Tech. report, Lund University, 1994.

Blasting Past Fusion Trees

Arne Andersson* Stefan Nilsson* Torben Hagerup†

10 October 1994

Abstract

We present an $O(n \log\log n)$ worst-case time algorithm for sorting arbitrary integers, a significant improvement over the bound achieved by the fusion tree of Fredman and Willard.

Model of computation

We will consider a unit cost random access machine, RAM, with word length w and a memory composed of 2^w words. The instruction set includes addition, subtraction, comparison, unrestricted shift and the bitwise boolean operations AND and OR.

The algorithm

We study the problem of sorting n integers in the range $1..2^b$ on a RAM. We say that $T(n,b) \le f(n)$ if there exists an algorithm with worst-case time complexity $f(n)$ that solves this sorting problem. The assertion that $T(n,b_2) \le f(n) \Rightarrow T(n,b_1) \le f(n)$ will be abbreviated as $T(n,b_1) \le T(n,b_2)$. Similarly, $T(n,b) \le O(g(n))$ will be used to denote that there exists some f, $f(n) = O(g(n))$, such that $T(n,b) \le f(n)$.

A sequential version of an integer sorting algorithm by Albers and Hagerup [1] achieves

$$T(n, w/\log n) \le O(n \log\log n),$$

provided that $w > \log n \log\log n$. This can be seen by applying Theorem 1 [1] with parameters $k = \log n$, $m = 2^{w/\log n}$.

Also, using traditional bucket sort, we have

$$T(n, \log n) \le O(n).$$

The range reduction technique by Kirkpatrick and Reisch [6] achieves

$$T(n,b) \le T(n,b/2) + O(n).$$

*Department of Computer Science, Lund University, Box 118, S-221 00 Lund, Sweden
†Max-Planck-Institut für Informatik, D-66123 Saarbrücken, Germany

Let $c = \max\{\log n, w/\log n\}$ and consider the following algorithm for sorting n integers in the range $1..2^w$.

1. Use the range reduction technique until the length b of an integer is shorter than c.
2. If $b \leq \log n$ finish off the sorting using bucket sort, otherwise use the integer sorting algorithm by Albers and Hagerup.

This algorithm has the worst-case time complexity

$$\begin{aligned} T(n, w) &\leq O(n \log \frac{w}{c} + n \log\log n) \\ &= O(n \log\log n), \end{aligned}$$

where the first term comes from the range reduction step and the second from the sorting step.

The algorithm uses $\Theta(n + 2^{w/2})$ space; this can be reduced to $O(n)$ using universal hashing [4], yielding a randomized algorithm with the same expected time complexity.

Comments

The new bound is a significant improvement over the $O(n\sqrt{\log n})$ bound achieved by the fusion tree of Fredman and Willard [5].

The algorithm was first discovered as a recursive application of Forward Radix Sort [3]. Using this application, it is possible to achieve $T(n, b) \leq T(n, b/k) + O(kn)$ for any $k > 1$. This technique gives a worst-case cost of $O(\frac{k}{\log k} n \log\log n)$, while the space required is $\Theta(n + 2^{w/k})$.

The result can be extended in many ways. First, using the described algorithm as a subroutine of Forward Radix Sort [3], the results can be extended to efficiently handle strings of arbitrary length, i.e. longer than one machine word. Second, instead of using the range reduction technique by Kirkpatrick and Reisch we may use a modified version of Forward Radix Sort to reduce the space complexity, as sketched above. Third, the algorithm is well suited as a basis for new, efficient, parallel sorting algorithms.

These aspects will be treated in an extended article [2]. Just to mention one result, we claim that n integers can be sorted on an ARBITRARY CRCW PRAM using $O(\log n \log\log n)$ time and $O(n \log\log n)$ operations.

References

[1] S. Albers and T. Hagerup. Improved parallel integer sorting without concurrent writing. In *Proc. 3rd ACM-SIAM SODA*, pages 463–472, 1992.

[2] A. Andersson, T. Hagerup, and S. Nilsson. Manuscript. In preparation.

[3] A. Andersson and S. Nilsson. A new efficient radix sort. In *Proc. 35th Annual IEEE Symposium FOCS*, 1994. To appear.

[4] J. L. Carter and M. N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18:143–154, 1979.

[5] M. L. Fredman and D. E. Willard. Blasting through the information theoretic barrier with fusion trees. In *Proc. 22nd ACM STOC*, pages 1–7, 1990.

[6] D. Kirkpatrick and S. Reisch. Upper bounds for sorting integers on random access machines. *Theoretical Computer Science*, 28:263–276, 1984.