A Parallel Application in Grid Computing for the Longest Common Subsequence

C. E. R. Alves
Univ. São Judas Tadeu
E. N. Cáceres
Univ. Fed. Mato Grosso do Sul
S. W. Song
Universidade de São Paulo

Abstract

Given two strings, obtention of the longest subsequence common (LCS) to both strings is an important problem with applications in DNA sequence comparison, data compression, pattern matching, etc. Consider two strings X and Y of lengths m and n, respectively. The LCS problem can be solved sequentially in $O(mn)$ time. In a shared memory PRAM, it can be solved in $O(\log n)$ time with $mn/\log n$ processors, when $\log^2 m \log\log m \leq \log n$. In this paper we present a parallel algorithm in the Coarse-Grained Multicomputer model for computing the longest common subsequence between two strings. The proposed algorithm uses wavefront or systolic communication such that each processor needs to send data only to two other processors. This makes it suitable as a potential application for grid computing. The longest common subsequence of two strings can be computed on a parallel computer of p processors, each with local $O(nm/p)$ memory, in $O(p)$ communication rounds and $O(nm/p)$ local computing time. The algorithm is based on a compromise between the workload of each processor and the number of communication rounds required, expressed by a parameter called α. The algorithm is expressed in terms of this parameter that can be tuned to obtain the best overall parallel time in a given implementation. We show very promising experimental results obtained on a 64-node Beowulf machine.

Theme: Parallel Computing

References

[1] Alves, C. E. R., Cáceres, E. N., Dehne, F., and Song, S. W. A Parallel Wavefront Algorithm for Efficient Biological Sequence Comparison. The 2003 International Conference on Computational Science and its Applications. *Lecture Notes in Computer Science*, Springer-Verlag. Montreal, Canada, May 18-21, 2003, Accepted.

[2] Alves, C. E. R., Cáceres, E. N., Dehne, F. and Song, S. W. A CGM/BSP Parallel Similarity Algorithm. *Proceedings I Brazilian Workshop on Bioinformatics*. Gramado, RS, October 18, 2002, pp. 1-8.

[3] Apostolico, A. and Guerra, C. The Longest Common Subsequence Problem Revisited, Algorithmica, Vol. 2, 1987, pp. 315–336.

[4] Babu, K. Nandau and Saxena, Sanjeev. Parallel Algorithms for the Longest Common Subsequence Problem. Proceedings of the International Conference on High Performance Computing, 1997.

[5] Lu, Mi, and Lin, Hua. Parallel Algorithms for the Longest Common Subsequence Problem. *IEEE Transactions on Parallel and Distributed Systems*, Vol. 5, No. 8, 1994, pp. 835–848.

[6] Rick, Claus. New algorithms for the Longest Common Subsequence Problem. Technical Report 85123-CS, Institut fr Informatik, Universitt Bonn, October, 1994.