



# CSCI 1102 Computer Science 2

Meeting 17: Tuesday 3/30/2021

Huffman Coding; More on Binary Trees



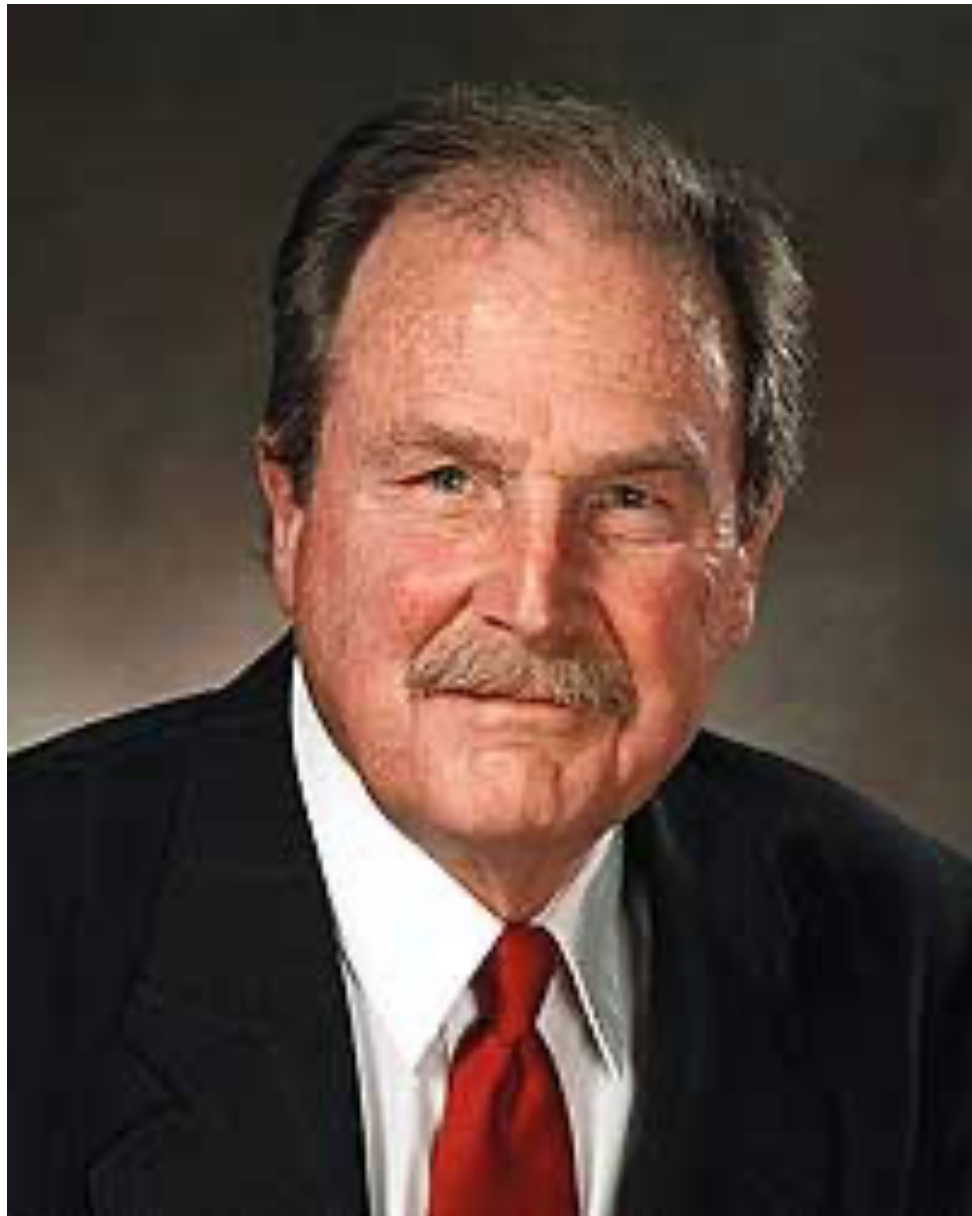


# John Backus

## Turing Award 1977

- Inventor of first high-level programming language FORTRAN
- Co-Inventor of Backus-Naur Form (BNF), a notation for describing syntax structure.

# Huffman Coding



CSCI 1102 Computer Science 2

# “ALABAMA”

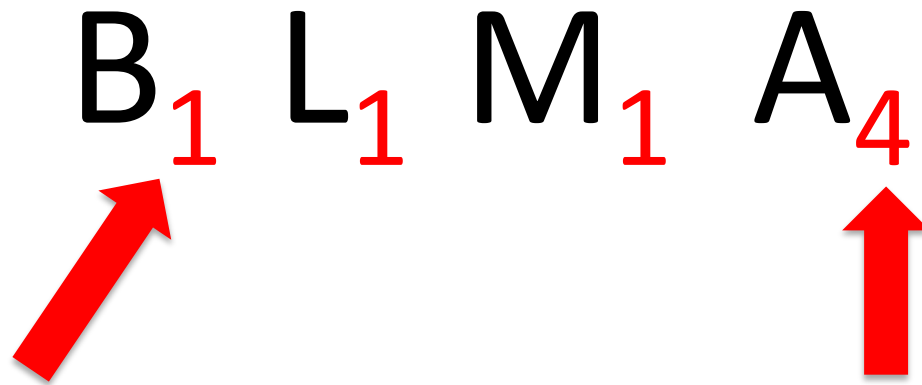
‘A’ is run of the mill, ‘B’, ‘L’, & ‘M’  
are more *interesting*, i.e., contain more  
Information than ‘A’.

Instead of an 8-bit fixed-length code for a letter, choose a variable-length code reflecting the amount of information in the letter.

“ALABAMA”

A<sub>4</sub> L<sub>1</sub> B<sub>1</sub> M<sub>1</sub>

*A Frequency Table*



Low frequency  
=> High Priority, head  
of the queue!  
**(letters arrived in  
alphabetical order)**

High frequency  
=> Low Priority,  
back of the queue!



B<sub>1</sub> L<sub>1</sub> M<sub>1</sub> A<sub>4</sub>

**Algorithm step:** while the priority queue has more than one entry, remove two entries, combine them and reinsert the result back into the priority queue.

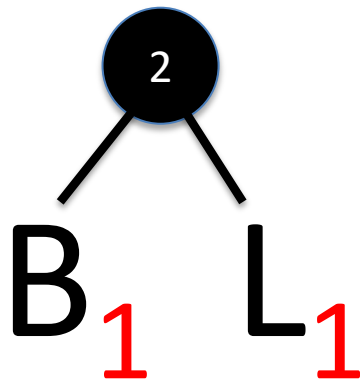
B<sub>1</sub> L<sub>1</sub> M<sub>1</sub> A<sub>4</sub>

B<sub>1</sub>

L<sub>1</sub>

M<sub>1</sub>

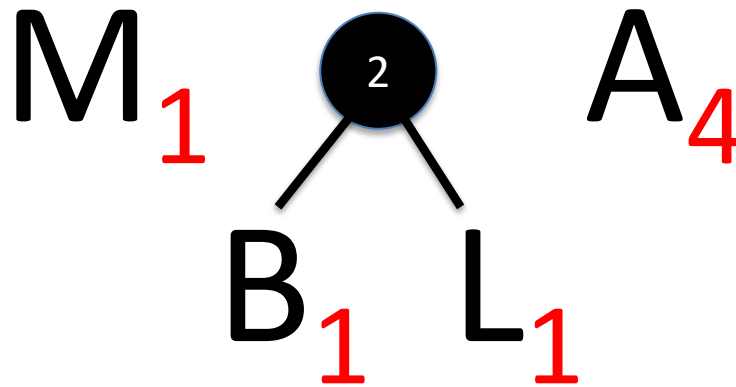
A<sub>4</sub>



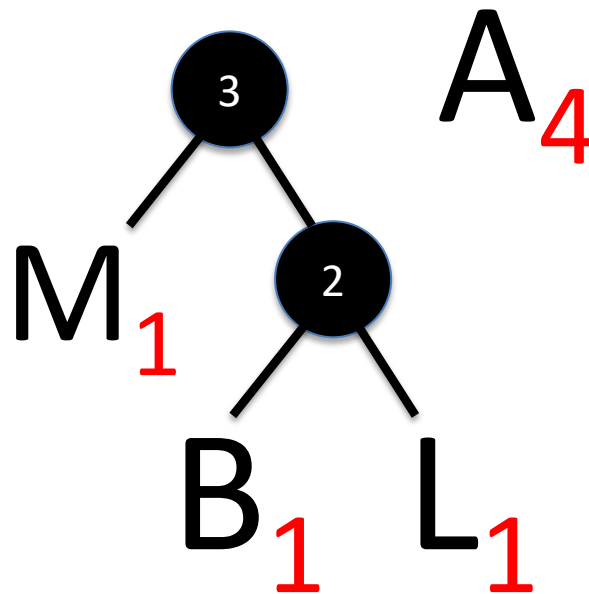
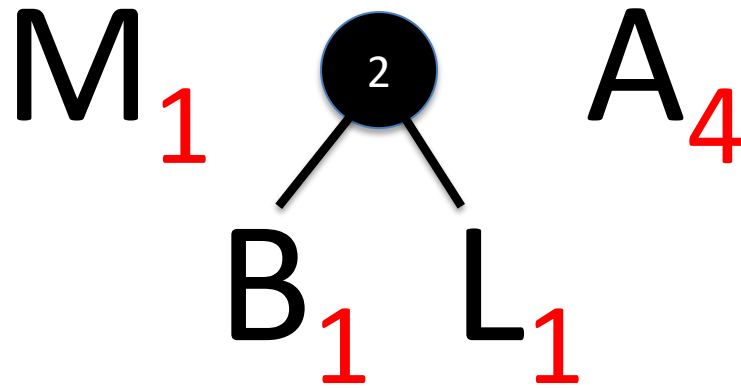
M<sub>1</sub> A<sub>4</sub>

The items in the priority queue are *weighted trees*.

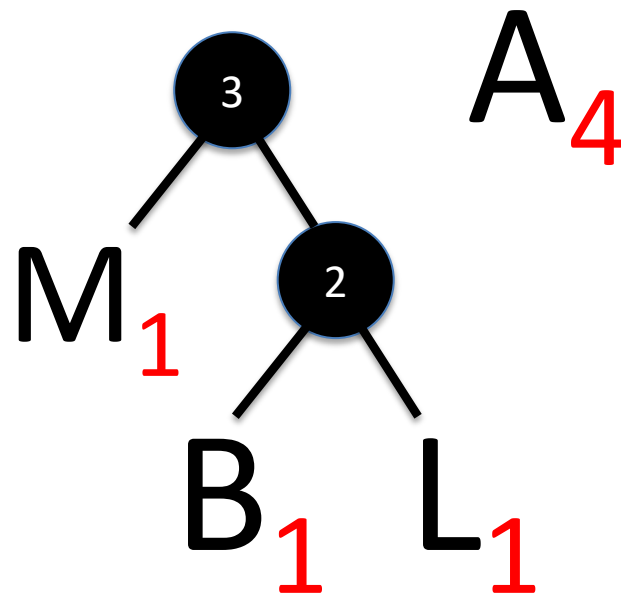
B<sub>1</sub> L<sub>1</sub> M<sub>1</sub> A<sub>4</sub>

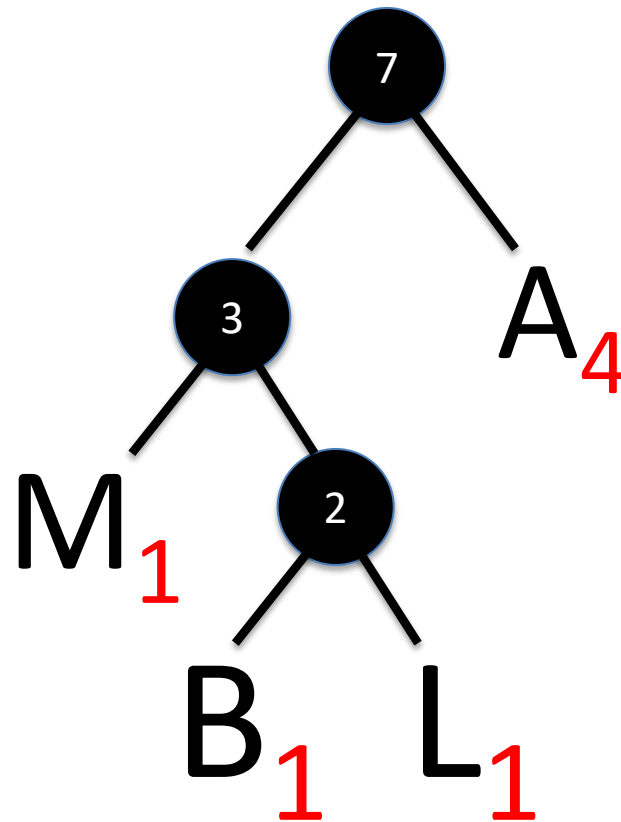


The items in the priority queue are *weighted trees*.

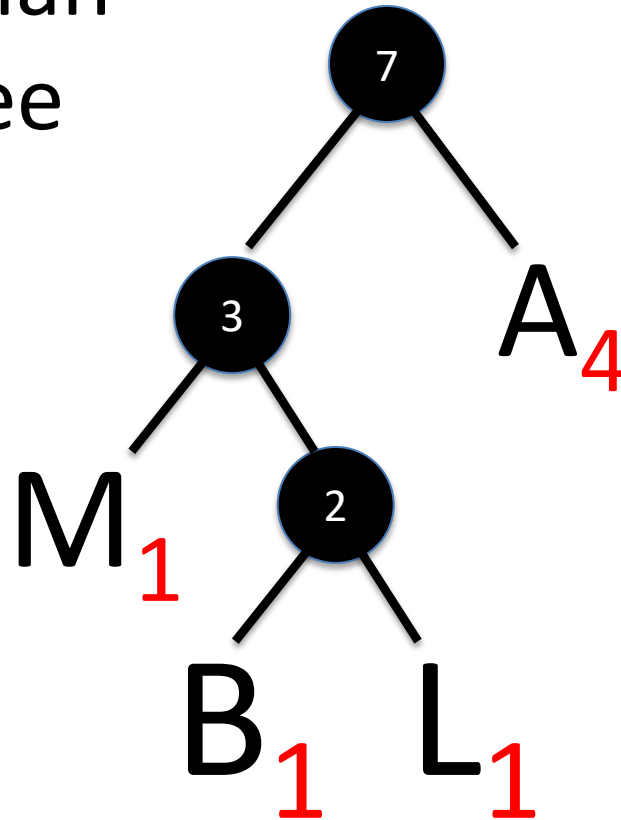




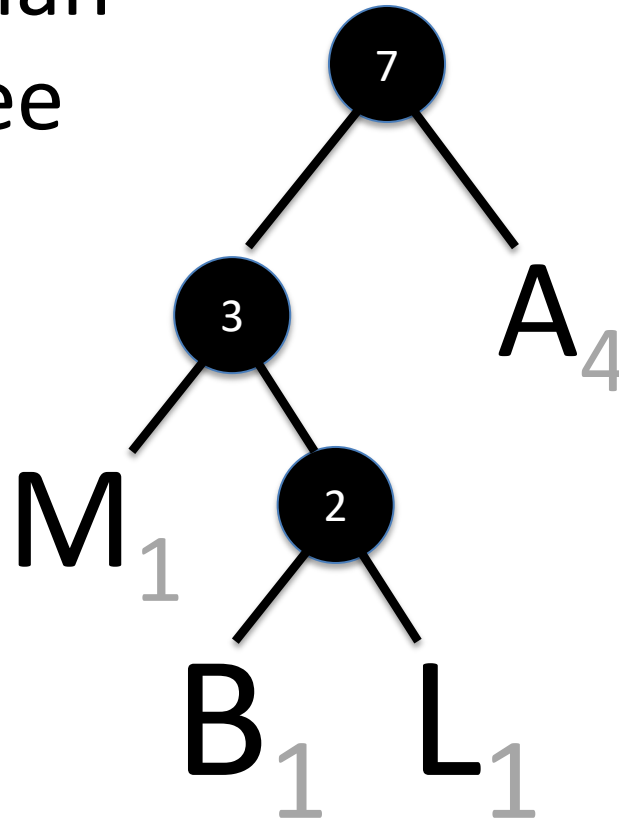




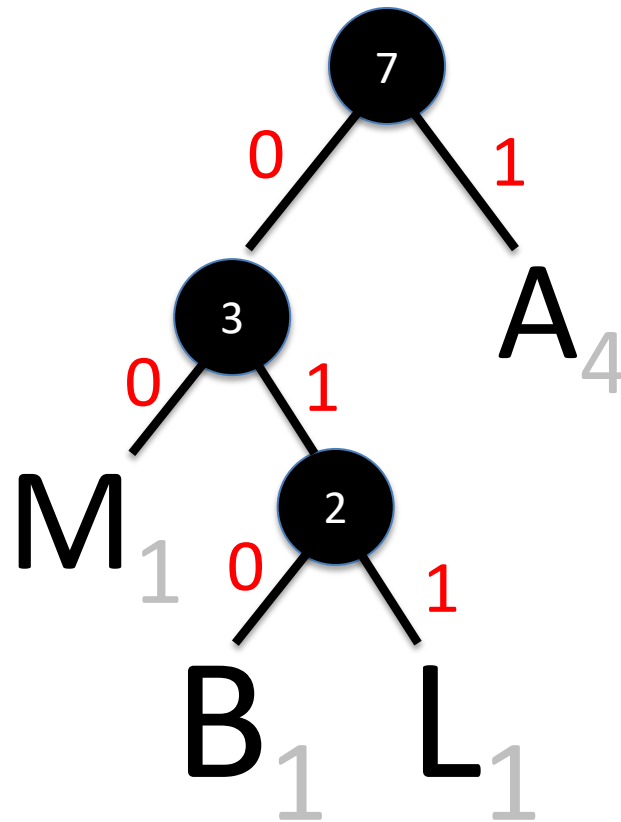
# The Huffman Coding Tree

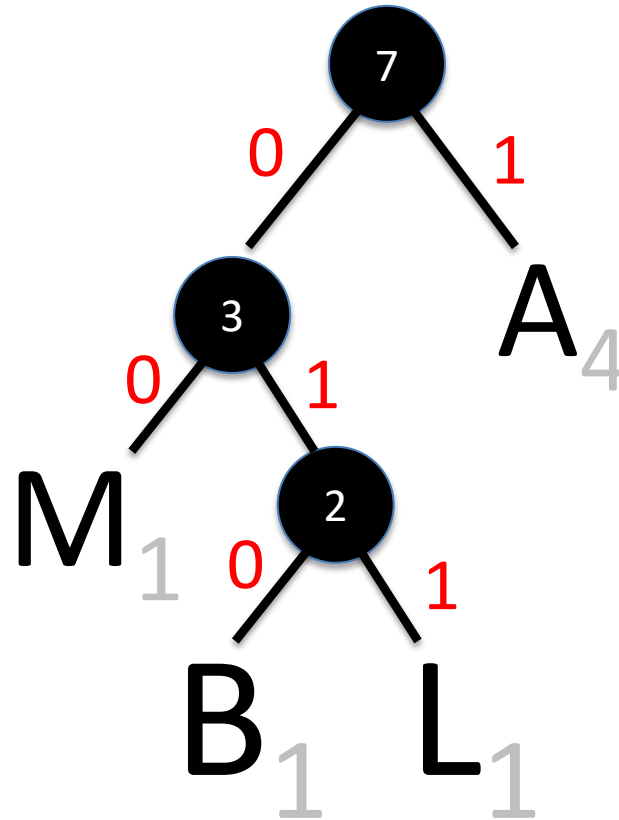


# The Huffman Coding Tree



Left = 0  
Right = 1





$M = 00, B = 010; L = 011; A = 1$



A L A B A M A

1 011 1 010 1 00 1

M = 00, B = 010; L = 011; A = 1

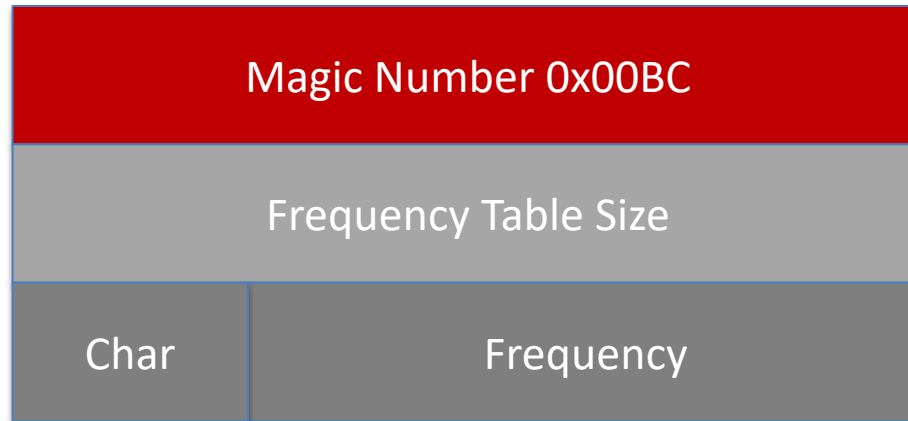
# ALABAMA

101110101001

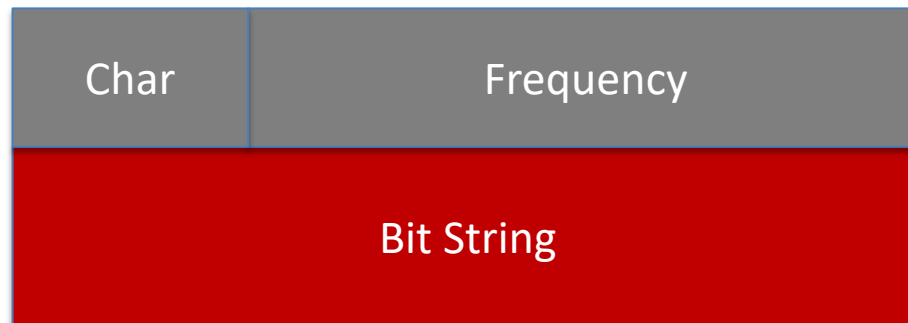
12 bits << 56 bits

M = 00, B = 010; L = 011; A = 1

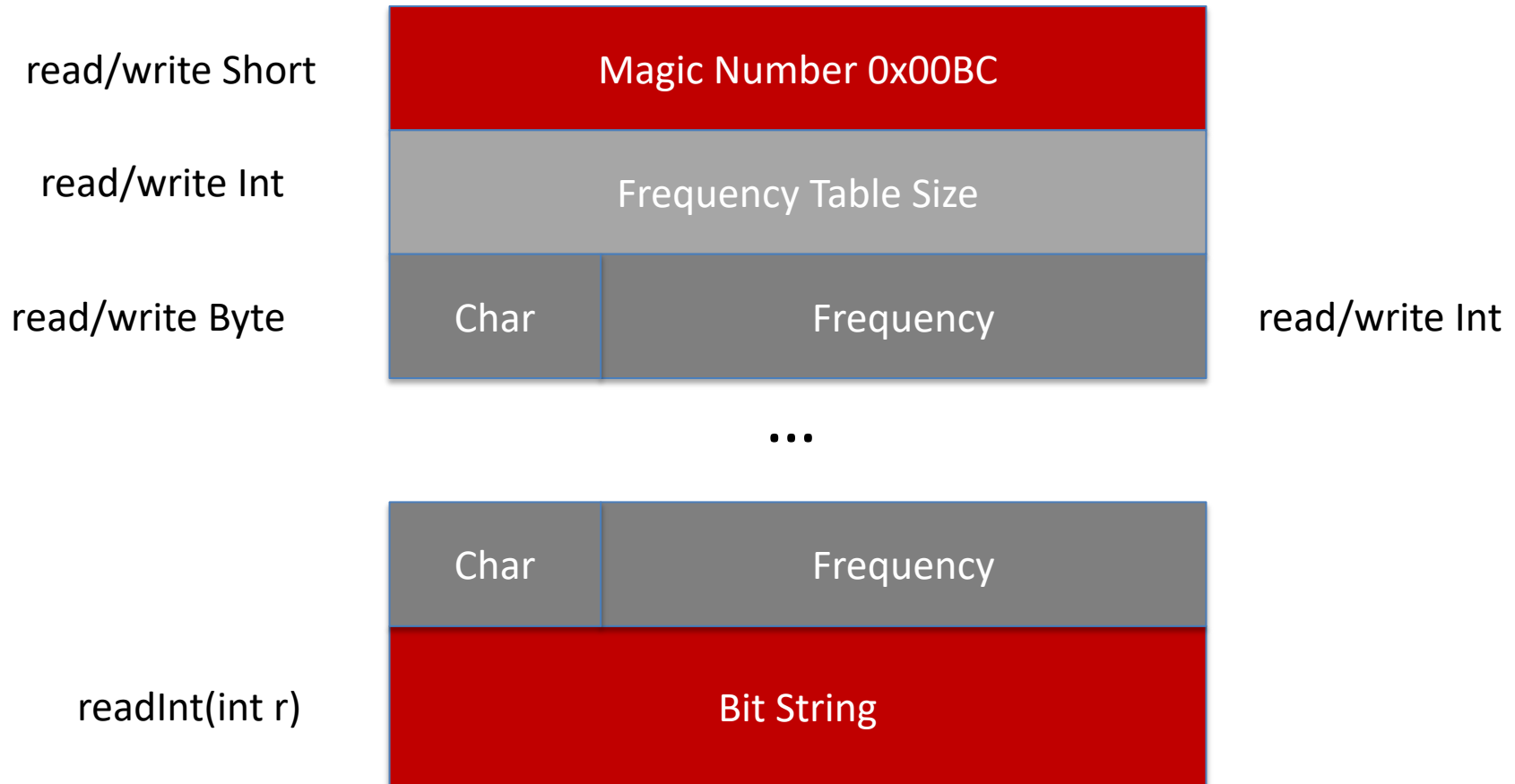
# The .zip File



...



# BinaryIn/BinaryOut



# Unix hexdump of lincoln.zip

```
00000000 0b c0 00 00 00 23 0a 00 00 00 19 20 00 00 00 f8
00000010 2c 00 00 00 18 2d 00 00 00 01 2e 00 00 00 0a 42
00000020 00 00 00 01 46 00 00 00 01 47 00 00 00 01 49 00
00000030 00 00 03 4e 00 00 00 01 54 00 00 00 02 57 00 00
00000040 00 02 61 00 00 00 66 62 00 00 00 0d 63 00 00 00
00000050 1f 64 00 00 00 3a 65 00 00 00 a5 66 00 00 00 1a
00000060 67 00 00 00 1b 68 00 00 00 50 69 00 00 00 41 6b
00000070 00 00 00 03 6c 00 00 00 2a 6d 00 00 00 0d 6e 00
00000080 00 00 4c 6f 00 00 00 5d 70 00 00 00 0f 71 00 00
00000090 00 01 72 00 00 00 4f 73 00 00 00 2c 74 00 00 00
000000a0 7c 75 00 00 00 15 76 00 00 00 18 77 00 00 00 1a
000000b0 79 00 00 00 0a cf 52 5a 78 92 51 fb 3d 78 ba 19
```

# magic number 0x0bc0

00000000	0b	c0	00	00	00	23	0a	00	00	00	19	20	00	00	00	f8
00000010	2c	00	00	00	18	2d	00	00	00	01	2e	00	00	00	0a	42
00000020	00	00	00	01	46	00	00	00	01	47	00	00	00	01	49	00
00000030	00	00	03	4e	00	00	00	01	54	00	00	00	02	57	00	00
00000040	00	02	61	00	00	00	66	62	00	00	00	0d	63	00	00	00
00000050	1f	64	00	00	00	3a	65	00	00	00	a5	66	00	00	00	1a
00000060	67	00	00	00	1b	68	00	00	00	50	69	00	00	00	41	6b
00000070	00	00	00	03	6c	00	00	00	2a	6d	00	00	00	0d	6e	00
00000080	00	00	4c	6f	00	00	00	5d	70	00	00	00	0f	71	00	00
00000090	00	01	72	00	00	00	4f	73	00	00	00	2c	74	00	00	00
000000a0	7c	75	00	00	00	15	76	00	00	00	18	77	00	00	00	1a
000000b0	79	00	00	00	0a	cf	52	5a	78	92	51	fb	3d	78	ba	19



Frequency table size  $0x23 = 35_{10}$

00000000	0b	c0	00	00	00	23	0a	00	00	00	19	20	00	00	00	f8
00000010	2c	00	00	00	18	2d	00	00	00	01	2e	00	00	00	0a	42
00000020	00	00	00	01	46	00	00	00	01	47	00	00	00	01	49	00
00000030	00	00	03	4e	00	00	00	01	54	00	00	00	02	57	00	00
00000040	00	02	61	00	00	00	66	62	00	00	00	0d	63	00	00	00
00000050	1f	64	00	00	00	3a	65	00	00	00	a5	66	00	00	00	1a
00000060	67	00	00	00	1b	68	00	00	00	50	69	00	00	00	41	6b
00000070	00	00	00	03	6c	00	00	00	2a	6d	00	00	00	0d	6e	00
00000080	00	00	4c	6f	00	00	00	5d	70	00	00	00	0f	71	00	00
00000090	00	01	72	00	00	00	4f	73	00	00	00	2c	74	00	00	00
000000a0	7c	75	00	00	00	15	76	00	00	00	18	77	00	00	00	1a
000000b0	79	00	00	00	0a	cf	52	5a	78	92	51	fb	3d	78	ba	19

Table entry: 0x0a = newline, 0x19 = 25<sub>10</sub>

00000000	0b	c0	00	00	00	23	0a	00	00	00	19	20	00	00	00	f8
00000010	2c	00	00	00	18	2d	00	00	00	01	2e	00	00	00	0a	42
00000020	00	00	00	01	46	00	00	00	01	47	00	00	00	01	49	00
00000030	00	00	03	4e	00	00	00	01	54	00	00	00	02	57	00	00
00000040	00	02	61	00	00	00	66	62	00	00	00	0d	63	00	00	00
00000050	1f	64	00	00	00	3a	65	00	00	00	a5	66	00	00	00	1a
00000060	67	00	00	00	1b	68	00	00	00	50	69	00	00	00	41	6b
00000070	00	00	00	03	6c	00	00	00	2a	6d	00	00	00	0d	6e	00
00000080	00	00	4c	6f	00	00	00	5d	70	00	00	00	0f	71	00	00
00000090	00	01	72	00	00	00	4f	73	00	00	00	2c	74	00	00	00
000000a0	7c	75	00	00	00	15	76	00	00	00	18	77	00	00	00	1a
000000b0	79	00	00	00	0a	cf	52	5a	78	92	51	fb	3d	78	ba	19

# Frequency Table

00000000	0b	c0	00	00	00	23	0a	00	00	00	19	20	00	00	00	f8
00000010	2c	00	00	00	18	2d	00	00	00	01	2e	00	00	00	0a	42
00000020	00	00	00	01	46	00	00	00	01	47	00	00	00	01	49	00
00000030	00	00	03	4e	00	00	00	01	54	00	00	00	02	57	00	00
00000040	00	02	61	00	00	00	66	62	00	00	00	0d	63	00	00	00
00000050	1f	64	00	00	00	3a	65	00	00	00	a5	66	00	00	00	1a
00000060	67	00	00	00	1b	68	00	00	00	50	69	00	00	00	41	6b
00000070	00	00	00	03	6c	00	00	00	2a	6d	00	00	00	0d	6e	00
00000080	00	00	4c	6f	00	00	00	5d	70	00	00	00	0f	71	00	00
00000090	00	01	72	00	00	00	4f	73	00	00	00	2c	74	00	00	00
000000a0	7c	75	00	00	00	15	76	00	00	00	18	77	00	00	00	1a
000000b0	79	00	00	00	0a	cf	52	5a	78	92	51	fb	3d	78	ba	19



# Start of Bit Stream

00000000	0b	c0	00	00	00	23	0a	00	00	00	19	20	00	00	00	f8
00000010	2c	00	00	00	18	2d	00	00	00	01	2e	00	00	00	0a	42
00000020	00	00	00	01	46	00	00	00	01	47	00	00	00	01	49	00
00000030	00	00	03	4e	00	00	00	01	54	00	00	00	02	57	00	00
00000040	00	02	61	00	00	00	66	62	00	00	00	0d	63	00	00	00
00000050	1f	64	00	00	00	3a	65	00	00	00	a5	66	00	00	00	1a
00000060	67	00	00	00	1b	68	00	00	00	50	69	00	00	00	41	6b
00000070	00	00	00	03	6c	00	00	00	2a	6d	00	00	00	0d	6e	00
00000080	00	00	4c	6f	00	00	00	5d	70	00	00	00	0f	71	00	00
00000090	00	01	72	00	00	00	4f	73	00	00	00	2c	74	00	00	00
000000a0	7c	75	00	00	00	15	76	00	00	00	18	77	00	00	00	1a
000000b0	79	00	00	00	0a	cf	52	5a	78	92	51	fb	3d	78	ba	19