



Κώδικες Huffman

- τεχνική συμπίεσης διδομένων (20% - 90% κέρδος)
- ακολουθία αλφαριθμητικών χαρακτήρων
- αναπαράσταση κάθε χαρακτήρα υπό τη μορφή δυαδικής συμβολοσειράς
 - βάσει των συχνοτήτων εμφάνισης των χαρακτήρων
- βέλτιστος τρόπος αναπαράστασης

Κώδικες Huffman

<i>Χαρακτήρες</i>	a	b	c	d	e	f
Συχνότητα (σε χιλιάδες)	45	13	12	16	9	5
Κωδικός σταθερού μήκους	000	001	010	011	100	101
Κωδικός μεταβλητού μήκους	0	101	100	111	1101	1100

Κώδικες Huffman

Χαρακτήρες	a	b	c	d	e	f
Συχνότητα (σε χιλιάδες)	45	13	12	16	9	5
Κωδικός σταθερού μήκους	000	001	010	011	100	101
Κωδικός μεταβλητού μήκους	0	101	100	111	1101	1100
Αρχείο: 100 000 χαρακτήρες	300000 bits		/	224000 25%		

Κωδικοποίηση

- abc

- 0 . 101 . 1101 (concatenation)

- Δηλαδή 01011101

- Εύκολη αποκωδικοποίηση

- Ουδεμία ασάφεια

Αποκωδικοποίηση

- 001011101

- 0 . 0 . 101 . 1101

- Δηλαδή aabe

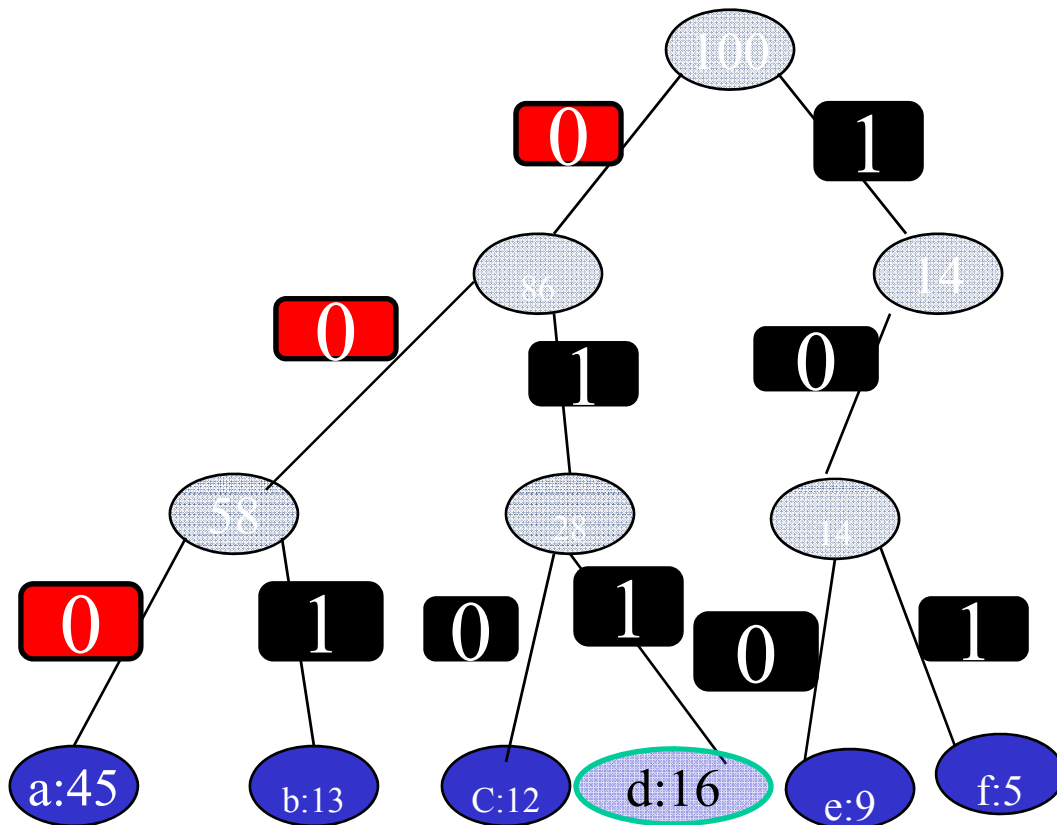
- αναπαράσταση μέσω ενός δυαδικού δέντρου

- Φύλλα αντιστοιχούν στους δεδομένους χαρακτήρες

Αποκωδικοποίηση / δυαδικό δέντρο

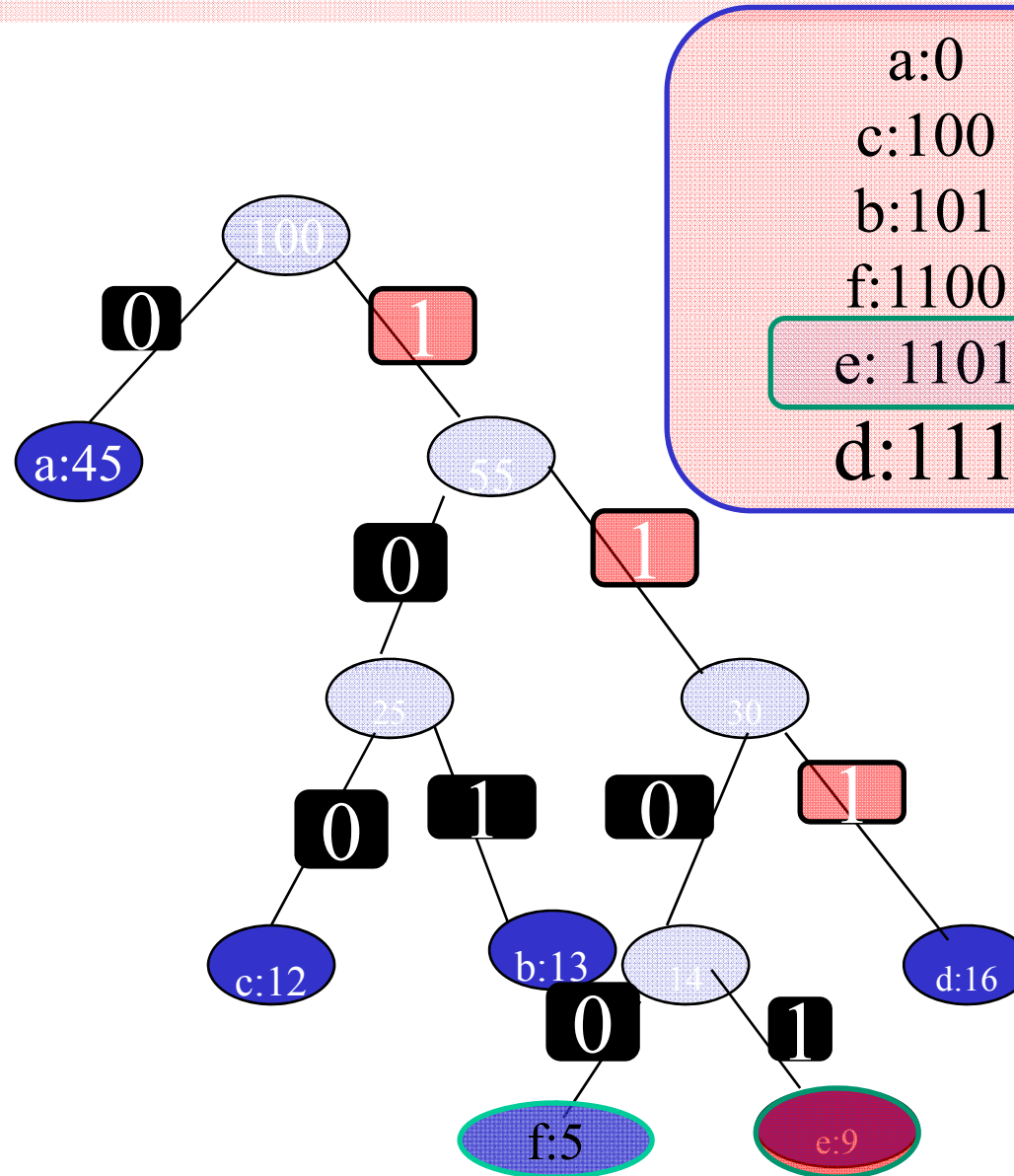
- ο δυαδικός κωδικός ενός χαρακτήρα συνίσταται στη διαδρομή από τη ρίζα του δέντρου μέχρι τον συγκεκριμένο χαρακτήρα
- «0» σημαίνει «μετάβαση στον αριστερό θυγατρικό κόμβο»
- «1» σημαίνει «μετάβαση στον δεξιό θυγατρικό κόμβο»

Αναπαράσταση Δυαδικό δέντρο (σταθερού μήκους κώδικες)



a: 000
c: 010
b: 001
f: 101
e: 100
d: 011

Αναπαράσταση Δυαδικό δέντρο (μεταβλητού μήκους κώδικες)



Αλγόριθμος Huffman

HUFFMAN(C)

$n=|C|$ /* πλήθος χαρακτήρων */

$Q=C$ /* Σωρός */

for $i=1$ to $n-1$ do

allocate a new node z

$x=\text{extract-minimum}(Q)$, $\text{left}[z]=x$

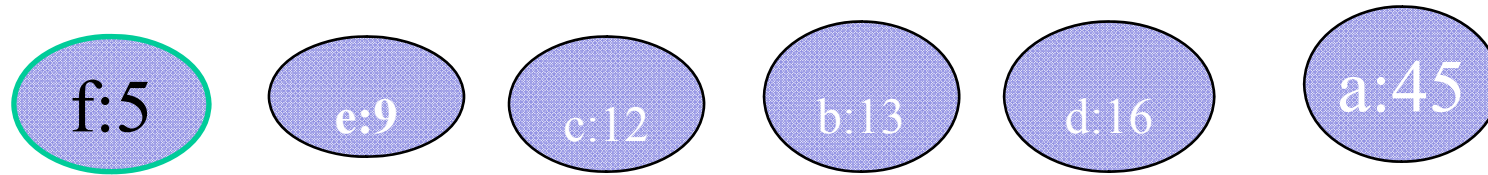
$y=\text{extract-minimum}(Q)$, $\text{right}[z]=y$

$f[z]=f[x]+f[y]$

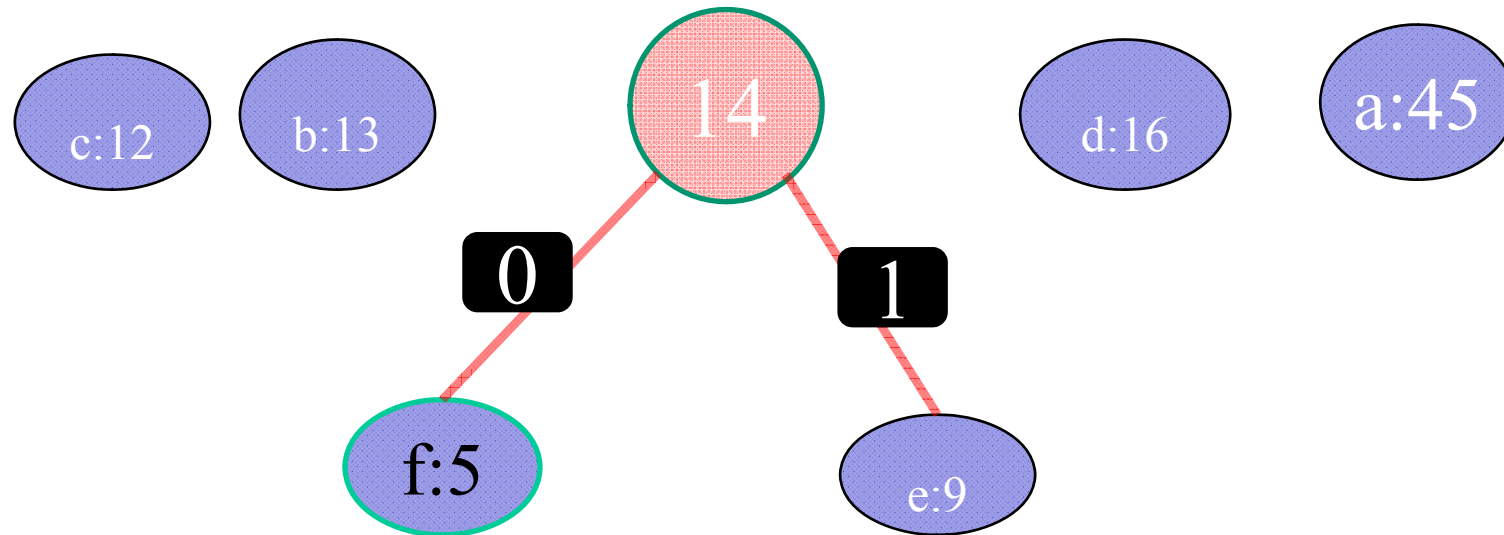
$\text{insert}(Q,z)$

return $\text{extract-minimum}(Q)$ /* ρίζα δέντρου */

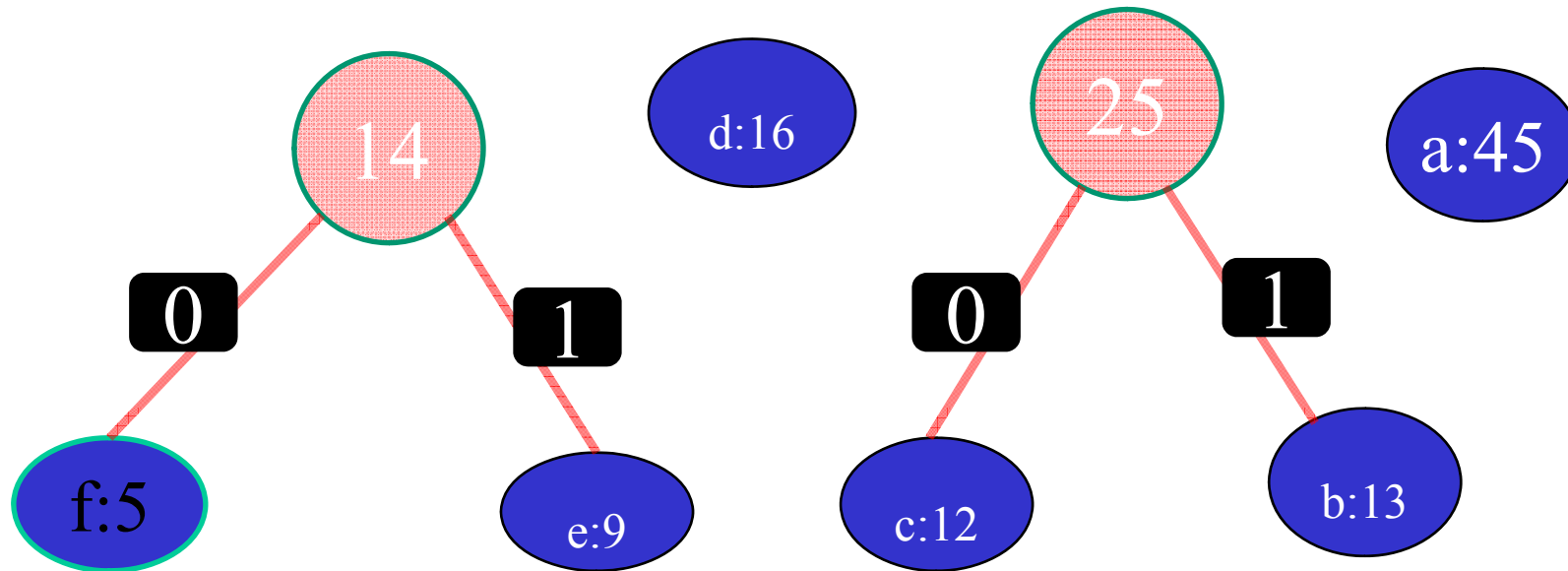
Αλγόριθμος Huffman/παράδειγμα



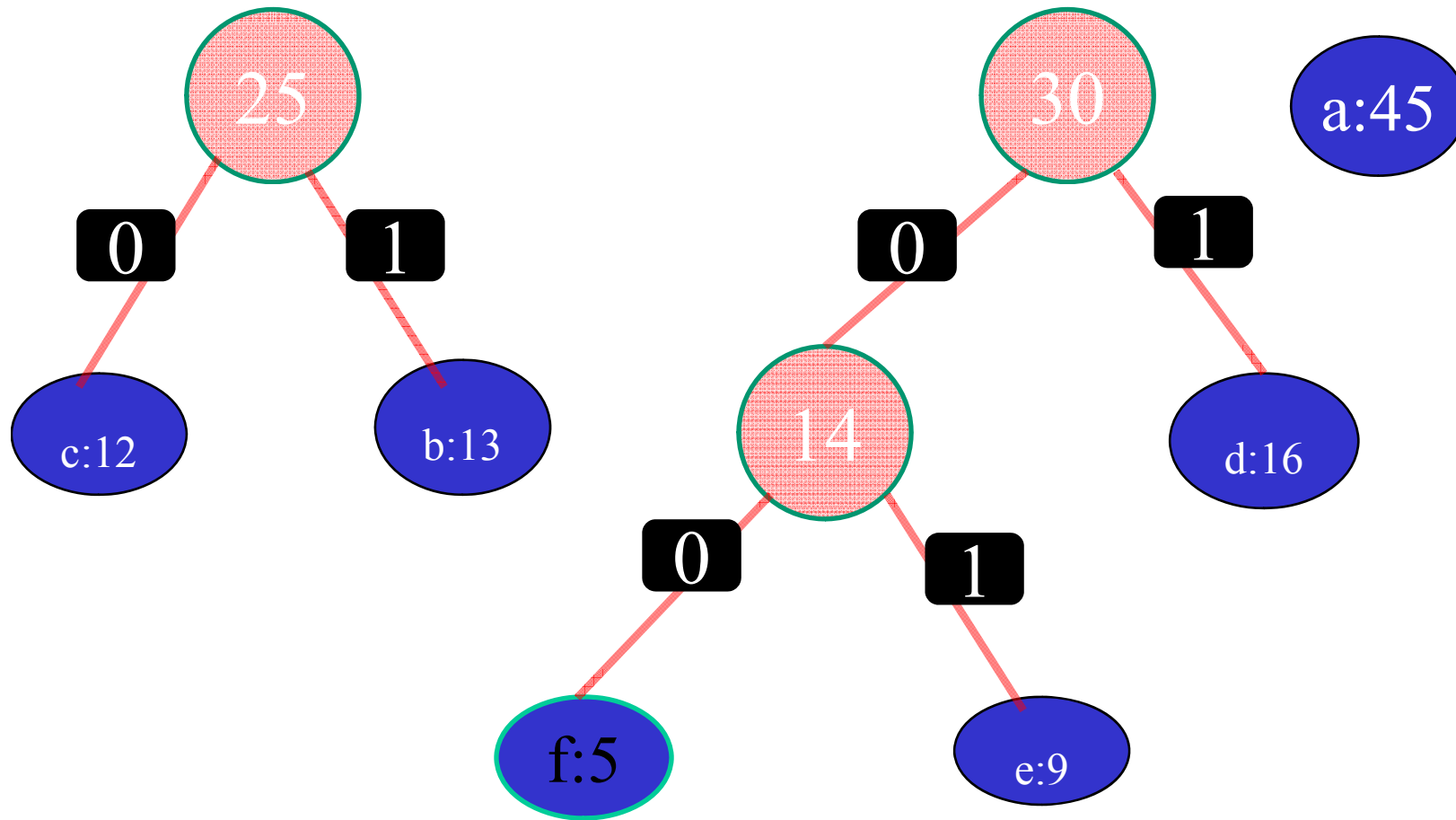
Αλγόριθμος Huffman/παράδειγμα



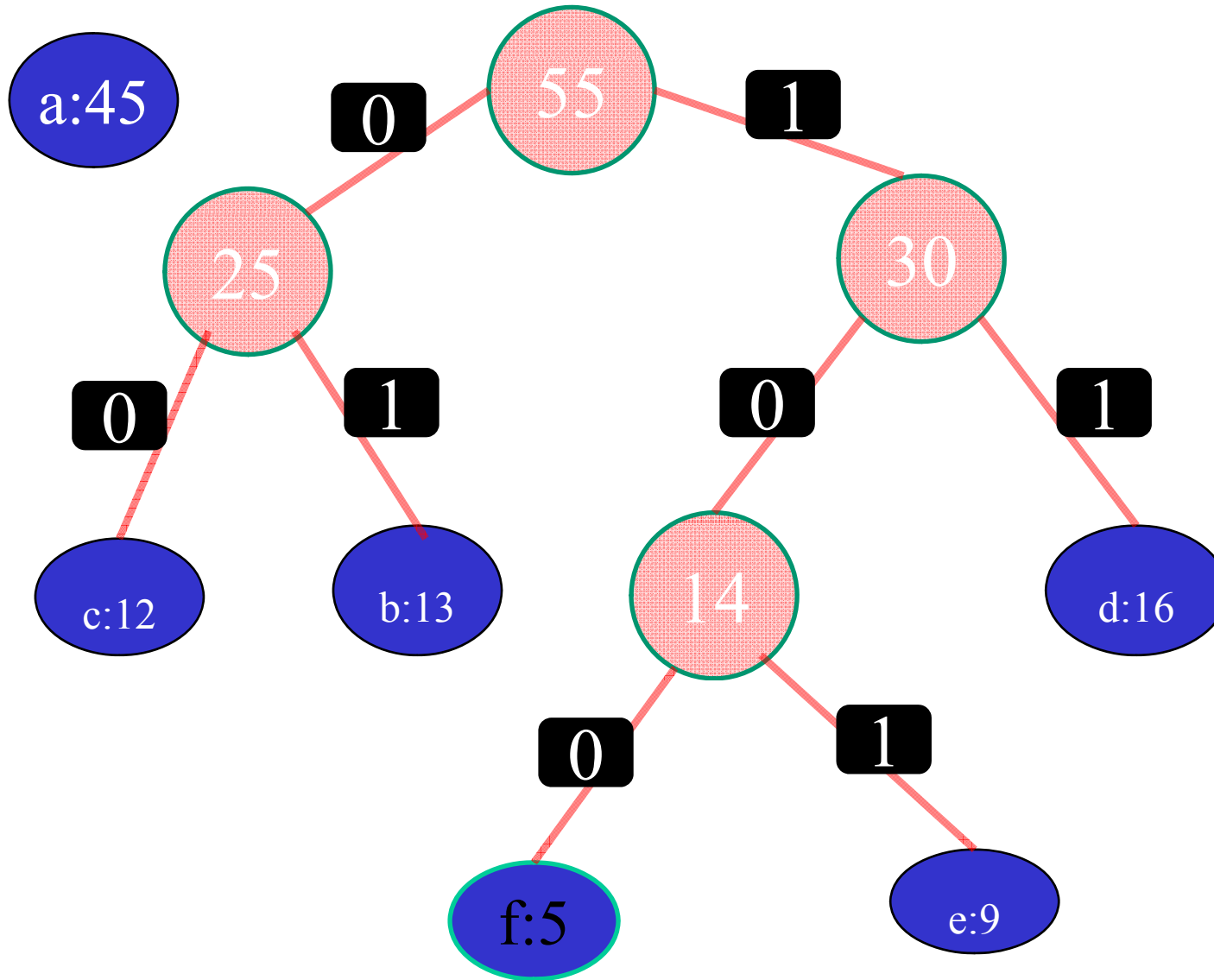
Αλγόριθμος Huffman/παράδειγμα



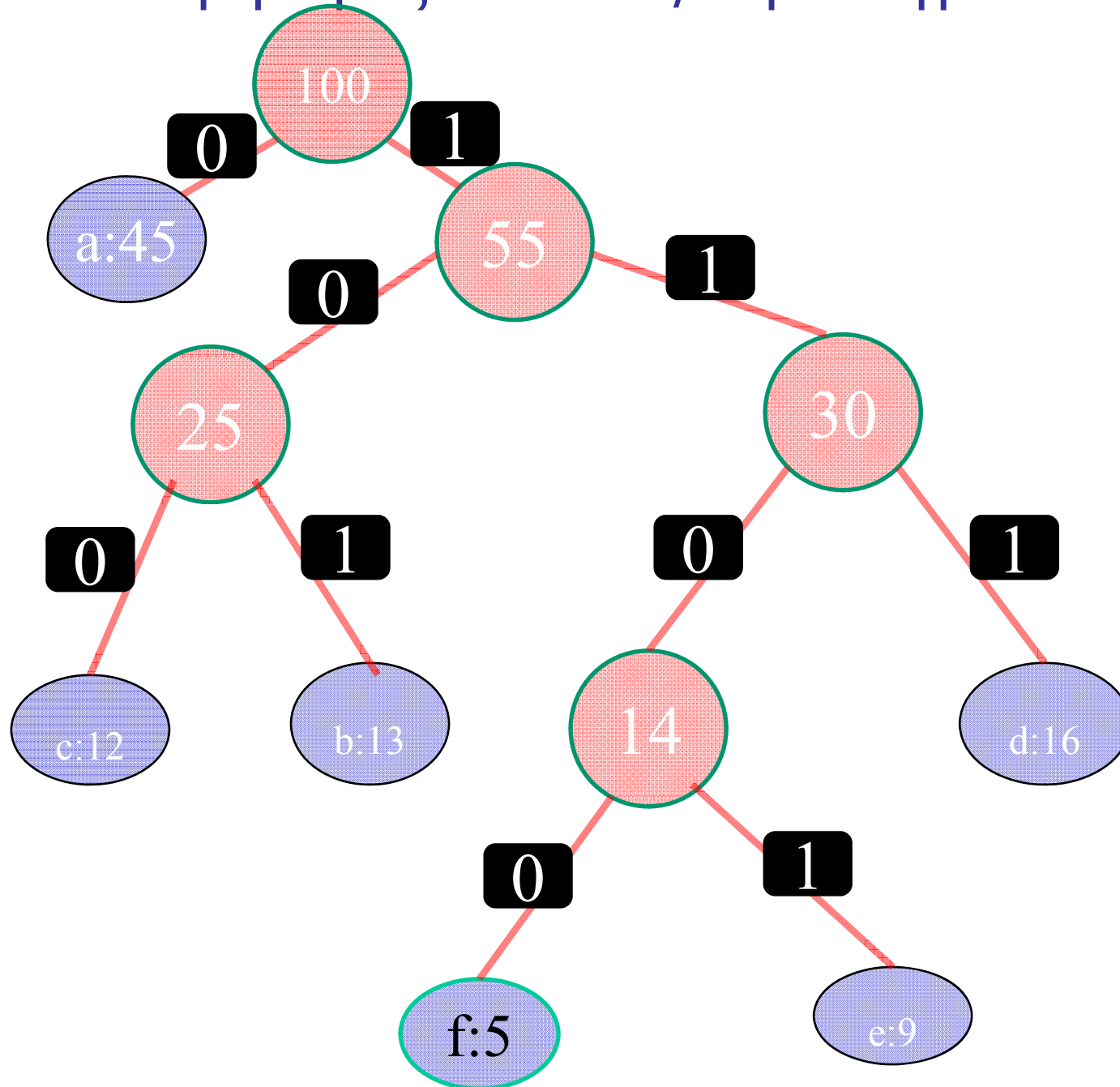
Αλγόριθμος Huffman/παράδειγμα



Αλγόριθμος Huffman/παράδειγμα

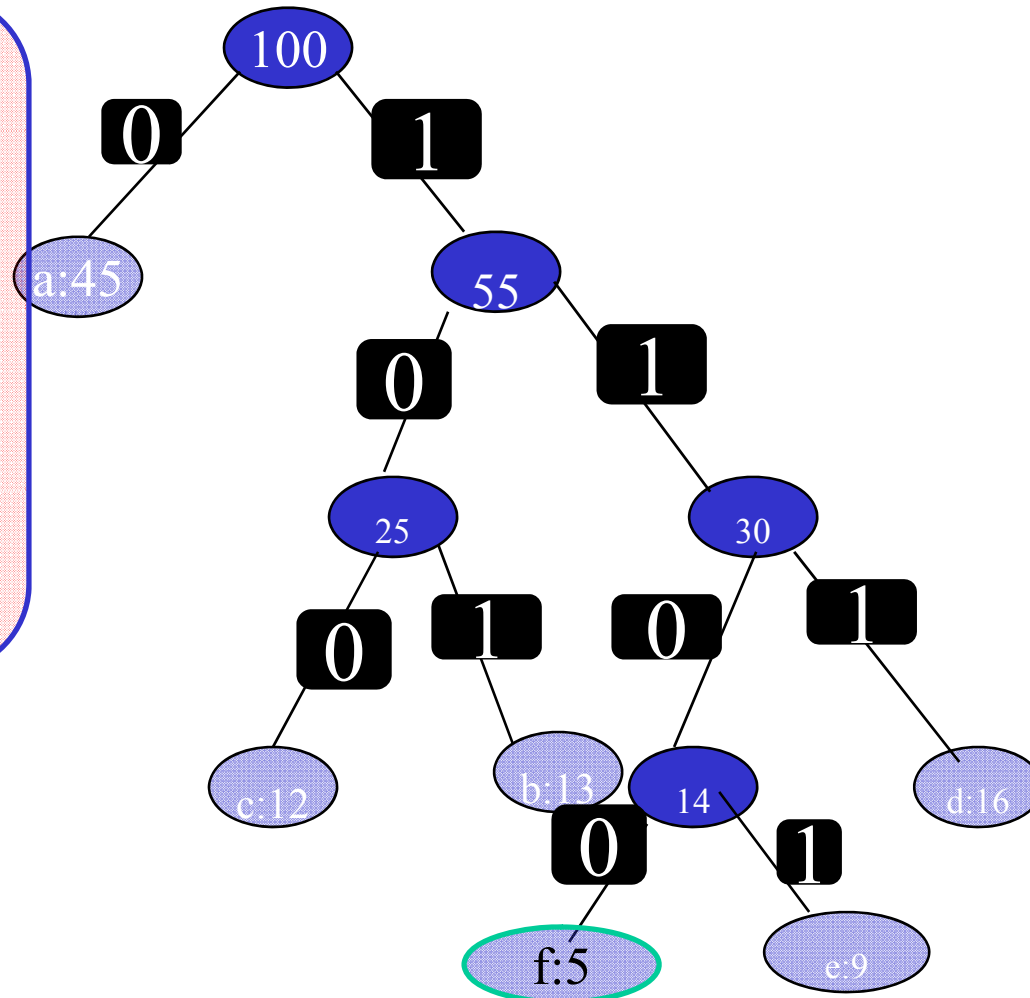


Αλγόριθμος Huffman/παράδειγμα



Αλγόριθμος Huffman/παράδειγμα

a:0
c:100
b:101
f:1100
e: 1101
d:111



Αλγόριθμος Huffman / Πολυπλοκότητα

HUFFMAN(C)

$n=|C|$ /* πλήθος χαρακτήρων */

$Q=C$ /* Σωρός */ $O(n)$

for $i=1$ to $n-1$ do

allocate a new node z

$x=\text{extract-minimum}(Q)$, $\text{left}[z]=x$ $O(\log n)$

$y=\text{extract-minimum}(Q)$, $\text{right}[z]=y$ $O(\log n)$

$f[z]=f[x]+f[y]$

$\text{insert}(Q,z)$ $O(\log n)$

return $\text{extract-minimum}(Q)$ /* ρίζα δέντρου */

Αλγόριθμος Huffman / Άσκηση

Σύμβολα: $\Sigma = \{0, 1\}$

INPUT:

X[1](49) X[2](26) X[3](12) X4 X[5](4) X[6](4) X[7](2)

/*Σε παρένθεση οι συχνότητες του χαρακτήρα X[]*/

OUTPUT: Ένας βέλτιστος κώδικας Huffman

Αλγόριθμος Huffman / Άσκηση

Σύμβολα: $\Sigma = \{0, 1\}$

INPUT:

a(1), b(1), c(2), d(3), e(5), f(8), g(13), h(21)

/*Σε παρένθεση οι συχνότητες του αντίστοιχου χαρακτήρα, 8
πρώτοι όροι της ακολουθίας FIBONACCI */

OUTPUT: Ένας βέλτιστος κώδικας Huffman

Γενικεύστε για να βρείτε το βέλτιστο κώδικα Huffman όταν
οι συχνότητες είναι οι n πρώτοι FIBONACCI αριθμοί.

Κώδικες Huffman

Σύμβολα: $\Sigma = \{0, 1, 2\}$

INPUT:

a(1), b(1), c(2), d(3), e(5), f(8), g(13), h(21)

/*Σε παρένθεση οι συχνότητες του αντίστοιχου χαρακτήρα*/

OUTPUT: Ένας βέλτιστος κώδικας Huffman με 3 σύμβολα
(ternary codes)

Γενικεύστε τον αλγόριθμο Huffman για ternary codes