

# Communicating Efficiently: A Computer Science Perspective

Nithin Varma  
Chennai Mathematical Institute

“the act or process of using words, sounds, signs, or behaviours to express or exchange information or to express your ideas, thoughts, feelings, etc., to someone else”

*–Merriam Webster*



# Ways to communicate

# Ways to communicate



# Ways to communicate



# Ways to communicate



# Ways to communicate



# Ways to communicate





# Brevity in communication

- Has always been considered desirable

# Brevity in communication

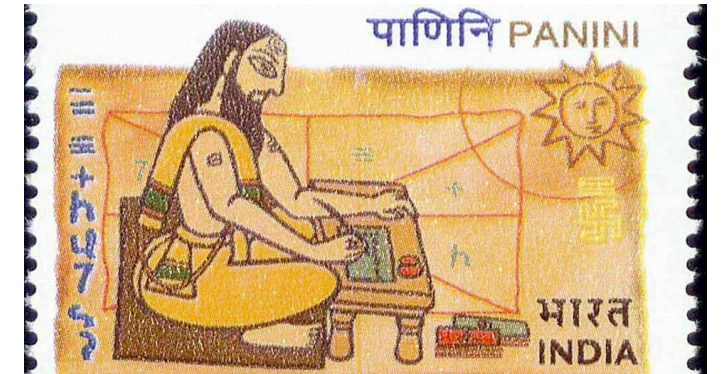
- Has always been considered desirable
  - In pre-writing ages, brevity in texts made memorisation easier

# Brevity in communication

- Has always been considered desirable
  - In pre-writing ages, brevity in texts made memorisation easier
  - Brevity in writing resulted in smaller books — easier to carry, easier to copy from, takes up less space ....

# Brevity in communication

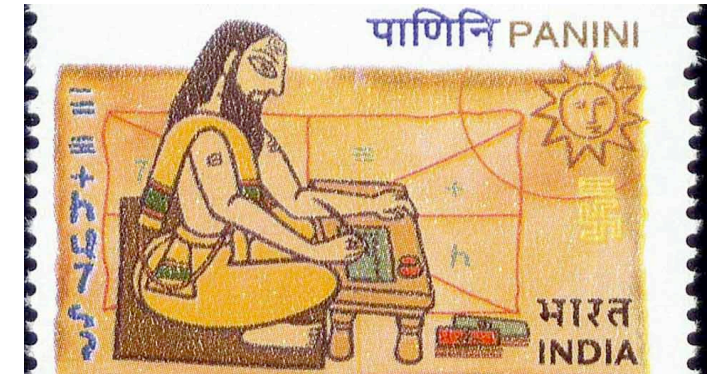
- Has always been considered desirable
  - In pre-writing ages, brevity in texts made memorisation easier
  - Brevity in writing resulted in smaller books — easier to carry, easier to copy from, takes up less space ....



*Sanskrit Grammarian*

# Brevity in communication

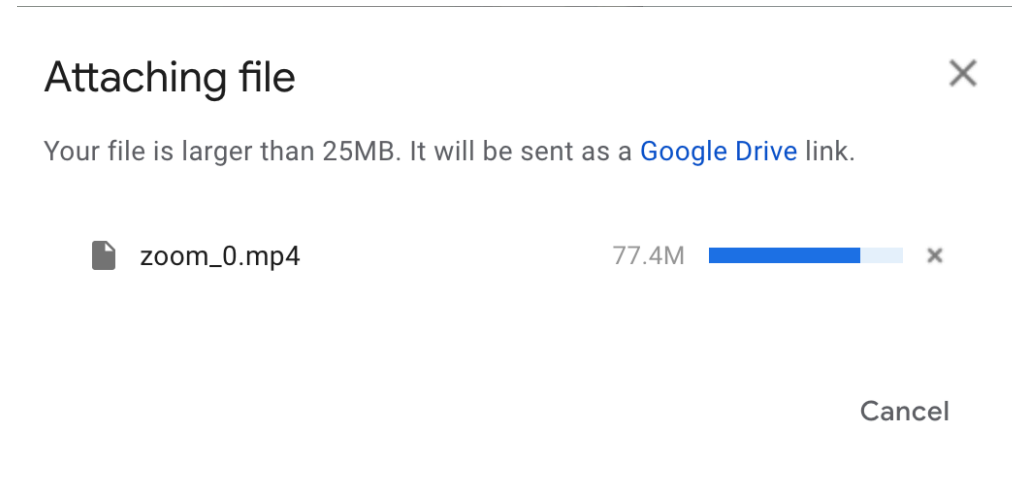
- Has always been considered desirable
  - In pre-writing ages, brevity in texts made memorisation easier
  - Brevity in writing resulted in smaller books — easier to carry, easier to copy from, takes up less space ....
- Is brevity or succinctness still desirable in the modern age ?



Is succinctness still important?

# Is succinctness still important?

- While sending an email...

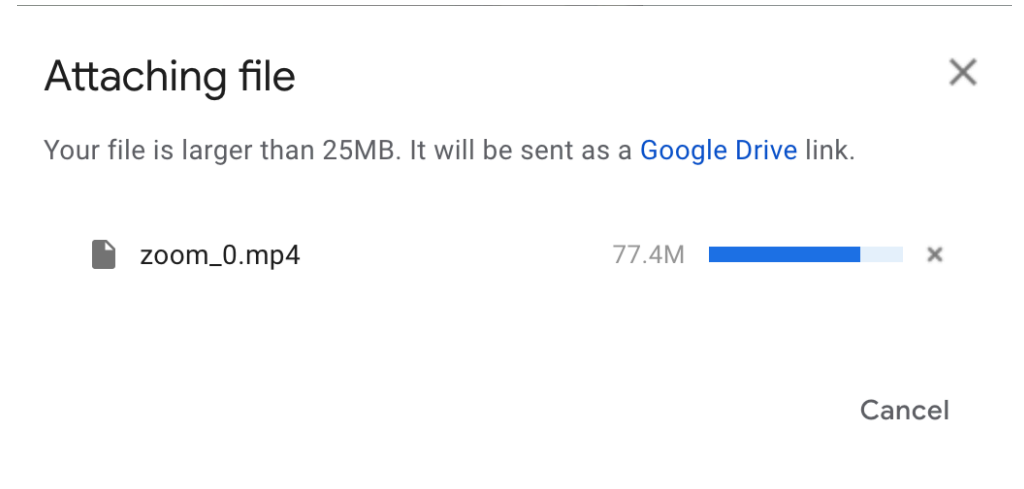


---

Sending or receiving large files is a problem

# Is succinctness still important?

- While sending an email...
- Downloading or watching a movie with limited data budget
- .....



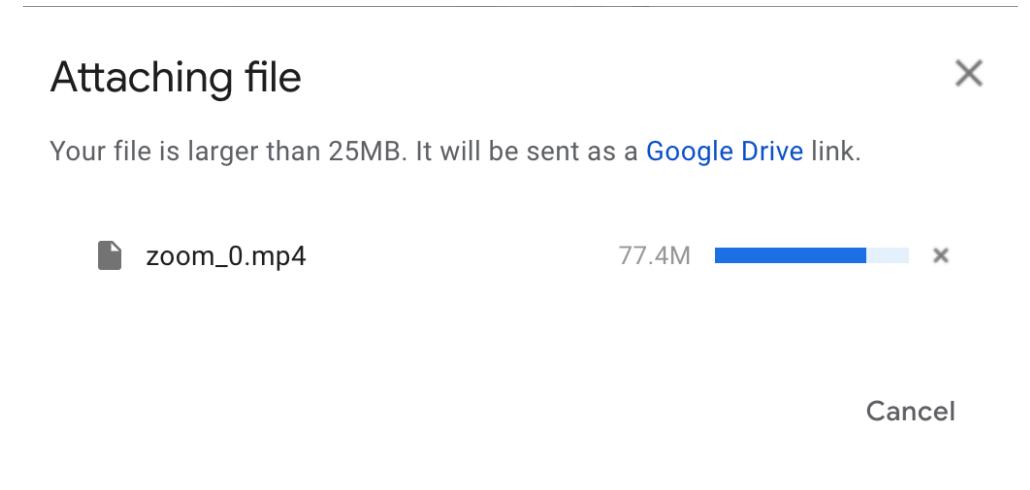
---

Sending or receiving large files is a problem



# Is succinctness still important?

- While sending an email...
- Downloading or watching a movie with limited data budget
- .....



Sending or receiving large files is a problem

**Why are such limits imposed?**

Communication has some cost..

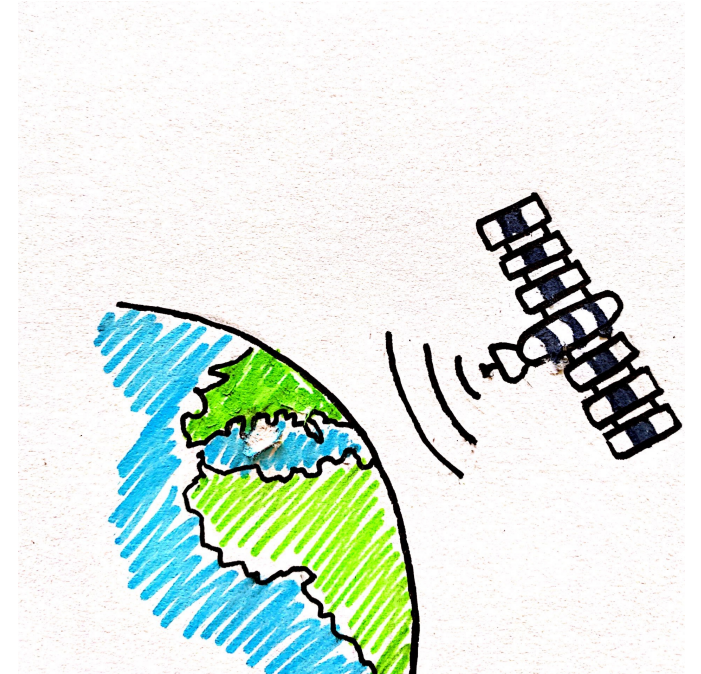
# Communication has some cost..

- Infrastructure installation and maintenance

# Communication has some cost..

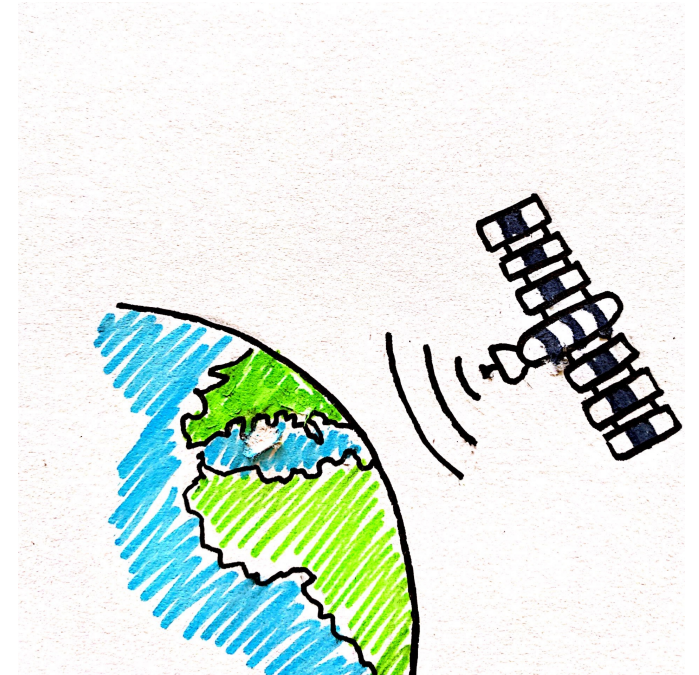
- Infrastructure installation and maintenance
- Energy costs
- ....

# Larger scale communication



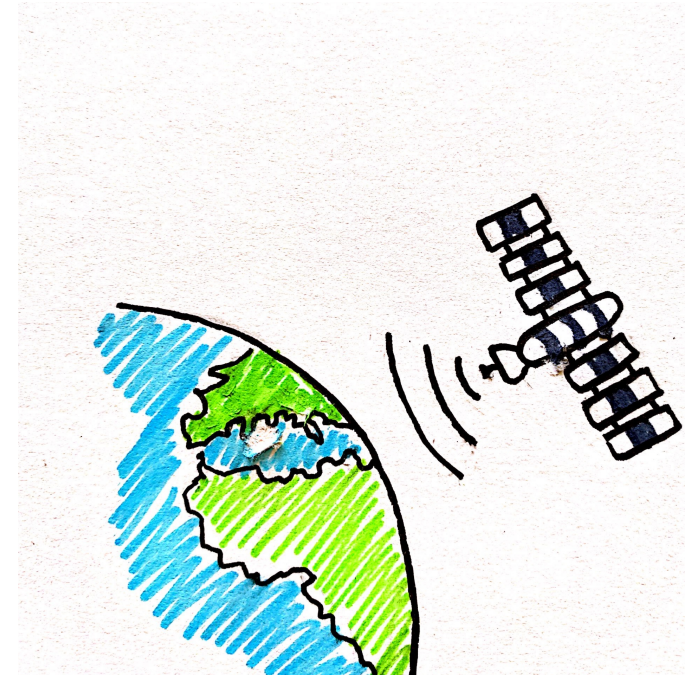
# Larger scale communication

- Satellites send large amounts of data daily to the ground stations



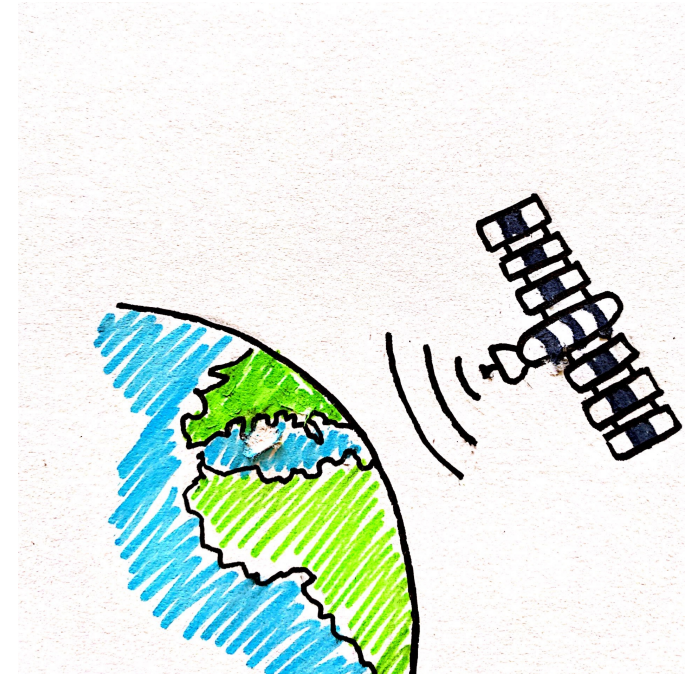
# Larger scale communication

- Satellites send large amounts of data daily to the ground stations
- Images of earth surface
  - Useful for understanding weather, climate, agriculture, urbanisation, ....



# Larger scale communication

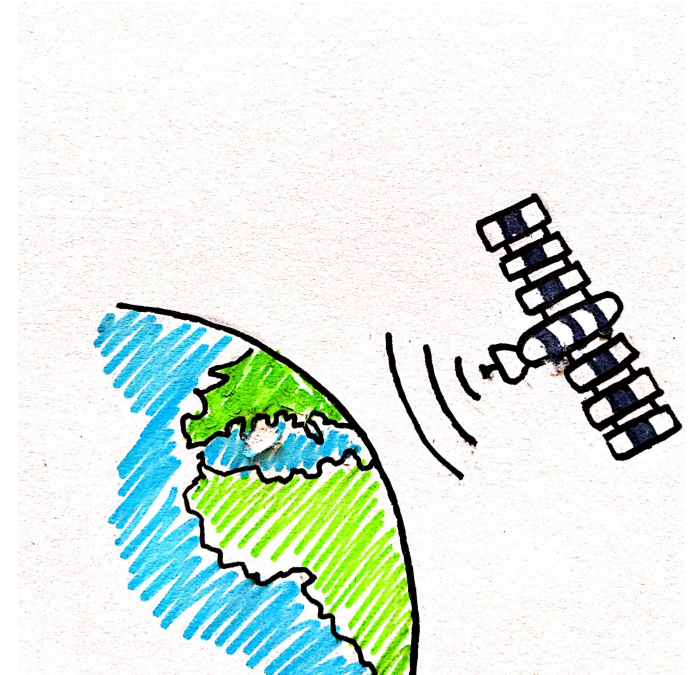
- Satellites send large amounts of data daily to the ground stations
- Images of earth surface
  - Useful for understanding weather, climate, agriculture, urbanisation, ....
- 100 Tera Bytes or more per day





# Larger scale communication

- Satellites send large amounts of data daily to the ground stations
  - Images of earth surface
    - Useful for understanding weather, climate, agriculture, urbanisation, ....
  - 100 Tera Bytes or more per day
    - 47200 hours of video
    - 13000000 digital photos



What can we do to address this problem?

# Transporting chairs

# Transporting chairs



# Transporting chairs



# Transporting chairs



Many more chairs can be transported with the same cost when chairs are folded!

Can we do the same with information being  
communicated?

# Can we do the same with information being communicated?

- Goal: Compress information before communicating



# Can we do the same with information being communicated?

- Goal: Compress information before communicating
- Today: How to compress text messages ?

# Text compression: Simple scheme



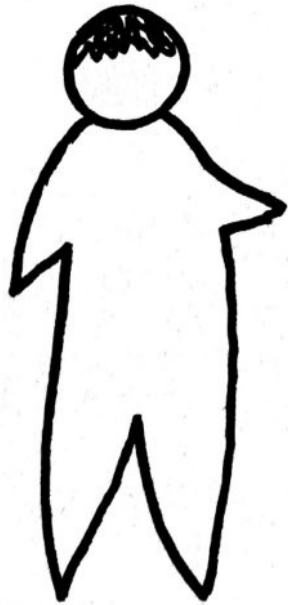
# Text compression: Simple scheme



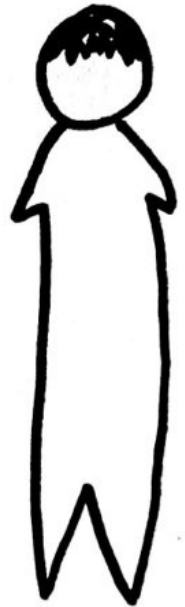
Hppy Brthdy nd Mrry Chrstms!



# Text compression: Simple scheme



Hppy Brthdy nd Mrry Chrstms!



*Frittatas*     *intimidating*     *easy*

Frttts r ntmdtng t frst, bt thy r sy t mk  
nc y knw wht y r dng

*once*

# Text compression: Features

# Text compression: Features

- Want to get considerable reduction in length

# Text compression: Features

- Want to get considerable reduction in length
- Should be able to decode the original message correctly and easily from the compressed text

# Text compression: Features

- Want to get considerable reduction in length
- Should be able to decode the original message correctly and easily from the compressed text

**Rest of the talk: Discuss a good scheme used to compress text messages**



# A simplification

# A simplification

- Assume that you only want to send text messages written using A, S, D, F, G, H

# A simplification

- Assume that you only want to send text messages written using A, S, D, F, G, H
- A message could look like

ASSASFSHGHGHGHDGFGS

# A simplification

- Assume that you only want to send text messages written using A, S, D, F, G, H
- A message could look like

ASSASFSHGHGHGHDGFGS

- Need to also encode messages in binary (0s and 1s)

# Binary encoding

- How many bits are needed to encode A, S, D, F, G, H?

A  $\longleftrightarrow$  000  
S  $\longleftrightarrow$  001  
D  $\longleftrightarrow$  010  
F  $\longleftrightarrow$  100  
G  $\longleftrightarrow$  101  
H  $\longleftrightarrow$  110

ASFDSS F

000 001 100 010 001 001 100

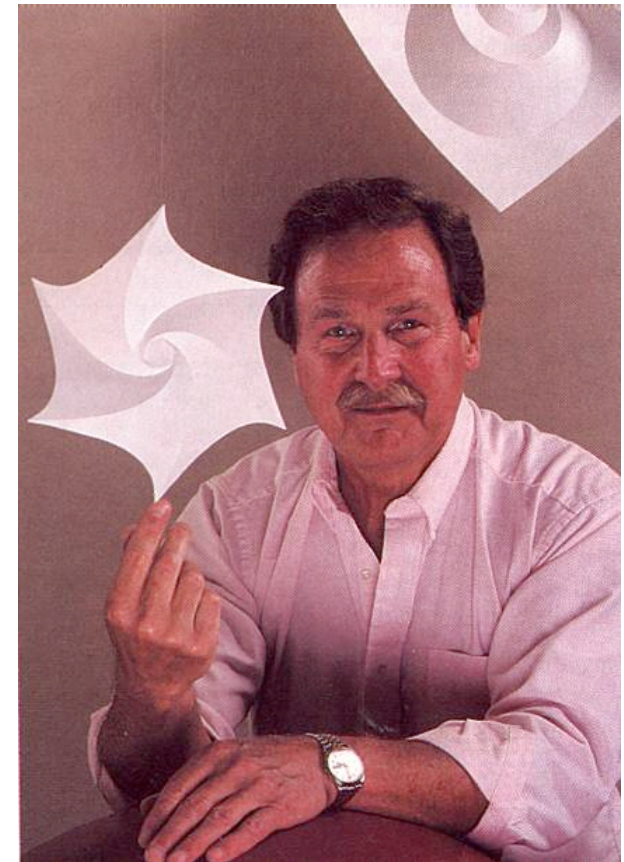
# Binary encoding

- If you have a text of  $k$  characters, you use \_\_\_\_\_ bits
- Can we do better ?

- Yes, we can do better!

# Huffman's compression scheme

- Yes, we can do better!
- Scheme developed by David Huffman in 1952, when he was still a student





# Huffman's compression: Idea 1

- Do all letters appear with the same frequency ?

-

# Huffman's compression: Idea 1

- Do all letters appear with the same frequency ?
  - No. Some letters occur more frequently than others.
  -

# Huffman's compression: Idea 1

- Do all letters appear with the same frequency ?
  - No. Some letters occur more frequently than others.
  - E, A, T, S, N, H, I, O occur much more frequently than others in the English language

# Huffman's compression: Idea 1

- Do all letters appear with the same frequency ?
  - No. Some letters occur more frequently than others.
  - E, A, T, S, N, H, I, O occur much more frequently than others in the English language
- Idea: Use fewer bits to encode characters that appear more frequently

# Frequency table

- The frequencies of these characters are as shown in the table

ASSASFSHGHGHGHDGFGS

1  
01

0101

DD

GSD

0

DGS

GSGS

A	10%	2
S	25%	5
D	5%	1
F	10%	2
G	30%	6
H	20%	4

# Frequency table

- The frequencies of these characters are as shown in the table

ASSASFSHGHGHGHDGFGS

- Can we come up with an encoding scheme that takes these frequencies into account?

A	10%
S	25%
D	5%
F	10%
G	30%
H	20%

Your scheme

# Huffman compression

A 10% →

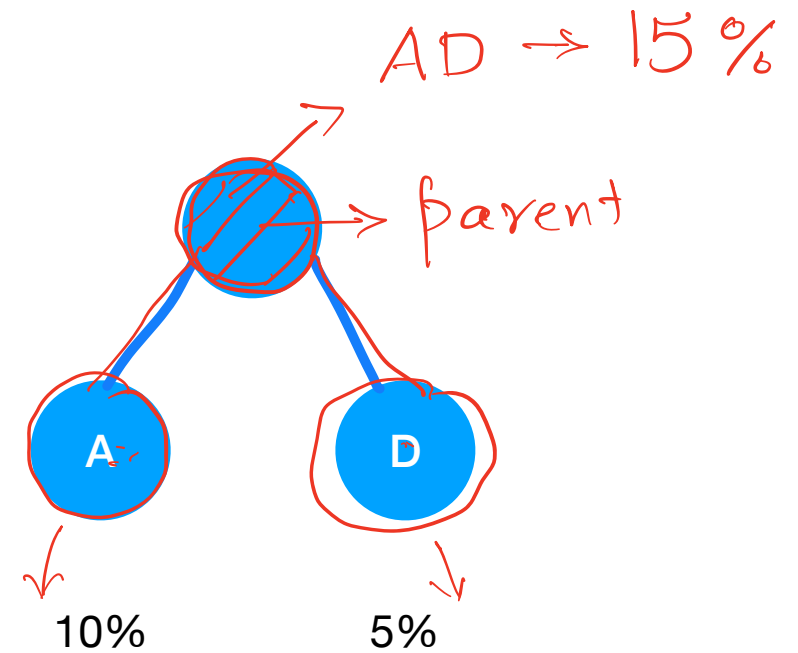
S 25%

D 5% →

F 10%

G 30%

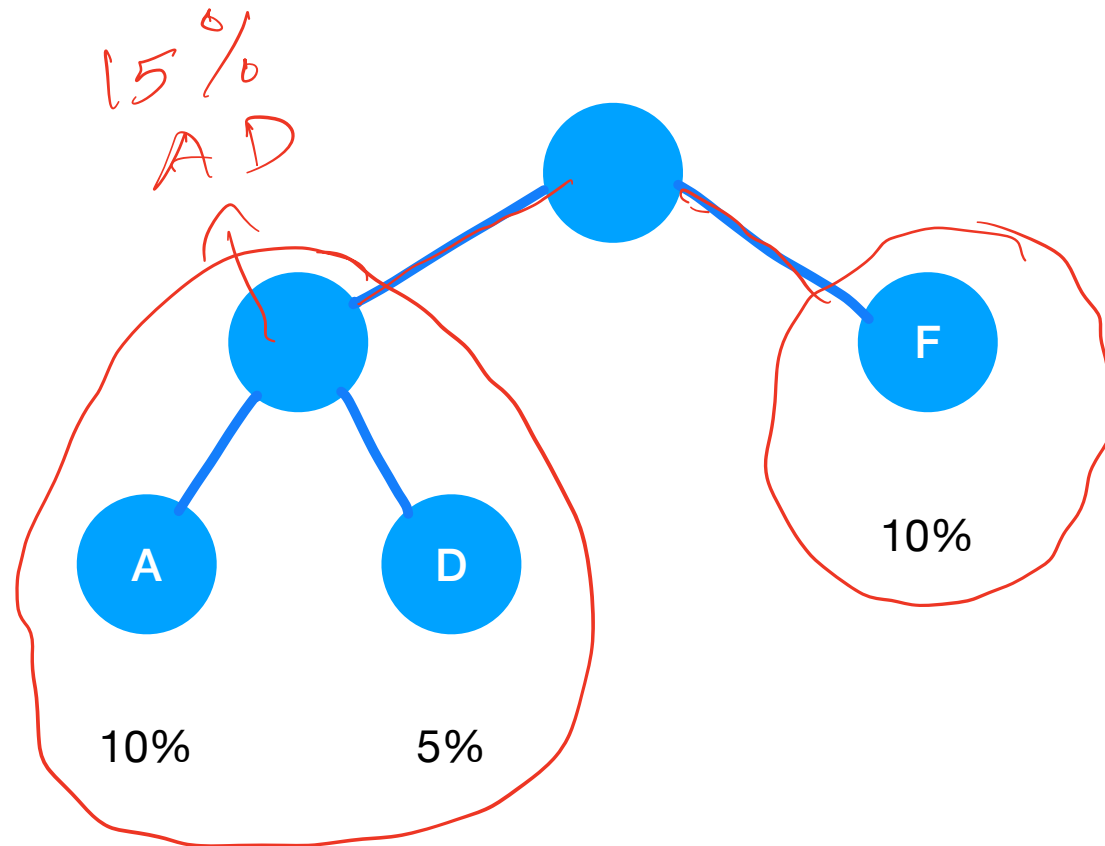
H 20%





# Huffman compression

AD	15%
S	25%
F	10%
G	30%
H	20%



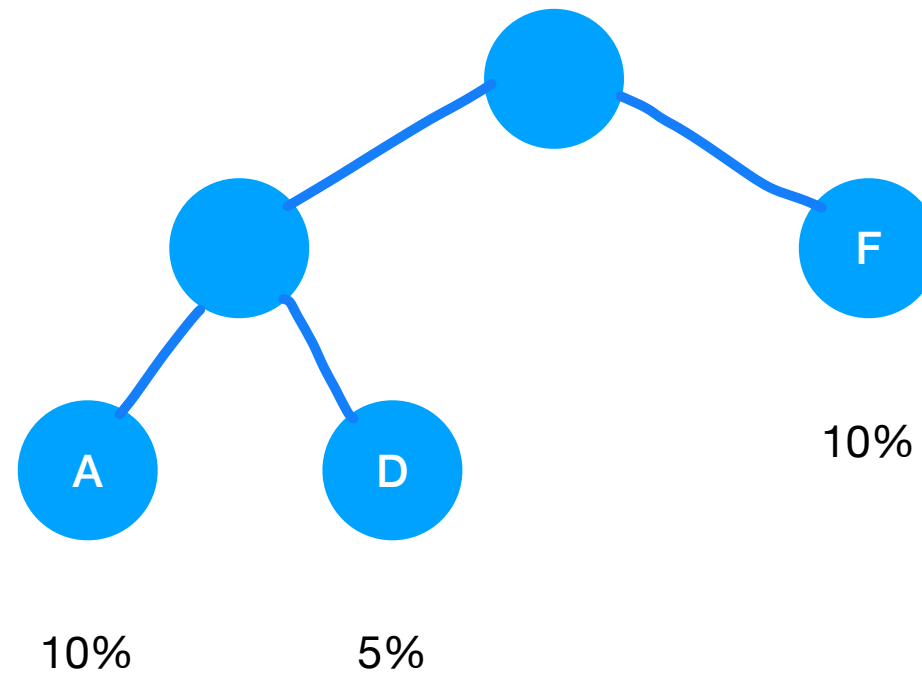
# Huffman compression

ADF 25%

S 25%

G 30%

H 20%



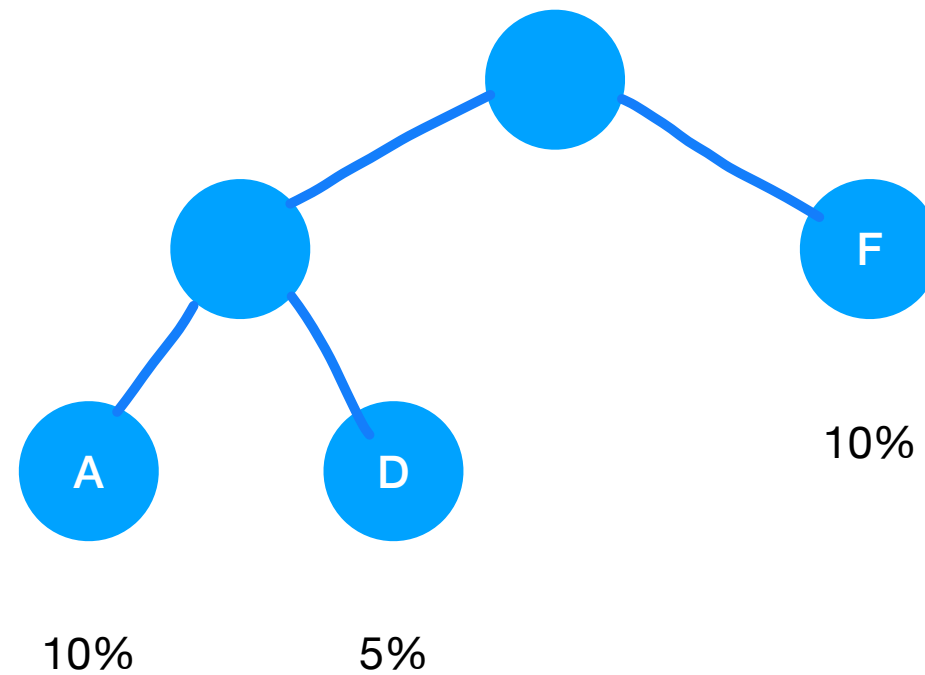
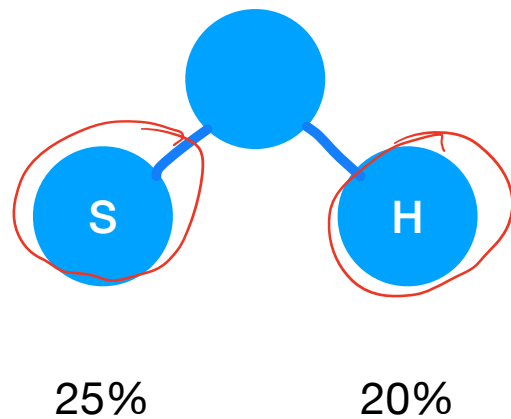
# Huffman compression

ADF 25%

S 25%

G 30%

H 20%

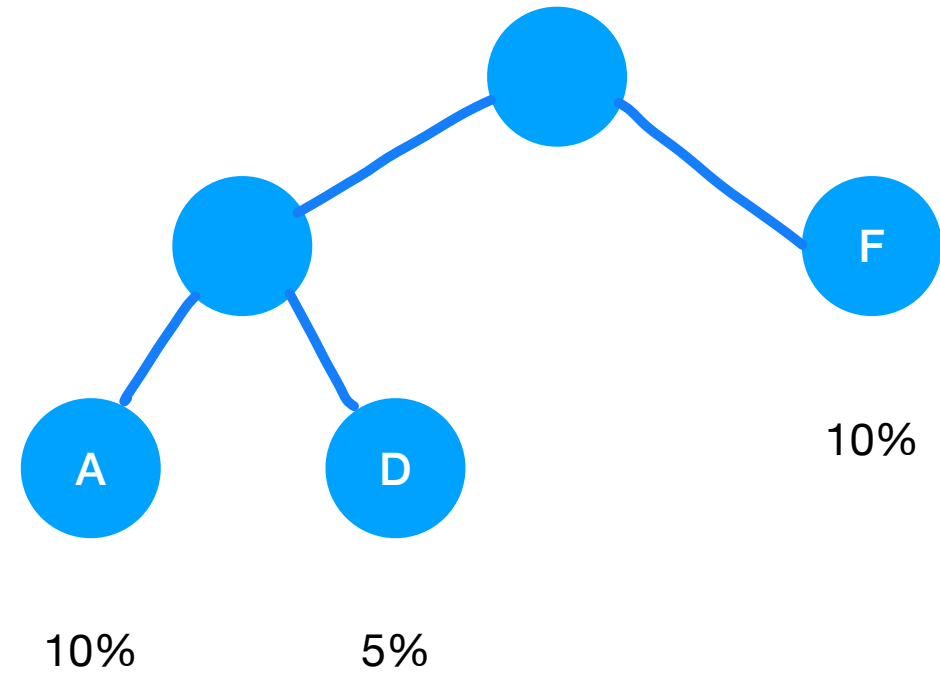
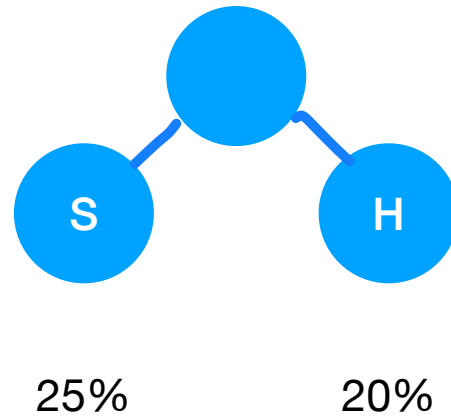


# Huffman compression

ADF 25%

SH 45%

G 30%

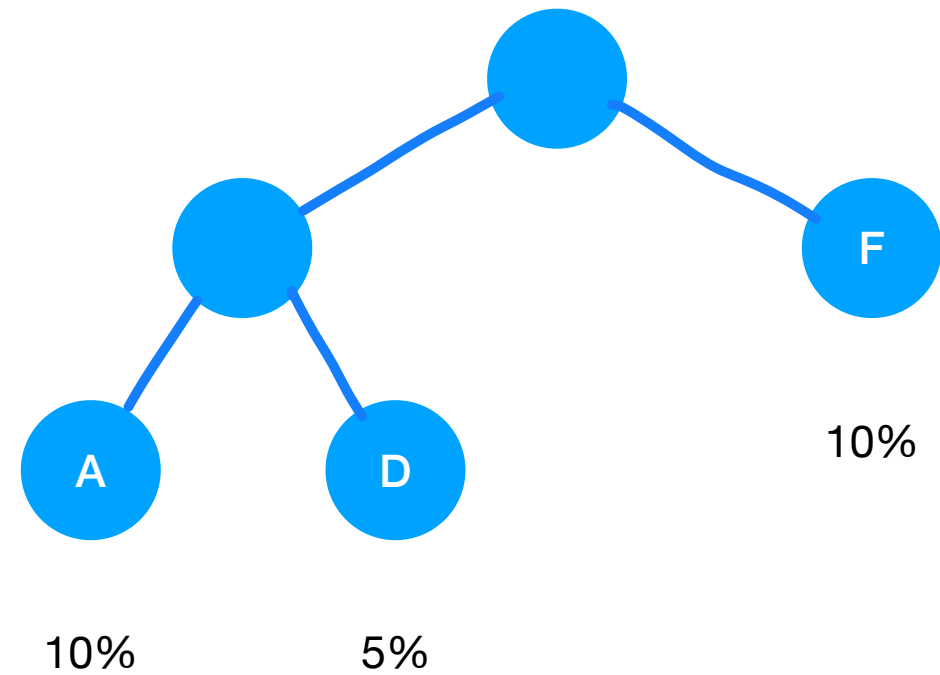
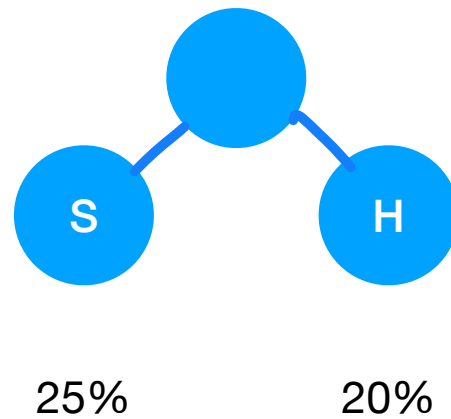


# Huffman compression

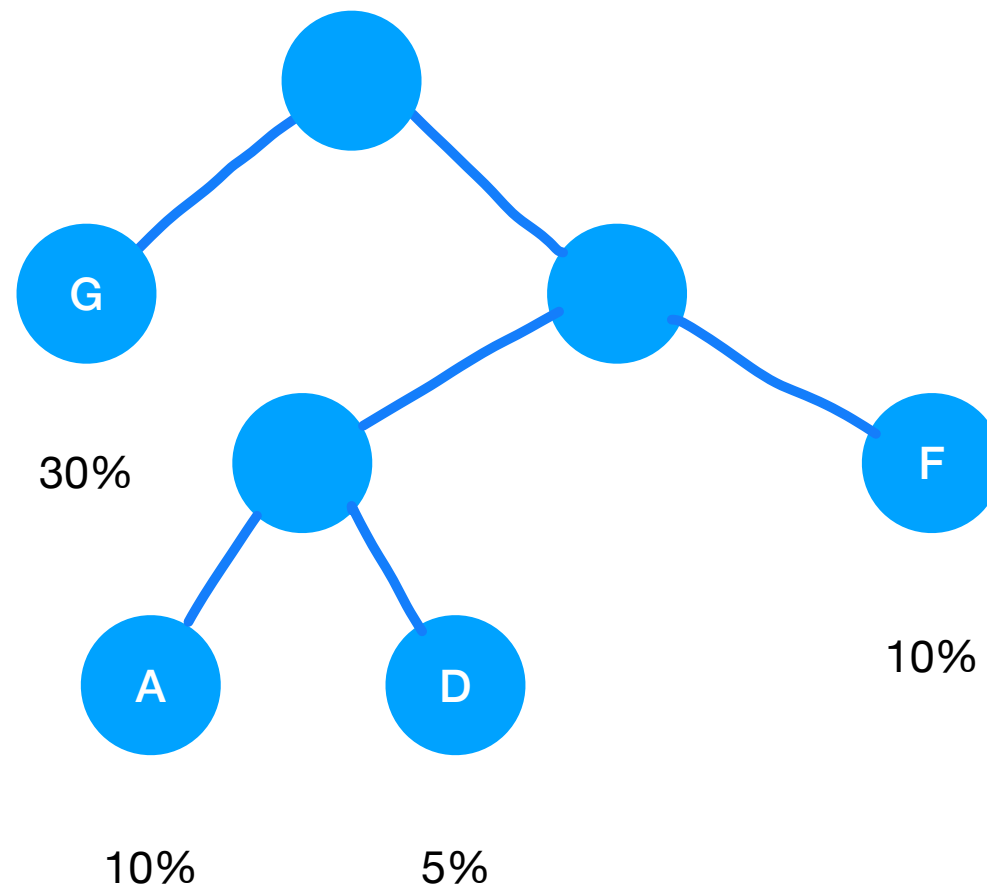
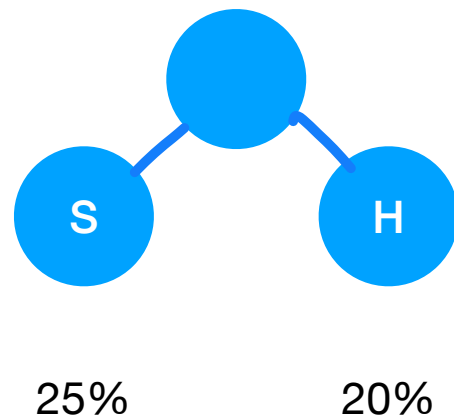
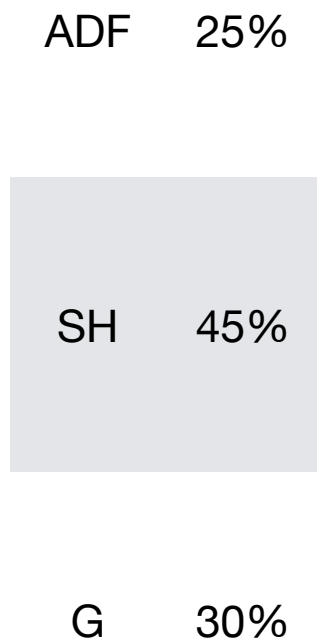
ADF 25%

SH 45%

G 30%



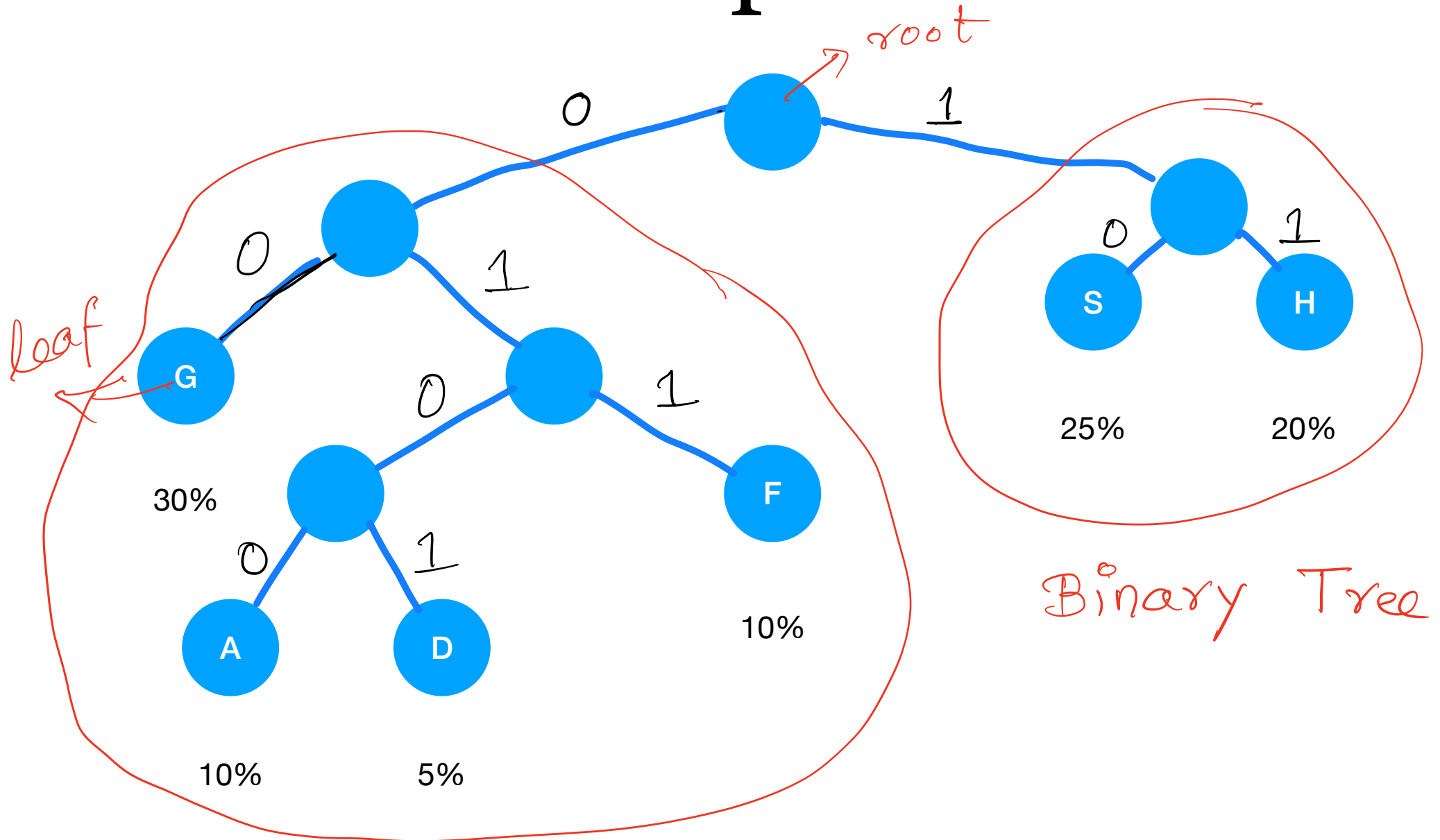
# Huffman compression



# Huffman compression

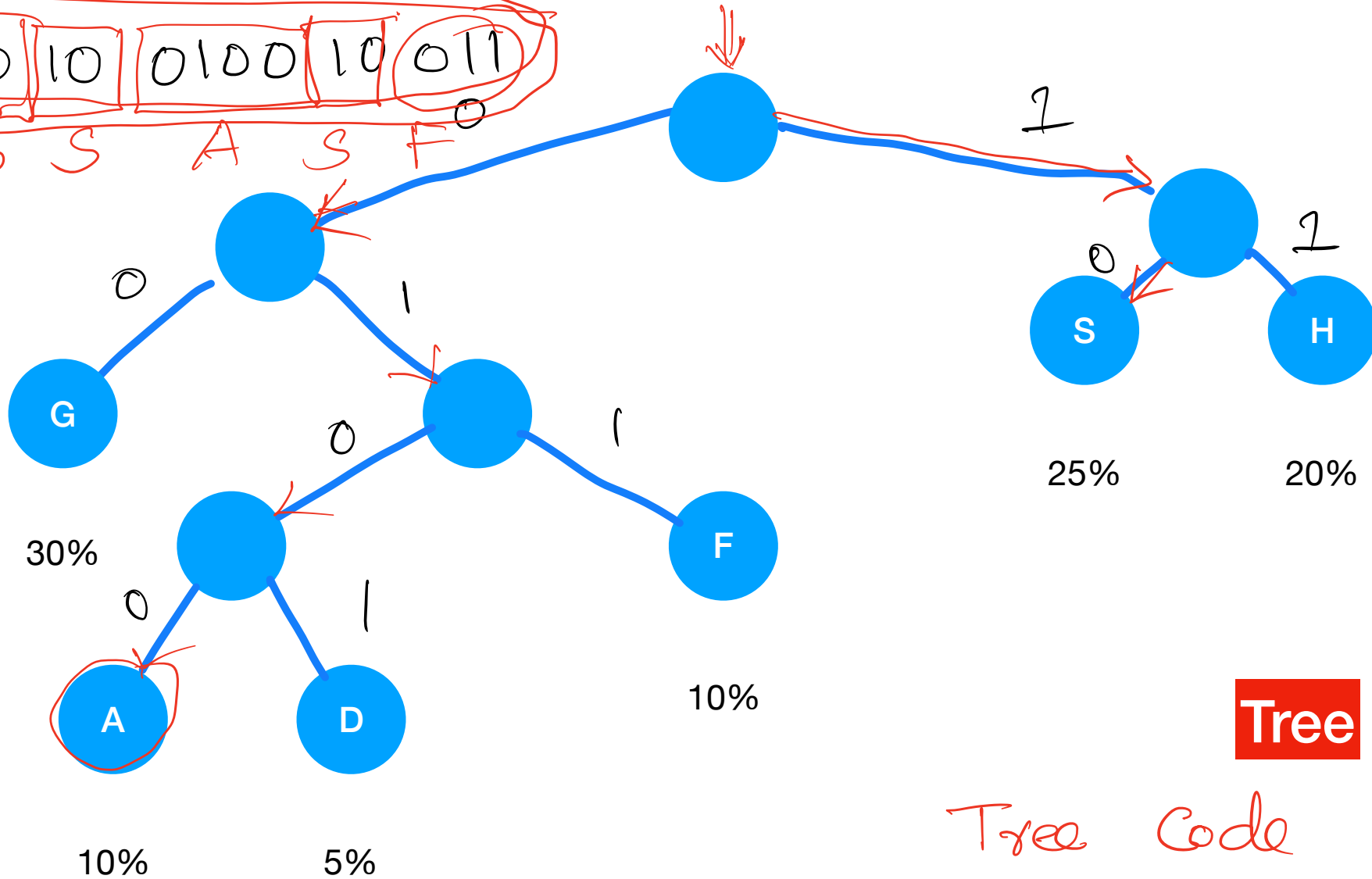
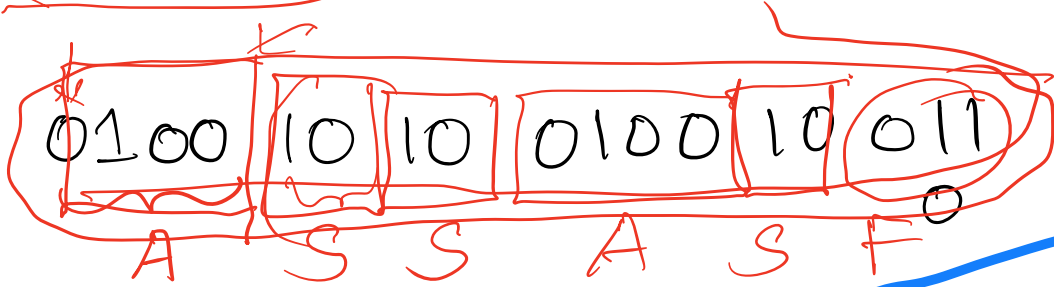
ADFG 55%

SH 45%



ASSASF

# Huffman compression



Tree

Tree Code

n ↓



# Huffman compression

Prefix-free Code

	A	2	10%	0100
	S	5	25%	10
→	D	1	5%	0101
	F	2	10%	011
→	G	6	30%	00
	H	4	20%	11

So, what is the encoding of

ASSASFSHGHGHGHGDGFGS

20 characters

• encoding each with 3 bits :

60 bits

• Huffman encoding :

$$8 + 10 + 4 + 6 + 12 + 8 = 48 \text{ bits}$$

# Final remarks..

**Huffman's scheme is one of the best compression methods known!**

Questions?