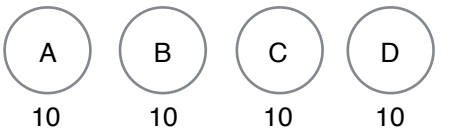
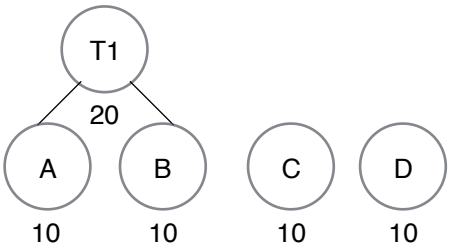
Writeup PA2

CLÉMENT ROCHE
U07034316
ax003915@acsmail.ucsd.edu

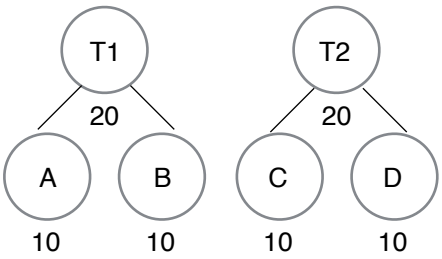
Let's "compress" the checkpoint1.txt file. The file is simply : "abcd" repeated 10 times.



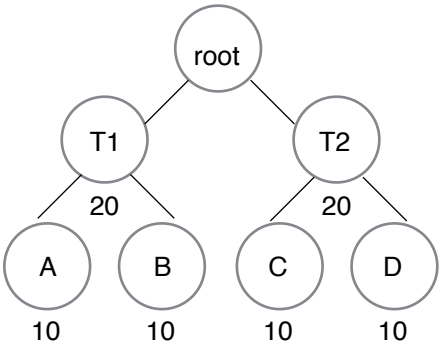
At the beginning of the Huffman's algorithm, the tree looks like this. In my implementation, when nodes have equals count, they're compared on their symbol.



We construct the first sub tree with the two lowest count and symbol.



Then we get two subtrees.



This is the last step of the construction of the tree. We know the algorithm is finished because there is only one element left in our priority queue (the root).

How to get the the code word for each byte ? In my implementation, I created a recursive function getEncodedSymbol(std::string &, HCNode *). This function fill the string by appending '0' or '1' depending on its way from the leaf to the root. For example, to encode 'C' with the tree we just created, we start from the 'C' leaf. It is the left side of its parent node, so a '0' will be appended to the string. T2 however is the right side of its parent node, so a '1' will be appended to the string. Since the '0' or '1' are added to the string after the recursive call with the parent node, the string will be in the correct order ("10" for C).