CS515 Parallel Programming: Assignment #2

Due on May 4th, 2016 at 11:59pm

Jingke Li Spring 2016

Konstantin Macarenco

Primes

As you can see from the graph - parallel Primes algorithm performed better than sequential version in all cases, with one exception: OMP with one thread is always slower than the sequential implementation. This can be explained by overhead that is introduced by using OMP.
There are two possible solutions to the homework

1. Each thread works independently and starts crossing out multiples as fast as it finds a non zero entry. The main disadvantage of this approach is that, since there is no synchronization some thread will perform redundant jobs, and cross out some entries multiple times.

2. Implement synchronization to avoid redundant job. In this case we need set of explicit locks, waits, array partitioning, and a way to notify other threads about index being crossed out. This most likely will result in worse performance than the first case.

I used the first solution, and you can observe that increasing number of threads gives small performance boost, with 128 being the fastest.

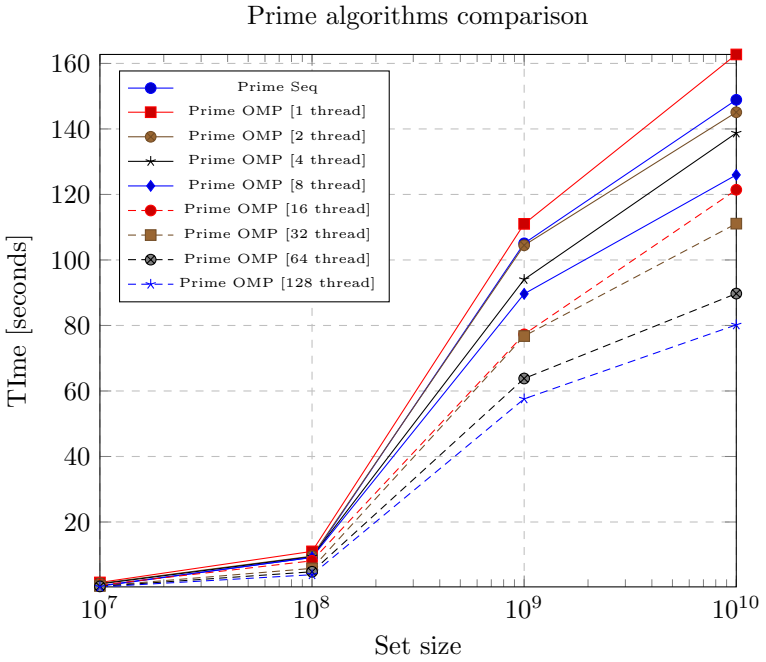


Table 1: Sequential execution (Time [sec])

	10^7	10^8	10^9	10^{10}
prime	0.533	9.616	105.080	148.891
quicksort	4.247	39.907	425.748	653.039

Table 2: Prime omp (Time [sec])

	10^7	10^8	10^9	10^{10}
1	1.548	11.032	111.028	162.742
2	1.268	9.572	104.474	145.092
4	1.097	9.413	94.059	138.766
8	0.400	9.203	89.660	125.969
16	0.753	8.149	77.292	121.433
32	0.478	5.845	76.765	111.086
64	0.337	4.807	63.811	89.739
128	0.192	3.904	57.598	80.190

Quick sort

Quick sort is much more suitable for parallelization since we only need to wait for the current chunk of the array to be partitioned and no other synchronization. At first I had trouble with placing OMP statements in the correct place. When pragma's are added to the quickSort routine, every time the method called new parallel region is declared, which creates exponential of number of threads, and leads to poor performance.

After some research I used "# pragma omp task firstprivate(array, low, high)" to declare a region that should be executed by a separate thread thread, and firstprivate indicates that array, low and high must be initialized before entering parallel region. In this case an existing thread is assigned new parallel region.

This approach nearly doubles the performance as soon as number of threads is two or more. Further increase of threads, gives slower performance boost, about %20 percent with each level (up to 16 threads). After number of threads reached 32 we can see some performance degradation by %8 in average (with each threads increase). I assume that this problems comes from frequent task switches. Each time new task is created some thread is reassigned to it and the Program performs task switch, as number of array partitions grows task switches executed much more frequent. Each task switch most likely is performed along local cache update, since threads are reassigned to random regions of array.
Hence the slow down. Another thing worth mentioning, one thread OMP quicksort struggles from the same issue as prime, it is slower than sequential version by a constant factor.

qsort-pthd

Pthread version of qsort is faster than sequential, but ($\approx 2x$ in the best case) slower than OMP. This is expected, since qsort-pthd has multiple bottlenecks, such as shared number of producing methods, and queue. This version's performance peak is at NUMTHREADS=8, as number of threads increased performance drops to sequential level or much worse. This probably happens due to couple of reasons. First, bottlenecks, despite queue being unlimited, it must be locked for read/write operations, hence large number of threads will be waiting before it can push a task to a queue, this will also delay tasks consumption. Second, same problem as OMP version - cache locality. None of the tested arrays will fit into the L1,L2 and L3 caches, therefore when task is pushed by one thread and consumed by another, data must be propagated through all levels.

Other issues

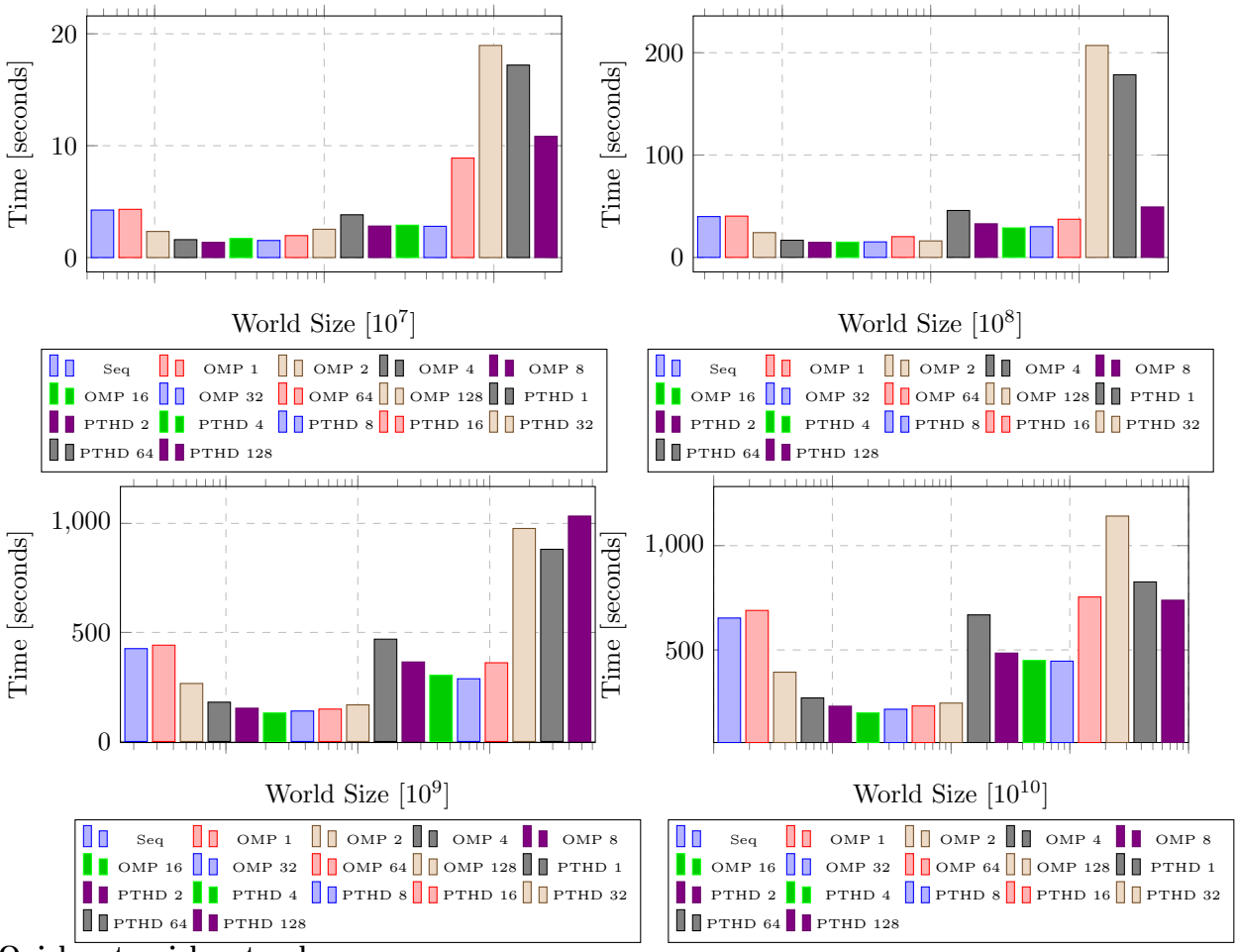
When number of threads is greater than number of available cores, there most likely be slowdown, unless some cores are waiting and can run additional threads.

Table 3: Quick Sort omp (Time [sec])

	10^7	10^8	10^9	10^{10}
1	4.313	40.367	441.530	689.361
2	2.336	24.254	265.905	393.473
4	1.603	16.793	180.163	270.030
8	1.359	14.699	152.552	230.375
16	1.710	14.954	130.645	197.896
32	1.527	15.191	139.620	215.855
64	1.965	20.295	148.494	231.565
128	2.531	16.134	167.827	245.310

Table 4: Quick Sort pthread (Time [sec])

	10^7	10^8	10^9	10^{10}
1	3.825	45.857	469.146	668.053
2	2.820	32.870	364.625	485.091
4	2.897	28.845	303.976	449.671
8	2.796	30.014	287.594	446.297
16	8.892	37.327	361.015	753.872
32	18.957	207.082	976.707	1141.691
64	17.203	178.471	881.260	826.079
128	10.841	49.304	1033.700	738.495



Quicksort quicksort only