CS307 - HW 4

Samuel Lee

24774

1/3/2021

In this report, I will read the given file (loremipsum.txt), and after successfully reading the file and counting the character ‘a,’ I will analyze and compare the differences between the three methods. The three methods are implemented into a cpp file that uses fstream or istream, a C file that uses fopen for reading the file and a C file that utilizes memory mapping with mmap function on the file.

First of all, the cpp file took a total of 5.87 seconds to execute the program. Out of all the three implementations, it took the longest time. In general, the c++ fstream inherits functionality where it converts objects into human readable text, while fopen in C deals with blocks of bytes. This may be the cause or reason behind the slower speed in the C++ program compared to the C program. Needless to say, the text file that was given was over 250MB, so 5.87 is a considerably fast program.

The second implementation that is written in C using fopen for reading the file took a total of 3.59 seconds for the program execution. It took around 60% of the time compared to the C++ program during execution. Although I originally expected the fstream and fopen to be quite similar in terms of performance, the C program with fopen performed much better, and this may be because of the reason mentioned in the paragraph above, where fopen deals with block of bytes directly, while fstream in C++ is a file backed implementation of a stream of bytes.

Out of all the three implementations, the C file that uses mmap function was the fastest. In memory mapping, instead of reading blocks of data, it maps the content of the file to a pointer, and the OS handles filling in the data. The reason for its speed has to do with the fact that the data on the disk can be mapped directly to the memory without any unnecessary copying. This allows it to give the fastest execution time with 1.37 seconds, when compared to

the 5.87 seconds of the C++ program. This is a huge improvement, taking around 23% or less than 1/4 of the time to the C++ program. One disadvantage that could arise from using mmap could be that it can be unstable, where a slight mistake with memory mapping can lead to program crash, where with fstream of C++ or fopen of C, you are less likely to have bus errors.

In conclusion, the mmap function outperformed both implementation of fstream in C++ and fopen in C by a large margin. It was able to execute the program in 1.37 seconds by directly mapping the content of the “loremipsum.txt” file to a pointer without any unnecessary copying. The compilation, execution and results of the three different programs can be found in the figure below.

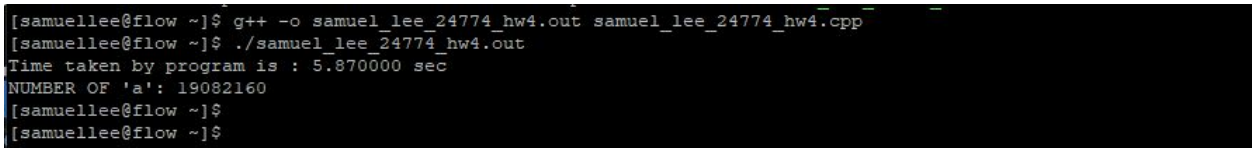

Figure 1. C++ File

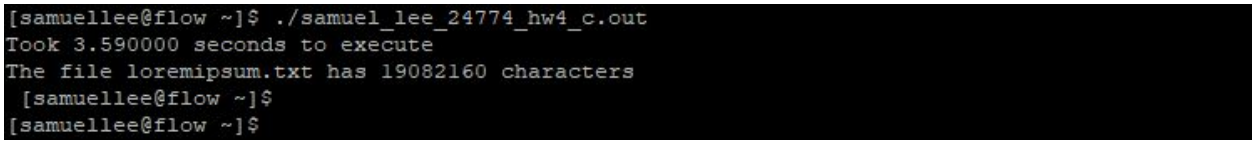

Figure 2. C file that uses fopen

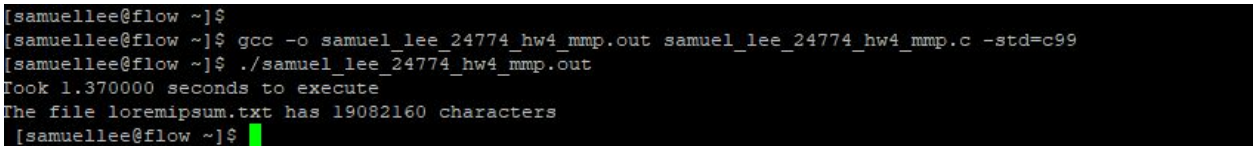

Figure 3. C File using mmap