Martin Klampfer, 1526110 & Calvin Claus,

TU Wien | LVA: 184.710 Parallel Computing Einführung paralleles Rechnen

Fast parallel Quicksort Project

Hand-in

Index

[1. Problem statement: 2](#_Toc503348728)

[1.1. Problem definition: 2](#_Toc503348729)

[1.2. Task: 2](#_Toc503348730)

[1.3. General Notes: 2](#_Toc503348731)

[1.4. Referential, sequential implementation: 3](#_Toc503348732)

[2. Experimental set up 3](#_Toc503348733)

[2.1. Shared memory machine 3](#_Toc503348734)

[2.2. Computer cluster 3](#_Toc503348735)

[2.3. Test data 3](#_Toc503348736)

[3. General notes 3](#_Toc503348737)

[4. OpenMP 4](#_Toc503348738)

[4.1. Implementation 4](#_Toc503348739)

[4.1.1. Without parallelized partition 4](#_Toc503348740)

[4.1.2. With parallelized partition 4](#_Toc503348741)

[4.2. How to run 4](#_Toc503348742)

[4.3. Experimental results 4](#_Toc503348743)

[5. Cilk 4](#_Toc503348744)

[5.1. Implementation 4](#_Toc503348745)

[5.2. How to run 4](#_Toc503348746)

[5.3. Experimental results 4](#_Toc503348747)

[6. MPI 5](#_Toc503348748)

[6.1. Implementation 5](#_Toc503348749)

[6.2. How to run 5](#_Toc503348750)

[6.3. Experimental results 5](#_Toc503348751)

[7. Summary 5](#_Toc503348752)

# Problem statement:

## Problem definition:

Sorting an array A of n elements by the Quicksort algorithm. The parallel algorithms should not suffer from the O(n) time partition bottleneck. In other words, find parallel approaches to doing partition in parallel.

The input is given in an array of some C basetype (int or double, …) with “<“ as the comparison function.

## Task:

Implement the specified problem in all three frameworks: OpenMP, Cilk and MPI.

## General Notes:

According to the task, we have to take a close look on the partition function because this is the time bottleneck. Our straightforward and sequential approach to the partition function looks like this:

1 void partition**(**int a**[],** int start**,** int end**,** struct partitionResult **\*** result**,**

int pivotValue**)** **{**

2 int aa**,** i**,** j**;**

3 i **=** start**-**1**;** j **=** end**+**1**;**

4

5 **for** **(;;)** **{**

6 **while** **(++**i**<**j**&&**a**[**i**]** **<** pivotValue**);**

7 **while** **(**a**[--**j**]** **>** pivotValue **&&** j**>=**start**);**

8 **if** **(**i**>=**j**)** **break;**

9 aa **=** a**[**i**];** a**[**i**]** **=** a**[**j**];** a**[**j**]** **=** aa**;**

10 **}**

11

12 result**->**smaller **=** j**-**start**+**1**;**

13 result**->**larger **=** **((**end **-** start**)** **+** 1**)** **-** result**->**smaller**;**

14**}**

In this algorithm, start and end are the starting and end indices to partition in the array a. Before the end of the method, two integer values are set in the struct partitionResult: smaller contains the number of elements smaller than the pivot value whereas larger contains the values larger than the pivot value. This method is used in all our implementations (sequential, OpenMP, Cilk and MPI) but we use the values in the struct differently. To evade the time bottleneck here, we use different approaches in the implementations. These approaches are explained further down in this document.

## Referential, sequential implementation:

As sequential reference implementation, we used a straightforward approach, which looks like this in pseudocode:

1 void quicksortS**(**int arr**[],** int low**,** int high**)** **{**

2 **if** **(**low **<** high**)** **{**

3 pivotIndex **=** get random value between low and high

4 pivotValue **=** value at position pivotIndex

5 //switch pivot to first element

6 swap**(**pivotIndex**,** low**)**

7

8 partition**(**arr**,** low**+**1**,** high**,** **&**result**,** pivotValue**);**

9 pi **=** new pivot Value index

10 swap**(**low**,** pi**)**

11

12 quicksortS**(**arr**,** low**,** pi **-** 1**);**

13 quicksortS**(**arr**,** pi **+** 1**,** high**);**

14 **}**

15**}**

//TODO Zeiten dazu (sequentiell) oder nicht?

# Experimental set up

To test our solutions, we ran them on different systems. During the implementation phase, we tested them on our own machines. The final benchmarking values were then gathered from the machines of the Research Group for Parallel Computing, TU Wien. We started the algorithms in a benchmarking loop with 25 iterations to gather more than one value and our system outputs the best, worst and average runtime of the algorithm(s).

The machines:

## Shared memory machine

To benchmark the implementations of Cilk and OpenMP we used the machine *saturn.par.tuwien.ac.at*. This is a shared memory machine with four sockets: four AMD Opteron 6168 (12 Cores, 1.9 GHz, 12 MB cache) and 128GB DDR3 memory and a Debian System is running on this machine.

## Computer cluster

To benchmark the implementation of MPI we used the machine *jupiter.par.tuwien.ac.at*. This is a computer cluster consisting of 35 compute nodes where each node has two AMD Opteron 6134 (eight cores, 2.3 GHz, 12 MB cache) and 32 GB DDR3 memory per node. The network between the nodes is a QDR InfiniBand and Gigabit Ethernet and CentOS 6 is running on this cluster.

## Test data

//TODO write something about our test data (different data generators)

# General notes

* All our implementations for the three different frameworks have one thing in common: the use of a constant called *UNIT*. This constant is defined in every c file and is used, to check if a recursive call to the quicksort method is done with an array, which has less than *UNIT* elements. If so, the algorithm switches to the sequential implementation described in 1.4.
* To find a pivot Value, all our implementations use the same method: we use a random pivot element. This works fine in most cases but can be a huge performance killer in some cases. We could improve this, so that some elements are taken from the array and the median value is chosen as pivot. In our MPI implementation, we use the simple approach, that the process with rank 0 chooses a random pivot from its values and the pivot value is broadcasted afterwards. This could also be improved so that all processes choose a pivot value and they exchange them and agree on the “best” one.
* //TODO Explain parameters to call main and mpi

To generate the file *main* without the make-file use the following command:

gcc -fopenmp -fcilkplus -o main main.c generator.c quicksortS.c quicksortO.c quicksortC.c shared.c

# OpenMP

## Implementation

We implemented two different versions: one version, where the partition part is not parallelized, and one version with a parallelized partition part. With this approach, we can compare the two different implementations. In theory, the second implementation should not suffer from the O(n) time bottleneck in the partition part.

## Without parallelized partition

This implementation differs only a little bit from our sequential implementation as described in 1.4. The difference is that we call the recursive part inside an omp statement:

1 #pragma omp parallel

2 #pragma omp **for**

3 **for(**int k **=** 0**;** k **<** 2**;** k**++)** **{**

4 **if(**k **==** 0**)** **{**

5 quicksortOImpl**(**a**,** pi**,** maxThreads**);**

6 **}else** **{**

7 quicksortOImpl**(**a**+**pi**+**1**,** n**-**pi**-**1**,** maxThreads**);**

8 **}**

9 **}**

With this approach, we tried to mimic the behavior of a task parallel quicksort (like Cilk) to spawn two tasks that each sort a part of the array.

We did not intend to use this as our final solution for the OpenMP task but just to see the speedup or possible overhead of the OpenMP calls. Furthermore, we can compare this minimal approach to a minimal Cilk approach that just changes the loop to two cilk\_spawn statements and the rest of the algorithm stays the same.

## With parallelized partition

This implementation tries to circumvent the O(n) time bottleneck in the partition part. To achieve this, we split the array in parts, where each part is n/threads elements large. N is the array size of this recursive call and threads is the number of processes. Each process independently starts to partition its part of the array. Afterwards a single process builds a prefix sum overall all processes. In this step the sums of smaller and larger values are built. The next step is parallelized again: every process writes its partitioned array back into the main array to the desired position and the pivot value is written into the correct position. After this step, the pivot value is in the correct position and we can recursively advance. This is done in the same way as described in 4.1.1.

To be able to write back in parallel we need a second array. In our implementation there is just one helper array, which has the same size as the input array. Every process copies the part he partitioned to the helper array in every recursive call. Then he can write from the helper array to the correct index into the input array.

## How to run

To run this program, it has to be compiled and linked first. This can be done using our make-file with the command:

make main

This command creates a file *main* that can then be run with *./main* where we can also pass some arguments (see 3). To run one of the OpenMP implementations, specify the argument –a o or –A o when –a has another argument for comparison. To run the program without the make-file see section 3.

## Experimental results

//TODO plot usw

One thing that really surprised us is the difference in runtime on different systems. We implemented and tested both implementations on our own systems with CPUs that have much higher clock speeds than the Saturn machine and the runtimes with the same number of threads and the same input sequences differ hugely. For example, with the sequence parameter –s 0 (periodic sequence) and using only 2 threads the mean runtime on a system with 64 bit Windows 10 (Build 16299.192) running on an Intel Core i5-4670k on 3.4 GHz (overclocked to 4.2 GHz) with 16 GB DDR3 memory is 0.492 seconds with 10 million elements to be sorted. On comparison, the runtime with the same parameters on Saturn is 1.427 seconds. For sure, we have to take into account that the pivot value is chosen randomly but we ran this test multiple times and the Saturn machine is slower with a factor around 3.

# Cilk

## Implementation

## How to run

## Experimental results

# MPI

## Implementation

## How to run

## Experimental results

# Summary

**von seiner Angabe:**

Allowed: • Although sequential Quicksort is an in-place algorithm, your parallel solution may use extra arrays of size O(n); try to keep extra space small. Explain where and why needed.

• Optimal pivot (exact median) solution is not required. It is therefore acceptable to test on randomly generated input only (or pick pivots randomly). Introduce performance counters to measure the work imbalance (different sizes of partitioned arrays, depth of recursions)

Hint: Use MPI solution as idea also for OpenMP and Cilk; consider load balancing by prefix-sums (as in lecture)