Design and Analysis of Algorithms
Programming Assignment 1
Submission Date: Feb 15, 2018

The sorting algorithm that is capable of sorting large data that cannot fit in the main memory of the computer is called external sorting algorithm. This massive amount of data is stored on a slower external memory device usually a hard disk drive. One example of external sorting is the external merge sort algorithm, which is a K-way merge algorithm. It sorts chunks that each fit in RAM, then merges the sorted chunks together.

External sorting algorithms are analyzed using external memory model. unlike RAM model, external memory model has a cache or internal memory of size M and an external memory of unbounded size divided into blocks of size B. In this model read/write operations are much faster in cache as compare to the external memory. Also reading long contiguous blocks of external memory are faster than reading randomly. The running time of this model is calculated by counting the number of memory transfer operations between external memory and internal memory (cache).

The external merge sort algorithm first sorts *M* items at a time and puts the sorted lists back into external memory. It then performs a k-way merge on those sorted lists. To do this merge, *B* elements from each sorted list are loaded into internal memory, and the minimum is repeatedly outputted in an output buffer of size B. once this output buffer is full, the sorted data is spilled to hard disk in an output file. The k-way merge operation is performed using a min heap of size k.

Write a C++ function ExternalMerge that takes an input file of **N** integers, cache size **M**, block size **B** and name of output file as parameter. It first takes chunks of M integers from input file (stored on the disk) one by one, sort them individually using merge sort, and output them in different output files. There will be total **N/M** output files (sorted chunks). It then reads **B** integers from each chunk(file) one at a time and apply **N/M**-way merge on these sorted chunks using a binary min heap structure. The merged elements will be stored in a buffer of size **B** placed in cache (main memory RAM). Once **B** elements are merged, these merged **B** elements will be outputted to the output file on the external memory.

First generate 1 billion integers randomly (integers may repeat) and store them on a file, then call the external merge sort to sort these 1 billion element using M = 2^{24} and B=512.

Note: Plagiarism of any sort is not acceptable and will be severely punished.