

Report

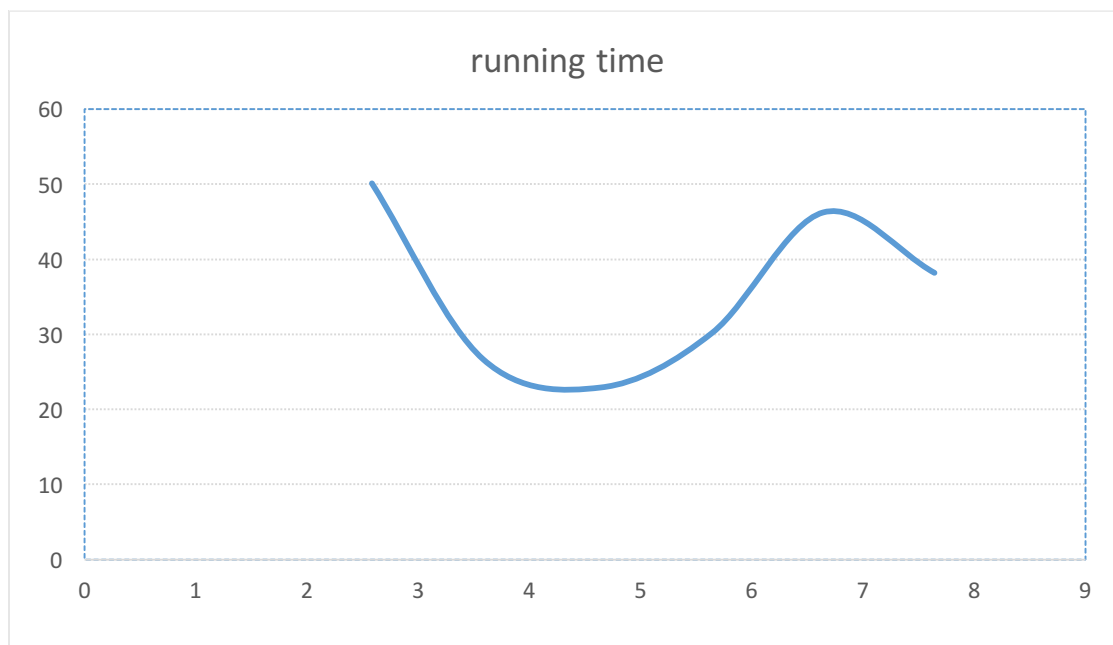
Test result for different memory size

Memory size (MB)	Elapsed time(m:ss)	Maximum resident set size(KB)
200	38.19	206224
100	46.29	103920
50	30.16	50708
25	22.88	26136
12	26.65	13812
6	50.14	5600

Black pass 2PMMS

Red only pass Phase I

Only Memory size of 200 MB and 100 MB can perform 2PMMS, others only only accomplish phase I because our optimal block size is 2MB and memory size is too small for merging in phase II.



Question:

1. Is there any difference in performance in your experiments? Explain why there is a difference or why there is no difference.

The elapsed time is different for different memory size. Any memory size less than or equal to 50MB which is too small for 2PMMS. Memory size with 200MB has the least elapsed time. Although only 200 MB and 100 MB satisfy 2PMMS, it shows a tendency that the elapsed time increases as memory size decreases. The reason is that large

memory size means that less sub arrays to be sorted separately in main memory, so it takes less time in phase I, and in phase II less file open streams needed to read each sub list file. Memory size 200MB has the most maximum resident set 201MB, and 100MB only takes half of it.

2. Which program is faster: your implementation or Unix sort? Which one uses less memory? Explain the difference (or the lack of difference) in performance. If there is a difference - what in your opinion could explain it?

My implementation is faster than Unix sort. Unix sort use 3286MB memory while my implementation only uses 201MB memory for 200MB allocated memory size. The elapsed time for Unix sort is 1:24.46 which is longer than my implementation. User time of Unix sort is 115 seconds, and my implementation takes around 10 seconds. In my implementation read binary file directly, while Unix sort read input line by line and extract sort keys from each line of input. Unix sort takes much time to read the inputs and then convert them into binary format.

Conclusion:

In 2PMMS, we perform 4 I/Os for per each block, so the running time depends on the memory size which can hold large unsorted array. More memory means faster sorting sub array in main memory. Unix sort take more memory and more time than 2PMMS.