Parallel Huffman Coding

High Performance Computing for Data Science - UniTn 2022/2023

Andrea Cristiano: 229370

Gabriele Padovani: 229207

# Introduction

Nowadays, the transmission through the network as well as the storage of a large amount of digital information has become a fundamental part of the human activity. Near those, as part of the process itself, it is possible to find algorithms of compression and decompression which play an essential role. For many applications the algorithms must be lossless, the information must be reconstructed as it was before the compression, and as fast as possible, some of them can operate in real time situations. Huffman coding[1] is likely one of the simplest approaches for data lossless compression. Parallelization instead, in many cases allows processes to reach better performances with respect to their serial version. On the one side, the project aims to develop an application that relays on Huffman coding algorithm to deal with a lossless compression of texts with variable length.  
On the other side, it aims to exploit the parallel paradigm[2] to improve the performances of the algorithm itself. For this purpose MPI[3] library and OpenMP[4] API are used. The application is entirely written in C and the source code is available on GitHub[5]. The document is organized as follows: the next section will briefly introduce the Huffman coding algorithm with all its main characteristics. This is followed by two parts, which will describe the approaches used to parallelize respectively the compression and the decompression phases. Section 3 will introduce the main obtained result, by focusing on specific parts of the algorithm as well as by giving a general overview of this last.  
The final section will report the conclusions of the project.

# Huffman coding

# Compression

# Decompression