

Data compression

anhtt-fit@mail.hut.edu.vn

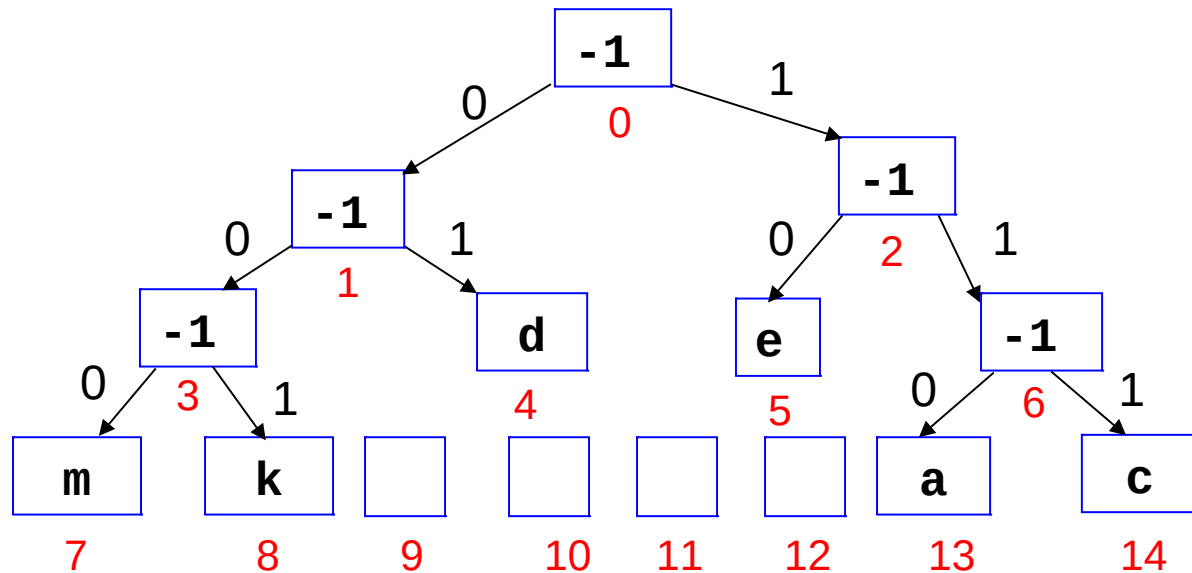
File Compression

- The Huffman Code can be used to compress a file.
- The compressed file may be organized as the following

HM	Huffman Tree	Size	Data ...
----	--------------	------	----------

- HM is the prefix code of the compressed file
- Huffman Tree is an array representation of the Huffman tree generated for the coding data
- Size specifies the compressed data's size in number of bits
- Data contains the bits of compressed data

Array representation of the Huffman Tree



All the leaf nodes in the Huffman tree are labeled as characters

15	-1	-1	-1	-1	d	e	-1	m	k					a	c
----	----	----	----	----	---	---	----	---	---	--	--	--	--	---	---

The array's size = $(2^0 + 2^1 + \dots + 2^n)$ where n is the tree's maximal level

The children of node i are at $2i+1$ and $2i+2$

The parent of node i is at $(i-1)/2$

Implementation

- Giving a data structure as the following to represent the Huffman tree in array

```
typedef struct {  
    int size;  
    int * nodes;  
} HuffmanTreeArray;
```

- Write a function to convert a Huffman tree in the array format
 - HuffmanTreeArray tree2array(HuffmanTree);

Quiz 1

- Rewrite the functions in the previous lab in order to compress files using the Huffman code.
- The program should be used to compress files in a command line mode as below
 - `$ compress in_file [out_file]`
- The following functions need to be implemented
 - `HuffmanTree makeHuffman (FILE * in);`
 - `void createHuffmanTable(HuffmanTree htree, Coding* htable);`
 - `HuffmanTreeArray tree2array(HuffmanTree);`
 - `void compressFile(FILE* in, FILE *out);`

Decoding the File

- Firstly, check the prefix “HM” of the file
- Read tree in the array representation
- Once receiver has tree it scans incoming bit stream
 - Data is bit based versus byte based
- Scanning algorithm
 - Set cursor at the root of the tree
 - If the current node has value -1, read new bit
 - 0 \Rightarrow move cursor to the left child
 - 1 \Rightarrow move cursor to the right child
 - Otherwise, get the new character at the node, move the cursor to the root.

Quiz 2

- Write a program to decompress files compressed as in Quiz 1
- Use the command line mode to decompress files
 - `$ decompress compressed_file [out_file]`