

9 no.s
6 bits/no.) 36 bits

$$\text{Compression Ratio} = \frac{288-36}{288} = 0.875 \rightarrow 87.5\% \text{ savings}$$

Text Compression Methods

1. Special purpose compression

- All numeric fields

123 \rightarrow 30 31 33 in hex ASCII

F0 F1 F3 EBCDIC

\downarrow
0 1 3
0000 0001 0011

All alphabetic

⌊ (L) 26 letters + blank = 27 characters

$$\lceil \log_2 27 \rceil = 5 \text{ bits}$$

2. Variable Length Compression

More frequent \Rightarrow allocate less no. of bits

Less frequent \Rightarrow more " " "

Char	Ambiguous	Not instantaneous	
		Unambiguous	Instantaneous
A	1	0	0
B	10	10	10
C	11	100	110
D	100	1000	111

Huffman Coding

F: .05

Space: .10

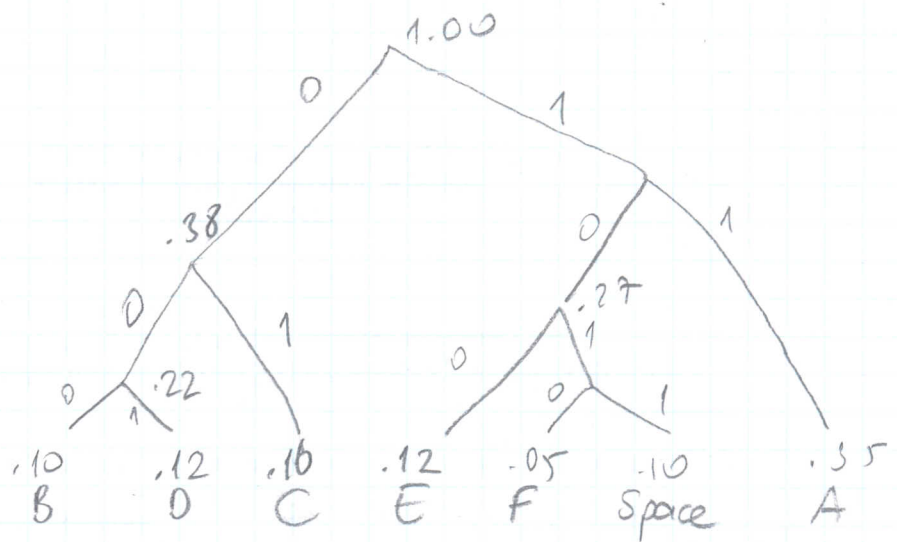
B: .10

D: .12

E: .12

C: .16

A: .35
1.00



B: 000 (3)

E: 100 (3) A: 11 (2)

D: 001 (3)

F: 1010 (4)

C: 01 (2)

Space: 1011 (4)

A BAD FEED

The average weighted code length

$$= \sum_{i=1}^n (\text{length of char}_i) * \text{freq} \leftarrow \text{probability}$$

$$= 3 \times .10 + 3 \times .12 + \dots + 4 \times .35 = 2.64 \text{ bits/char}$$

Non technical English \rightarrow 4.5 bits/char

$$\text{Compression Rate (w.r.t ASCII)} = \frac{8 - 4.5}{8} = \frac{3.5}{8} \approx 0.44$$

Horizontal Partitioning

Signature File Partitioning

Partitioned Signature Files = Design Issues

D.L. Lee, C.W. Leng, D.H. Tu

S₁: 0111 1000

S₂: 1000 1011

S₃: 0011 1100

S₄: 1100 0011

S₅: 0110 1100

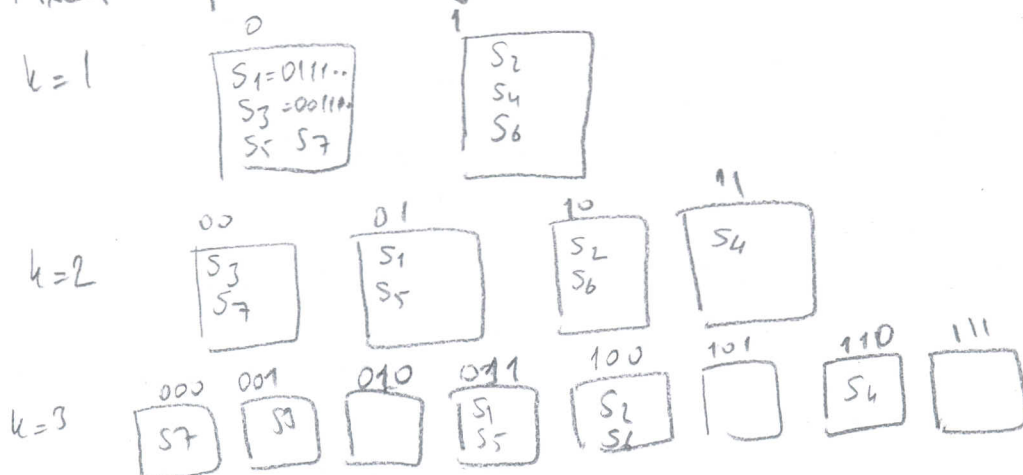
S₆: 1001 0011

S₇: 0000 1111

Assumes that signatures are random bits.

50% 50%
0 1

Fixed Prefix Partitioning



Example: Q₁ = 1110 0001

Q₂ = 0000 1111

Q₃ = 0110 0011

Partition Activation Ratio = $\frac{\text{\# of activated partitions}}{\text{total \# of partitions}}$

of activated signatures = $2^{\text{\# of 0's in query signature}}$

Signature Activation Ratio = $\frac{\text{\# of activated signatures}}{\text{total \# of signatures}}$

Query	k=1	k=2	k=3	PAL(k=1) SAR	PAL(k=2)	PAL(k=3)
Q ₁	1(1)	1(11)	1(111)	$\frac{1}{2}$ $\frac{3}{7}$	$\frac{1}{6}$ $\frac{1}{7}$	$\frac{1}{8}$ $\frac{0}{7}$
Q ₂	2(0,1)	4(00,01 10,11)	8(All)	$\frac{2}{2}$	$\frac{4}{6}$	$\frac{8}{8}$
Q ₃	2(0,1)	2(01,11)	2(011, 111)	$\frac{2}{2}$	$\frac{2}{6}$	$\frac{2}{8}$