los temas afines de procesamiento paralelo y distribuido. Esta tarea se continúa, y también se está avanzando en los postgrados de maestría y doctorado. De hecho en el transcurso de este año ya se ha finalizado un doctorado en la Universidad Autónoma de Barcelona y se han dictado cursos de postgrado en la Universidad Nacional de La Plata que otorga créditos válidos para el doctorado.

Por otro lado, siempre se tiene disponible la posibilidad de incorporación de alumnos y graduados a las tareas de investigación relacionadas con procesamiento paralelo en clusters. En este sentido, no se establece como requisito la inscripción en un postgrado o la finalización de una carrera de grado sino la motivación misma de las propias tareas de investigación.

También los proyectos de cooperación con universidades de España y Brasil implican la participación de alumnos avanzados de la Licenciatura en Informática de nuestra Facultad así como de graduados con interés en investigación. Varios de los alumnos de postgrado (doctorado y maestría) están o estarán relacionados directa o indirectamente con estos proyectos de cooperación. Por un lado, en la evaluación de desarrollos del exterior y por otro en la propuesta y desarrollos para otras Universidades.

Referencias

- [1] Anderson E., Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, D. Sorensen, "LAPACK: A Portable Linear Algebra Library for High-Performance Computers", Proceedings of Supercomputing '90, pages 1-10, IEEE Press, 1990.
- [2] Anderson E., Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, D. Sorensen, LAPACK Users' Guide (Second Edition), SIAM Philadelphia, 1995.
- [3] Anderson T., D. Culler, D. Patterson, and the NOW Team, "A Case for networks of Workstations: NOW", IEEE Micro, Feb. 1995.
- [4] Baker M., R. Buyya, "Cluster Computing at a Glance", in R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 3-47, 1999.
- [5] Barbieri A., "CLICLUX: Cloner for LINUX CLUsters. Una herramienta para instalación de clusters usando Linux", 4to. Workshop de Procesamiento Distribuido y Paralelo, CACIC 2003, La Plata, 6 al 10 de Octubre de 2003.
- [6] Barreiro Paz M., V. M. Gulias, "Cluster Setup and its Administration", in R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 48-67, 1999.
- [7] Basney J., M. Livny, "Deploying a High Throughput Computing Cluster", in R. Buyya Ed., High Performance Cluster Computing: Architectures and Systems, Vol. 1, Prentice-Hall, Upper Saddle River, NJ, USA, pp. 116-134, 1999.
- [8] Bischof C. H., C. Van Loan, "The WY Representation for Products of Householder Matrices", SIAM Journal of Scientific and Statistical Computing, 8: s2-s13, 1987.
- [9] Blackford L., J. Choi, A. Cleary, E. Dázevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, R. Whaley, ScaLAPACK Users' Guide, SIAM, Philadelphia, 1997.
- [10] Buyya R., T. Cortes, H. Jin, Single System Image (SSI), en Cluster Computing White Paper, Version 2.0, December 2000, Editor: Mark Baker.
- [11] Choi J., J. Demmel, I. Dhillon, J. Dongarra, S. Ostrouchov, A. Petitet, K. Stanley, D. Walker, R. Whaley, "LAPACK Working Note 95. ScaLAPACK: A Portable Linear Algebra Library for Distributed Memory Computers - Design Issues and Performance", 1996. Available at <http://www.netlib.org/lapack/lawns/index.html>
- [12] Choi J., J. Dongarra, R. Pozo, D. Walker, "ScaLAPACK: A Scalable Linear Algebra Library for Distributed Memory Concurrent Computers", Proc. 4th Symposium on the Frontiers of Massively Parallel Computation, IEEE Computer Society Press, pp. 120-127, 1992.
- [13] Dongarra J., J. Du Croz, S. Hammarling, R. Hanson, "An extended Set of Fortran Basic Linear Subroutines", ACM Trans. Math. Soft., 14 (1), pp. 1-17, 1988.
- [14] Dongarra J., D. Walker, "Libraries for Linear Algebra", in Sabot G. W. (Ed.), High Performance Computing: Problem Solving with Parallel and Vector Architectures, Addison-Wesley Publishing Company, Inc., pp. 93-134, 1995.
- [15] Foster I., Designing and Building Parallel Programs, Addison-Wesley, Inc., 1995. Versión html disponible en <http://www-unix.mcs.anl.gov/dbpp>.
- [16] Golub G., C. Van Loan, Matrix Computations, 2nd Edition, The John Hopkins University Press, 1989.
- [17] Institute of Electrical and Electronics Engineers, Local Area Network - CSMA/CD Access Method and Physical Layer Specifications ANSI/IEEE 802.3 - IEEE Computer Society, 1985.

- [18] Kumar V., A. Grama, A. Gupta, G. Karypis, Introduction to Parallel Computing. Design and Analysis of Algorithms, The Benjamin/Cummings Publishing Company, Inc., 1994.
- [19] Lawson C., R. Hanson, D. Kincaid, F. Krogh, "Basic Linear Algebra Subprograms for Fortran Usage", ACM Transactions on Mathematical Software 5, pp. 308-323, 1979.
- [20] Litzkow M., T. Tannenbaum, J. Basney, M. Livny, Checkpoint and Migration of UNIX Processes in the Condor Distributed Processing System, 1997.
- [21] Livny M., J. Basney, R. Raman, T. Tannenbaum, "Mechanisms for High Throughput Computing", SPEEDUP Journal, Vol. 11, No. 1, June 1997.
- [22] Mutka M., M. Livny, "The Available Capacity of a Privately Owned Workstation Environment", Performance Evaluation 12, 269-284, 1991.
- [23] Schreiber R., C. Van Loan, "A storage efficient WY Representation for products of Householder matrices", SIAM Journal of Scientific and Statistical Computing, 10: 53-57, 1989.
- [24] Tannenbaum T., D. Wright, K. Miller, M. Livny, "Condor - A Distributed Job Scheduler", in Thomas Sterling, editor, Beowulf Cluster Computing with Linux, The MIT Press, 2002. ISBN: 0-262-69274-0.
- [25] Thain D., T. Tannenbaum, M. Livny, "Distributed Computing in Practice: The Condor Experience", Concurrency and Computation: Practice and Experience, to appear in 2004.
- [26] Tinetti F., Barbieri A., "An Efficient Implementation for Broadcasting Data in Parallel Applications over Ethernet Clusters", Proceedings of the 17th International Conference on Advanced Information Networking and Applications (AINA 2003), IEEE Press, ISBN 0-7695-1906-7, March 2003.
- [27] Tinetti F., Barbieri A., "Cómputo y Comunicación: Definición y Rendimiento en Redes de Estaciones de Trabajo", Workshop de Investigadores en Ciencias de la Computación (WICC 2001), San Luis, Argentina, 22-24 de Mayo de 2001, pp. 45-48.
- [28] Tinetti F., Barbieri A., "Cómputo Paralelo en Clusters: Herramienta de Evaluación de Rendimiento de las Comunicaciones", Proceedings VIII Congreso Argentino de Ciencias de la Computación (CACIC), Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, 15 al 18 de Octubre de 2002, p. 123.
- [29] Tinetti F., Denham M., "Algebra Lineal en Clusters Basados en Redes Ethernet", Workshop de Investigadores en Ciencias de la Computación (WICC 2003), Tandil, Argentina, 22-23 de Mayo de 2003, pp. 575-579.
- [30] Tinetti F., Denham M., De Giusti A., "Parallel Matrix Multiplication and LU Factorization on Ethernet-based Clusters", High Performance Computing, 5th International Symposium, ISHPC 2003, Tokyo-Odaiba, Japan, October 20-22, 2003, Proceedings. Series: Lecture Notes in Computer Science, Vol. 2858. Veidenbaum, A.; Joe, K.; Amano, H.; Aiso, H. (Eds.), 2003, XV, 566 p. ISBN: 3-540-20359-1.
- [31] Tinetti F. G., E.Luque, "Parallel Matrix Multiplication on Heterogeneous Networks of Workstations", Proceedings VIII Congreso Argentino de Ciencias de la Computación (CACIC), Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina, p. 122, Oct. 2002. Available at <http://lidi.info.unlp.edu.ar/~fernando/publis/pmm.pdf>
- [32] Wilkinson B., Allen M., Parallel Programming: Techniques and Applications Using Networking Workstations, Prentice-Hall, Inc., 1999.
- [33] Zhang X., Y. Yan, "Modeling and characterizing parallel computing performance on heterogeneous NOW", Proceedings of the Seventh IEEE Symposium on Parallel and Distributed Processing, (SPDP'95), IEEE Computer Society Press, San Antonio, Texas, October 1995, pp. 25-34.