

Ψηφιακές Τηλεπικοινωνίες

- Προθεματικοί κώδικες
- Αλγόριθμος Huffman

Προθεματικοί κώδικες

- Αλγόριθμοι κωδικοποίησης (συμπίεσης) πηγής
- Στόχος: Να επιτυγχάνουν ρυθμούς κωδικοποίησης κοντά στην εντροπία (στο όριο συμπίεσης χωρίς απώλειες)
- Κωδικοποίηση από σταθερό σε μεταβλητό μήκος:
 - είσοδος: μπλοκ συμβόλων σταθερού μήκους
 - έξοδος: μπλοκ bits μεταβλητού μήκους (κωδική λέξη)
- **Πρόβλημα:**

Συγχρονισμός: πώς μπορώ να βρω τα όρια των μπλοκ στην έξοδο για να γίνει η αποκωδικοποίηση
- **Λύση:**

Προθεματικός κώδικας: Καμία κωδική λέξη δεν αποτελεί πρόθεμα κάποιας άλλης λέξης

 - » μοναδικά αποκωδικοποιήσιμος (κάθε έξοδος αντιστοιχεί σε μοναδική είσοδο)
 - » άμεσος (επιτρέπει απευθείας αποκωδικοποίηση)

Παραδείγματα κωδίκων

Source Symbol	Probability	Code 1	Code 2	Code 3
s_0	0.5	0	0	0
s_1	0.25	1	01	10
s_2	0.125	10	011	110
s_3	0.125	11	0111	111
Uniquely decodable				
Prefix				

Παραδείγματα κωδίκων

Source Symbol	Probability	Code 1	Code 2	Code 3
s_0	0.5	0	0	0
s_1	0.25	1	01	10
s_2	0.125	10	011	110
s_3	0.125	11	0111	111
Uniquely decodable		NO	YES	YES
Prefix		NO	NO	YES

Μέσο Μήκος Κώδικα

- Έστω DMS με πιθανότητες εμφάνισης $p(s_i)$ και ένας κωδικοποιητής πηγής που αναθέτει $l(s_i)$ bits στο σύμβολο s_i

Εάν είναι προθεματικός κώδικας έχει τις εξής ιδιότητες:

- Ισχύει η Kraft-McMillan inequality:

$$\sum_{i=1}^N 2^{-l(s_i)} \leq 1$$

- Ικανή συνθήκη για την ύπαρξη προθεματικού κώδικα
- Αναγκαία συνθήκη για την ύπαρξη προθεματικού κώδικα (και γενικά για μοναδικά αποκωδ/μο κώδικα)

- Ισχύουν φράγματα για το μέσο μήκος. Συγκεκριμένα:

Για κάθε πηγή DMS μπορεί να κατασκευαστεί προθεματικός για τον οποίο:

$$H(X) \leq \bar{L} < H(X) + 1$$

Αποδοτικότητα Κώδικα

- Η αποδοτικότητα ενός κώδικα ορίζεται ως

$$\eta = \frac{H(X)}{\bar{L}} \leq 1$$

– και δείχνει πόσο κοντά βρίσκεται ο κωδικοποιητής στο όριο συμπίεσης της πηγής (εντροπία)

- Ένας κώδικας είναι αποδοτικός, όσο το η ...
- ... πλησιάζει στο 1

N-οστής Τάξης Επέκταση Πηγής

- Στην περίπτωση προθεματικής κωδικοποίησης μπορούν να δημιουργηθούν **επεκτεταμένες πηγές**. Δηλαδή:
- ως επεκτεταμένη είσοδος είναι ένα μπλοκ από σύμβολα
- που κωδικοποιούνται ως ένα σύνθετο σύμβολο, δηλαδή
 - έστω s_i και s_j
 - θέτει $\sigma_k = (s_i, s_j)$
 - με πιθανότητα εμφάνισης $p(\sigma_k) = p(s_i)p(s_j)$ για πηγή χωρίς μνήμη
- Για **n -οστή επέκταση** πηγής ισχύει η ανισότητα:

$$H(X^n) \leq \bar{L}_n < H(X^n) + 1$$

- Για επεκταμένη πηγή χωρίς μνήμη αποδεικνύεται ότι:

$$H(X^n) = nH(X)$$

N-οστής Τάξης Επέκταση Πηγής (2)

- Προκύπτει

$$H(X) \leq \frac{\overline{L}_n}{n} < H(X) + \frac{1}{n}$$

- **Συμπέρασμα:** Η n -οστής τάξης επέκταση μιας πηγής αποφέρει κώδικες που είναι ολοένα και πιο κοντά στο όριο συμπίεσης (εντροπία) της πηγής

$$\lim_{n \rightarrow \infty} \frac{1}{n} \overline{L}_n = H(X)$$

- **Ερώτηση:** Γιατί δε χρησιμοποιώ ένα πολύ μεγάλο n , ώστε να πετύχω συμπίεση κοντά στο όριο της εντροπίας;

Βήματα Αλγορίθμου Huffman (1)

- Δημιουργία Δυαδικού Δέντρου:

1. Διάταξε τις εισόδους κατά φθίνουσα σειρά πιθανοτήτων
2. Συγχώνευσε τα δύο σύμβολα με τις μικρότερες πιθανότητες και δημιούργησε νέο «σύμβολο», το οποίο έχει πιθανότητα:

Το άθροισμα των επιμέρους πιθανοτήτων

3. Ανάθεσε στα δύο συνιστώντα σύμβολα «0» και «1»
4. Ταξινόμησε εκ νέου τη λίστα των συμβόλων
5. Επανάλαβε τα παραπάνω μέχρις όλα τα σύμβολα συγχωνευτούν σε ένα τελικό σύμβολο

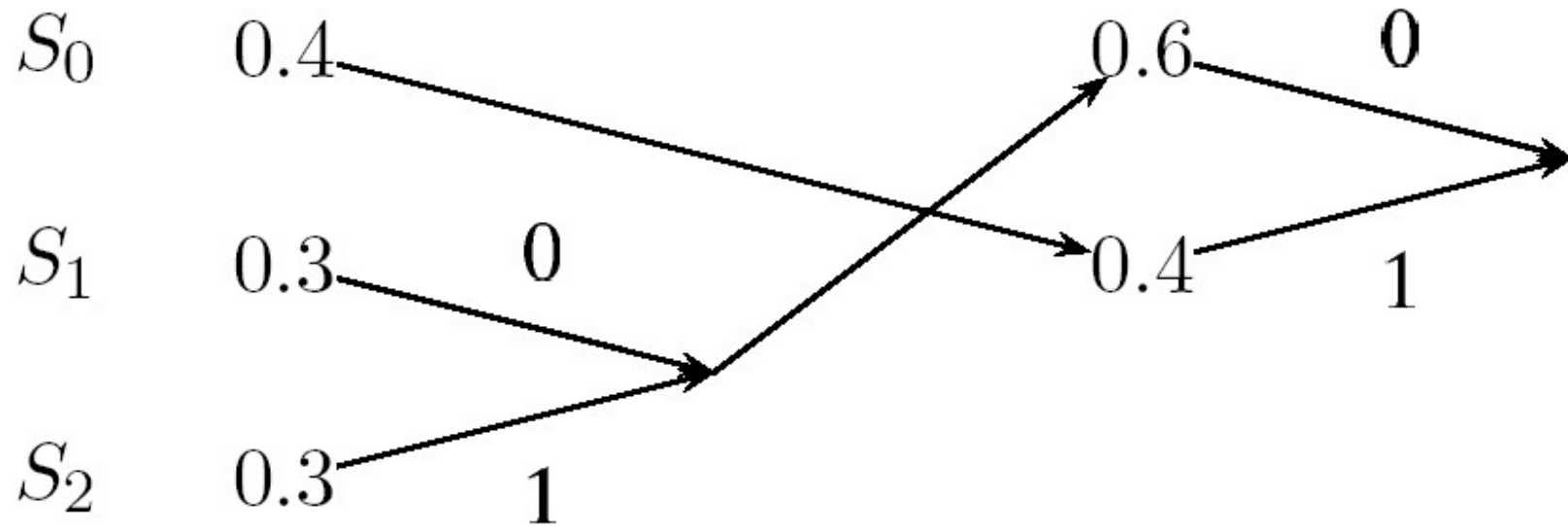
- Δημιουργήθηκε ένα δυαδικό δέντρο:

- ρίζα: το τελικό σύνθετο σύμβολο
- φύλλα: τα αρχικά σύμβολα
- ενδιάμεσοι κόμβοι: σύνθετα σύμβολα

Βήματα Αλγορίθμου (2)

- Ανάθεση Bits σε Σύμβολα Εισόδου
 1. Ξεκίνα από τη ρίζα και κινήσου προς ένα φύλλο
 2. Η ακολουθία των bits που συναντώνται είναι η ακολουθία κωδικοποίησης
 3. Επανάλαβε για όλα τα σύμβολα (φύλλα)

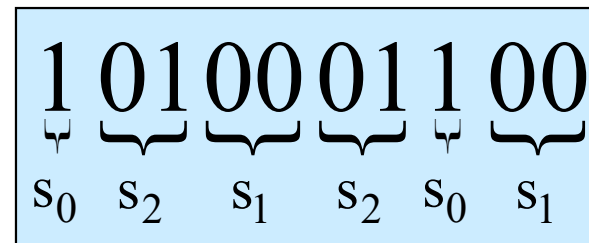
Παράδειγμα Huffman



- Προθεματική αντιστοίχιση:

- s_0 : 1
- s_1 : 00
- s_2 : 01

- Μονοσήμαντη και άμεση αποκωδικοποίηση



Χαρακτηριστικά Huffman

■ Μειονέκτημα:

- Απαιτεί να γνωρίζει εκ των προτέρων τις πιθανότητες εμφάνισης των συμβόλων της πηγής. Συνεπώς δεν είναι κατάλληλος για:
εφαρμογές πραγματικού χρόνου

■ Βέλτιστος:

- Για πηγές DMS, ανάμεσα σε όλους τους **symbol-by-symbol** προθεματικούς κώδικες πετυχαίνει το ελάχιστο μέσο μήκος κώδικα, όταν έχουμε:
symbol-by-symbol coding, zero-memory και γνωστή pdf

■ Συμβάσεις:

- Ο τρόπος ανάθεσης 0 και 1 (δεν παίζει ρόλο)
- Η τοποθέτηση στην ταξινομημένη λίστα σε περίπτωση «ισοβαθμίας» σύνθετου συμβόλου με άλλο σύμβολο (σχετίζεται με τη διασπορά του κώδικα)

Ένα ακόμη παράδειγμα

