

Δομές Αναζήτησης

χειρότερη περίπτωση

μέση περίπτωση

εισαγωγή

αναζήτηση

επιλογή

εισαγωγή

αναζήτηση

διατεταγμένος
πίνακας

$O(N)$

$O(\log N)$

$O(1)$

$O(N)$

$O(\log N)$

διατεταγμένη
λίστα

$O(N)$

$O(N)$

$O(N)$

$O(N)$

$O(N)$

μη διατεταγμένος
πίνακας

$O(1)$

$O(N)$

$O(N \log N)^\dagger$

$O(1)$

$O(N)$

μη διατεταγμένη
λίστα

$O(1)$

$O(N)$

$O(N \log N)^\dagger$

$O(1)$

$O(N)$

δένδρο αναζήτησης
(μη ισορροπημένο)

$O(N)$

$O(N)$

$O(N)$

$O(\log N)$

$O(\log N)$

τυχαίοποιημένο
δένδρο

$O(N)^*$

$O(N)^*$

$O(N)^*$

$O(\log N)$

$O(\log N)$

ισορροπημένο
δένδρο αναζήτησης

$O(\log N)$

$O(\log N)$

$O(\log N)$

$O(\log N)$

$O(\log N)$

κατακερματισμός

$O(1)$

$O(N)^*$

$O(N \log N)^\dagger$

$O(1)$

$O(1)$

(*) Συμβαίνει με εξαιρετικά μικρή πιθανότητα

(†) Μετά από ταξινόμηση. Χρόνος $O(N)$ είναι εφικτός με πιο περίπλοκους αλγόριθμους.

Πίνακες Διασποράς (Κατακερματισμός)

Χειριζόμαστε ένα σύνολο στοιχείων S όπου το κάθε στοιχείο έχει ένα κλειδί από ολικά διατεταγμένο σύνολο K

Θέλουμε να υποστηρίξουμε τις βασικές λειτουργίες:

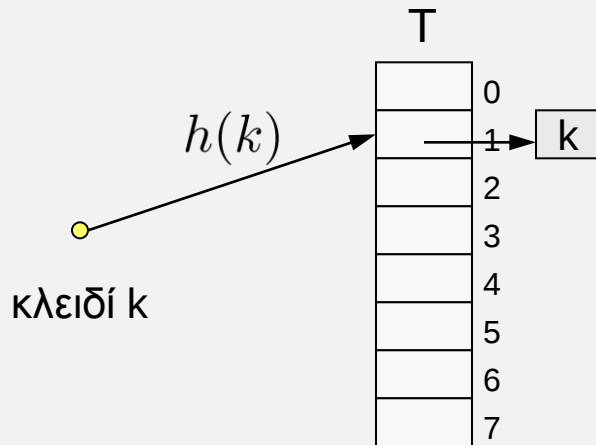
- Εισαγωγή ενός νέου στοιχείου $x \in S$ με κλειδί $k \in K$
- Αναζήτηση στοιχείου με δεδομένο κλειδί $k \in K$
- Διαγραφή ενός καθορισμένου στοιχείου

Ο κατακερματισμός δίνει μια πολύ αποδοτική λύση (απλή και γρήγορη στην πράξη) όταν χρειαζόμαστε μόνο τις παραπάνω λειτουργίες

Πίνακες Διασποράς

Χρησιμοποιούμε ένα πίνακα διασποράς T και μια συνάρτηση διασποράς h

Ένα στοιχείο με κλειδί k αποθηκεύεται στη θέση $T[h(k)]$



U : χώρος πιθανών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$h : U \rightarrow \{0, 1, \dots, m - 1\}$

Πίνακες Διασποράς

Μετατροπή σε ακέραιο στο διάστημα $[0, m - 1]$

- για αριθμό κινητής υποδιαστολής στο διάστημα $(0, 1)$

```
int m, k;  
float x;
```

π.χ. για $m = 67$

$x = .57641 \Rightarrow k = 38$

$x = .32178 \Rightarrow k = 21$

$x = .58113 \Rightarrow k = 38$

```
k = x*m;
```

Πίνακες Διασποράς

Μετατροπή σε ακέραιο στο διάστημα $[0, m - 1]$

- για αλφαριθμητικά με χαρακτήρες

κάθε χαρακτήρας αντιστοιχεί σε ένα ακέραιο σε κωδικοποίηση ASCII

έχουμε 'a' = 97, 'b' = 98, 'c' = 99, ... κλπ

π.χ. για $m = 67$

String $s = \text{"have"}$ $\Rightarrow$

$$k = (104 + 97 + 118 + 101) \bmod 67$$

$$= 420 \bmod 67 = 18$$

Πίνακες Διασποράς

Μετατροπή σε ακέραιο στο διάστημα $[0, m - 1]$

- για αλφαριθμητικά με χαρακτήρες

κάθε χαρακτήρας αντιστοιχεί σε ένα ακέραιο σε κωδικοποίηση ASCII

έχουμε 'a' = 97, 'b' = 98, 'c' = 99, ... κλπ

π.χ. για $m = 67$

String $s = \text{"have"}$ $\Rightarrow$

$$\begin{aligned} k &= (104 + 97 + 118 + 101) \bmod 67 \\ &= 420 \bmod 67 = 18 \end{aligned}$$

Αντιστοιχεί στον ίδιο ακέραιο αλφαριθμητικά που προκύπτουν από μεταθέσεις των ίδιων χαρακτήρων.

Π.χ. "stop" – "tops" – "pots" – "spot", «γραφή» – «φραγή»

Για να το αποφύγουμε πολλαπλασιάζουμε με ένα συντελεστή βάρους w^j σε κάθε θέση j

Πίνακες Διασποράς

Μετατροπή σε ακέραιο στο διάστημα $[0, m - 1]$

- για αλφαριθμητικά με χαρακτήρες των 7 bit ($w = 2^7 = 128$ χαρακτήρες)

κάθε χαρακτήρας αντιστοιχεί σε ένα ακέραιο σε κωδικοποίηση ASCII

έχουμε 'a' = 97, 'b' = 98, 'c' = 99, ... κλπ

π.χ. για $m = 67$

String $s = \text{"have"}$ $\Rightarrow$

$$\begin{aligned} k &= (104*128^3 + 97*128^2 + 118*128^1 + 101*128^0) \bmod 67 \\ &= 219708261 \bmod 67 = 52 \end{aligned}$$

για να αποφύγουμε το πρόβλημα της υπερχείλισης χρησιμοποιούμε τον αλγόριθμο του Horner και τις αριθμητικές ιδιότητες της συνάρτησης `mod`

$$\begin{aligned} &104*128^3 + 97*128^2 + 118*128^1 + 101*128^0 = \\ &((104*128 + 97)*128 + 118)*128 + 101 \end{aligned}$$

Πίνακες Διασποράς

Μετατροπή σε ακέραιο στο διάστημα $[0, m - 1]$

- για αλφαριθμητικά με χαρακτήρες των 7 bit ($w = 2^7 = 128$ χαρακτήρες)

κάθε χαρακτήρας αντιστοιχεί σε ένα ακέραιο σε κωδικοποίηση ASCII

έχουμε 'a' = 97, 'b' = 98, 'c' = 99, ... κλπ

π.χ. για $m = 67$

String s = "have" $\Rightarrow$

$$k = (((104 * 128 + 97) * 128 + 118) * 128 + 101) \bmod 67 = 52$$

```
int w = 128;
```

```
int k = 0;
```

```
for (i=0; i<n; i++){ k = (k*w + s.charAt(i)) % m; }
```

Πίνακες Διασποράς

Μετατροπή αντικειμένου σε ακέραιο στη Java

Μέθοδος `hashCode()` : Επιστρέφει ακέραιο 32-bit

Για να λάβουμε ακέραιο στο διάστημα $[0, m - 1]$ μπορούμε να υπολογίσουμε

```
private int hash(Key k)
{
    return ( ( k.hashCode() & 0x7fffffff ) % m );
}
```

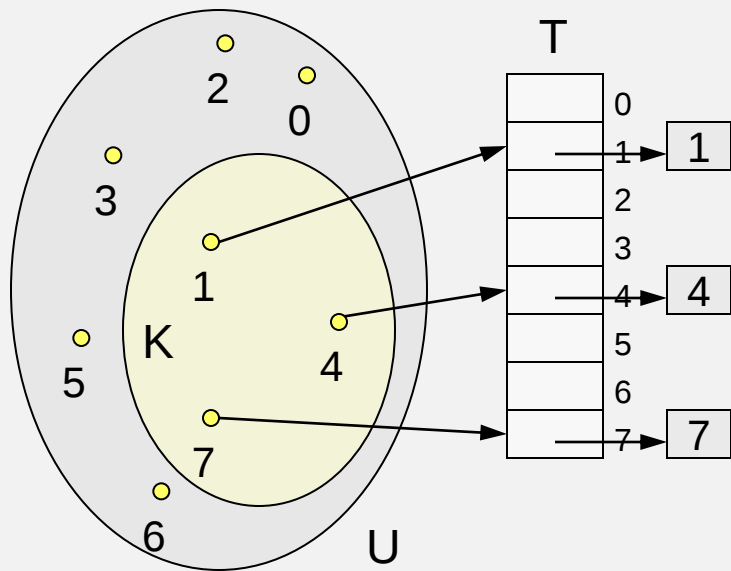
Προσοχή: Η μέθοδος `hashCode()` μπορεί να μην είναι καλή επιλογή για ορισμένα αντικείμενα. (Π.χ. μπορεί να επιστρέφει αναφορά σε θέση μνήμης του αντικειμένου.)

Πίνακες σταθερών διευθύνσεων

Απλή λύση όταν το U είναι μικρό:

$$m \leftarrow |U|$$

$$h : U \rightarrow \{0, 1, \dots, m - 1\} \quad \text{όπου} \quad h(k) = k$$



U : χώρος πιθανών κλειδιών

K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$$h : U \rightarrow \{0, 1, \dots, m - 1\}$$

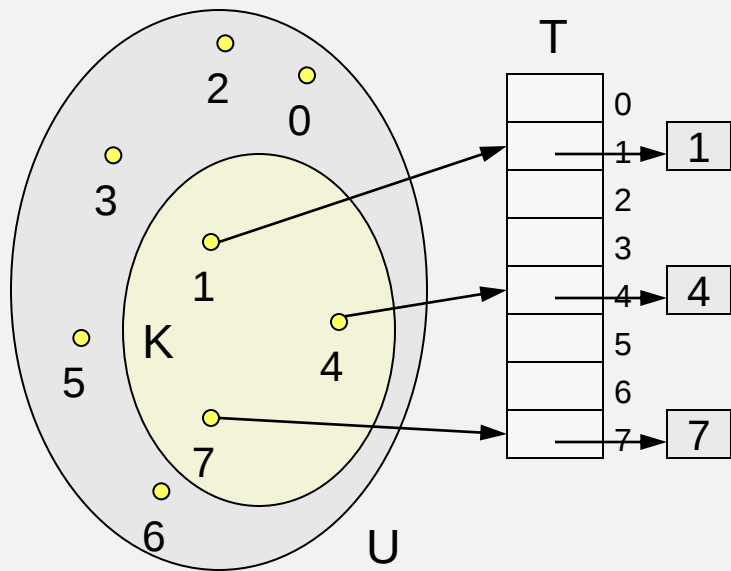
Πίνακες σταθερών διευθύνσεων

Απλή λύση όταν το U είναι μικρό:

$$m \leftarrow |U|$$

$$h : U \rightarrow \{0, 1, \dots, m - 1\} \quad \text{όπου} \quad h(k) = k$$

Δίνει $O(1)$ χρόνο ανά πράξη.



U : χώρος πιθανών κλειδιών

K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$$h : U \rightarrow \{0, 1, \dots, m - 1\}$$

Πίνακες σταθερών διευθύνσεων

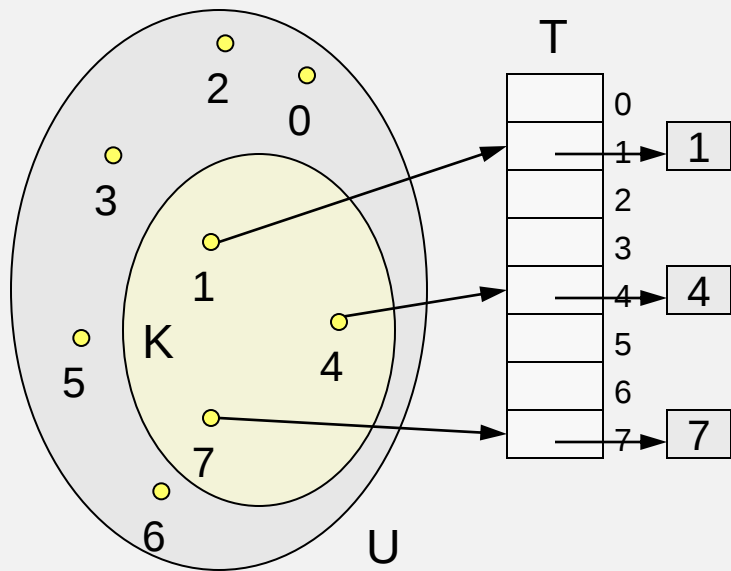
Απλή λύση όταν το U είναι μικρό:

$$m \leftarrow |U|$$

$$h : U \rightarrow \{0, 1, \dots, m - 1\} \quad \text{όπου} \quad h(k) = k$$

Δίνει $O(1)$ χρόνο ανά πράξη.

Τι γίνεται όμως αν το U είναι πολύ μεγάλο;



U : χώρος πιθανών κλειδιών

K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$$h : U \rightarrow \{0, 1, \dots, m - 1\}$$

Πίνακες σταθερών διευθύνσεων

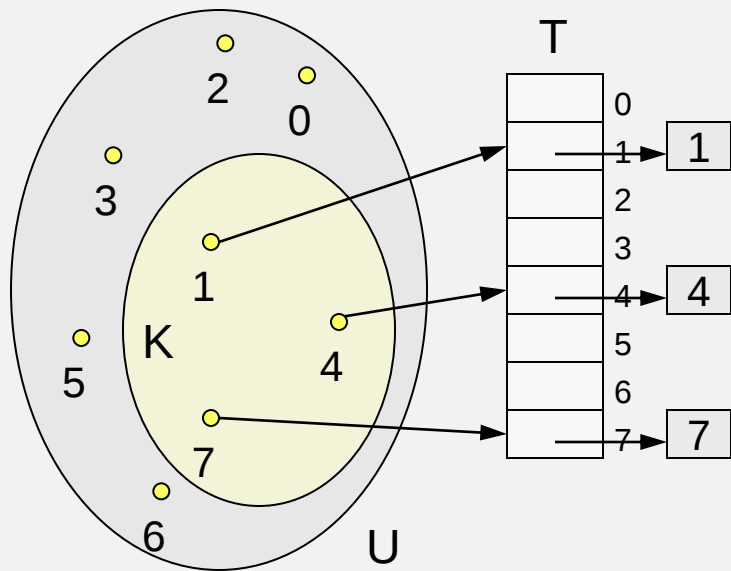
Απλή λύση όταν το U είναι μικρό:

$$m \leftarrow |U|$$

$$h : U \rightarrow \{0, 1, \dots, m - 1\} \quad \text{όπου} \quad h(k) = k$$

Δίνει $O(1)$ χρόνο ανά πράξη.

Τι γίνεται όμως αν το U είναι πολύ μεγάλο και το K πολύ μικρότερο;



U : χώρος πιθανών κλειδιών

K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

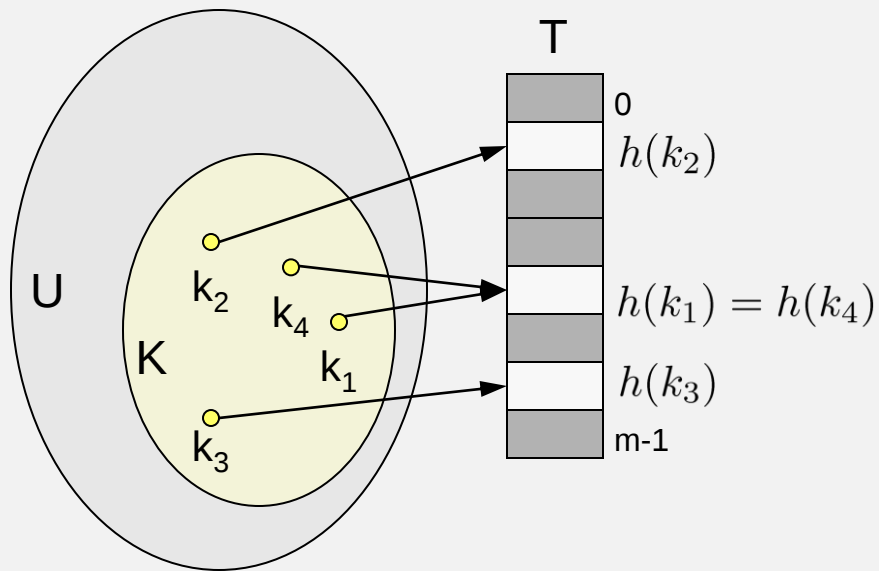
$$h : U \rightarrow \{0, 1, \dots, m - 1\}$$

Πίνακες Διασποράς

Επιθυμητές ιδιότητες:

$O(|K|)$ χώρος

$O(1)$ χρόνος ανά λειτουργία (αναμενόμενη περίπτωση)



U : χώρος πιθανών κλειδιών

K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$h : U \rightarrow \{0, 1, \dots, m - 1\}$

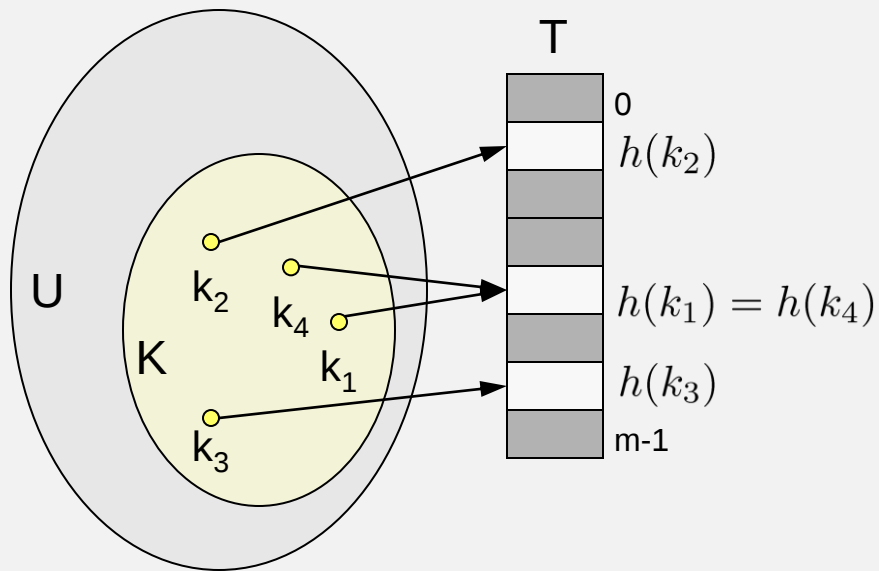
Πίνακες Διασποράς

Επιθυμητές ιδιότητες:

$O(|K|)$ χώρος

$O(1)$ χρόνος ανά λειτουργία (αναμενόμενη περίπτωση)

Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$



U : χώρος πιθανών κλειδιών

K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$h : U \rightarrow \{0, 1, \dots, m - 1\}$

Πίνακες Διασποράς

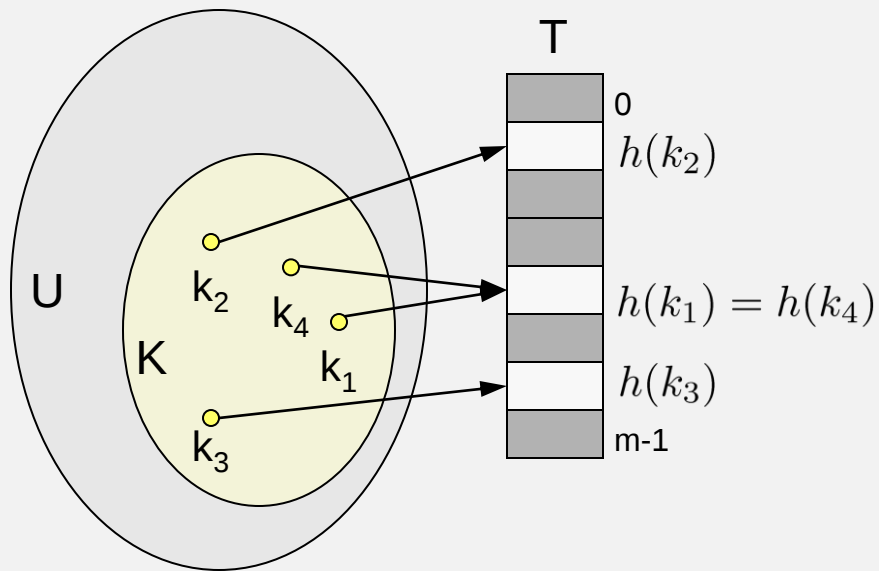
Επιθυμητές ιδιότητες:

$O(|K|)$ χώρος

$O(1)$ χρόνος ανά λειτουργία (αναμενόμενη περίπτωση)

Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Η h πρέπει να δίνει τυχαιόμορφη κατανομή



U : χώρος πιθανών κλειδιών

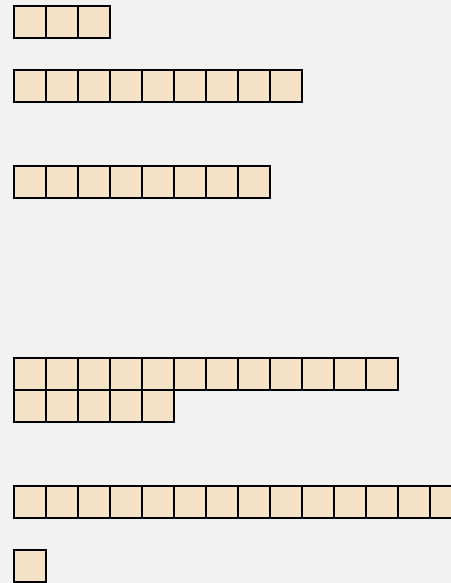
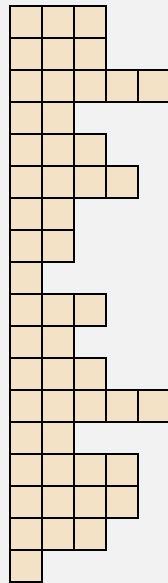
K : χώρος ενεργών κλειδιών

T : πίνακας μεγέθους m

συνάρτηση διασποράς

$h : U \rightarrow \{0, 1, \dots, m - 1\}$

Πίνακες Διασποράς



Ποιά συνάρτηση διασποράς είναι καλύτερη;

Πίνακες Διασποράς

Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Πρέπει να οριστεί ένας τρόπος διαχείρισης συγκρούσεων ώστε τα κλειδιά να μπορούν να βρεθούν αρκετά γρήγορα.

Χρησιμοποιούνται 2 κύριοι τρόποι:

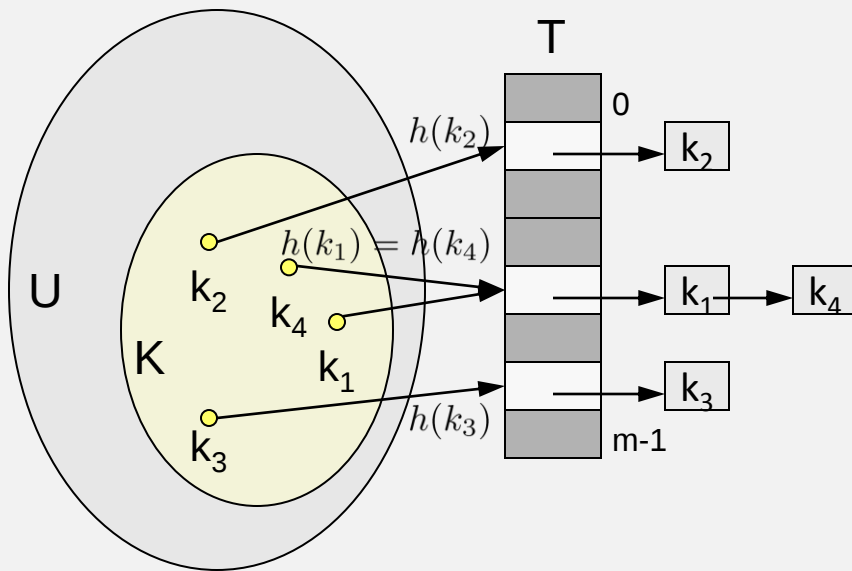
- Τοποθέτηση στην ίδια θέση (π.χ., σε αλυσίδα)
- Τοποθέτηση σε άλλη θέση

Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Άρση των συγκρούσεων μέσω **αλυσιδωτής σύνδεσης**

Κάθε $T[i]$ δείχνει σε μία αλυσίδα για τα κλειδιά k με διασπορά $h[k] = i$



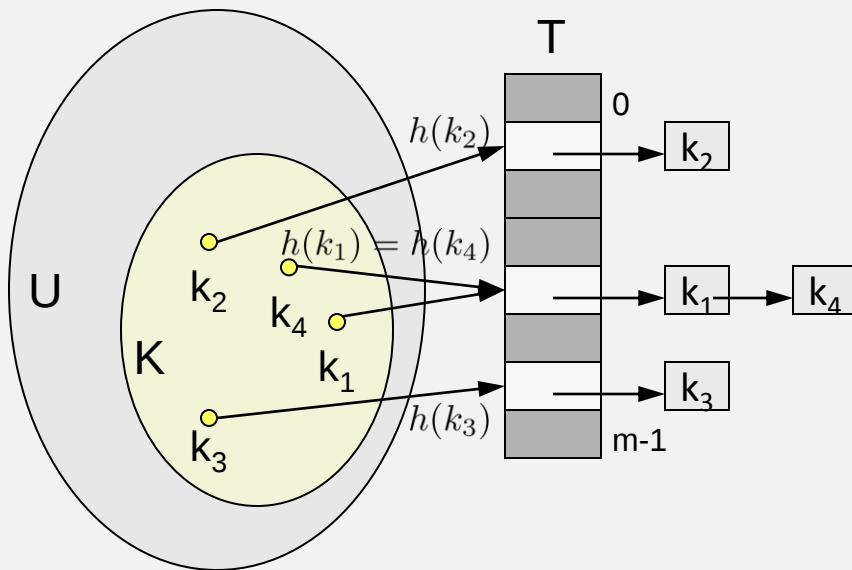
Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Άρση των συγκρούσεων μέσω **αλυσιδωτής σύνδεσης**

Κάθε $T[i]$ δείχνει σε μία αλυσίδα για τα κλειδιά k με διασπορά $h[k] = i$

εισαγωγή(T,x): εισαγωγή του x στη λίστα του $T[h(\kappa\lambda\epsilon\iota\delta\acute{\iota}[x])]$



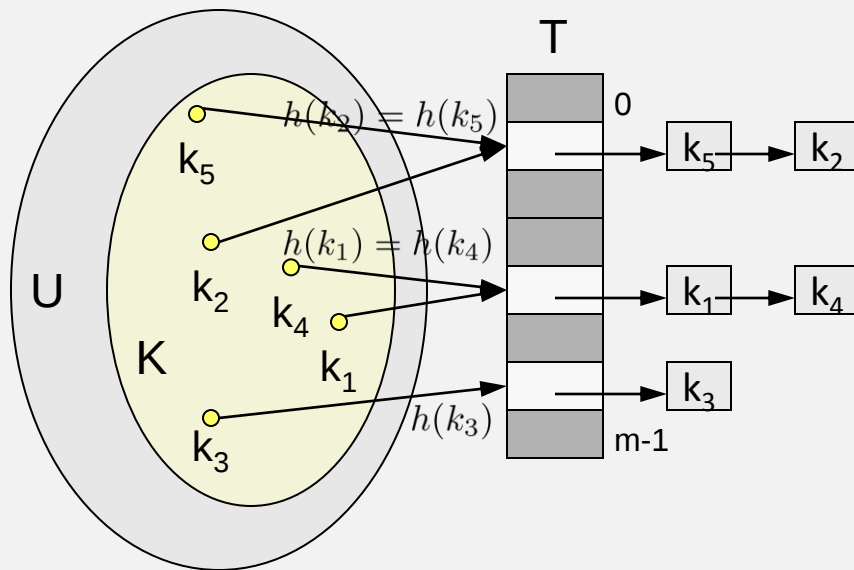
Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Άρση των συγκρούσεων μέσω **αλυσιδωτής σύνδεσης**

Κάθε $T[i]$ δείχνει σε μία αλυσίδα για τα κλειδιά k με διασπορά $h[k] = i$

εισαγωγή(T,x): εισαγωγή του x στη λίστα του $T[h(\kappaλειδί[x])]$



Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

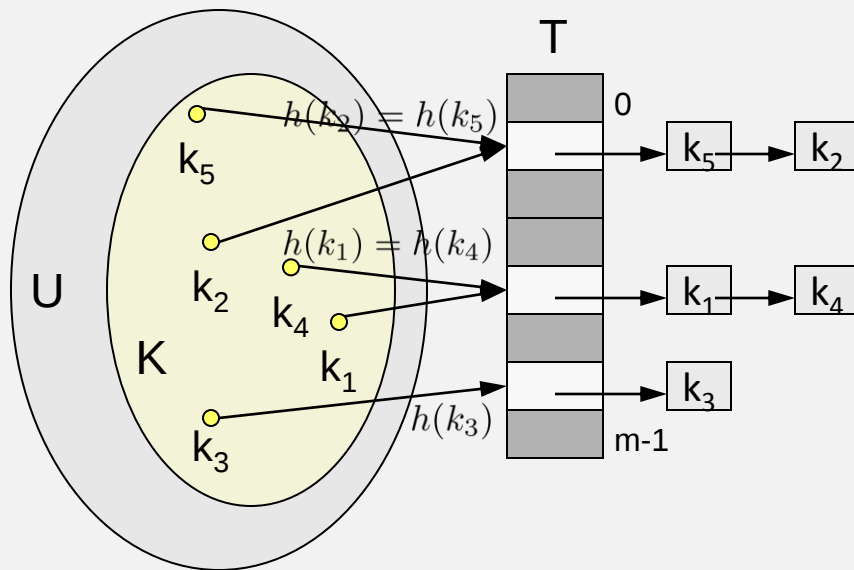
Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Άρση των συγκρούσεων μέσω **αλυσιδωτής σύνδεσης**

Κάθε $T[i]$ δείχνει σε μία αλυσίδα για τα κλειδιά k με διασπορά $h[k] = i$

εισαγωγή(T,x): εισαγωγή του x στη λίστα του $T[h(\text{κλειδί}[x])]$

αναζήτηση(T,k): αναζήτηση στοιχείου με κλειδί k στην αλυσίδα $T[h(k)]$



Η εισαγωγή παίρνει $O(1)$ χρόνο αν γίνεται στην αρχή της αλυσίδας και δεν κάνουμε αναζήτηση αν το x είναι ήδη καταχωρημένο. Διαφορετικά πρέπει να προηγηθεί αναζήτηση του x .

Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

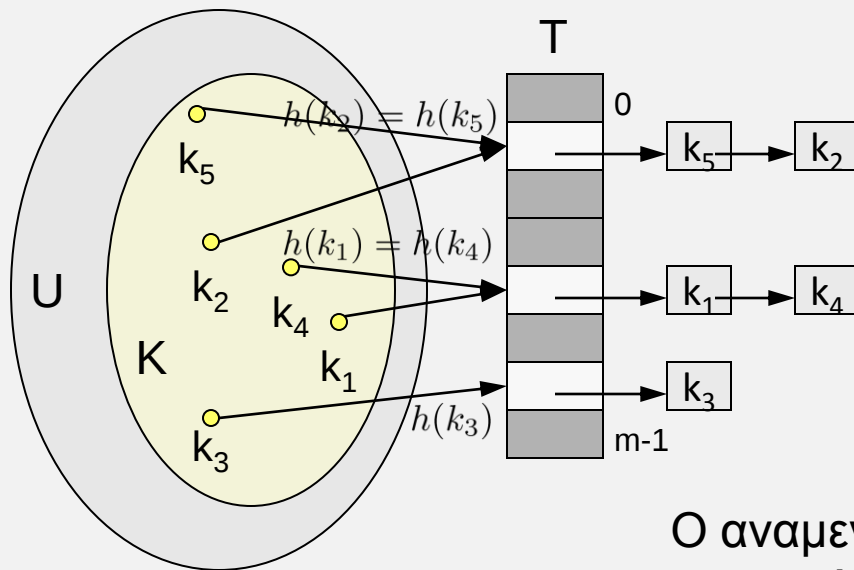
Σύγκρουση (collision): $h(k) = h(\ell)$ για κλειδιά $k \neq \ell$

Άρση των συγκρούσεων μέσω **αλυσιδωτής σύνδεσης**

Κάθε $T[i]$ δείχνει σε μία αλυσίδα για τα κλειδιά k με διασπορά $h[k] = i$

εισαγωγή(T,x): εισαγωγή του x στη λίστα του $T[h(\kappa\lambda\epsilon\iota\delta\acute{\iota}[x])]$

αναζήτηση(T,k): αναζήτηση στοιχείου με κλειδί k στην αλυσίδα $T[h(k)]$



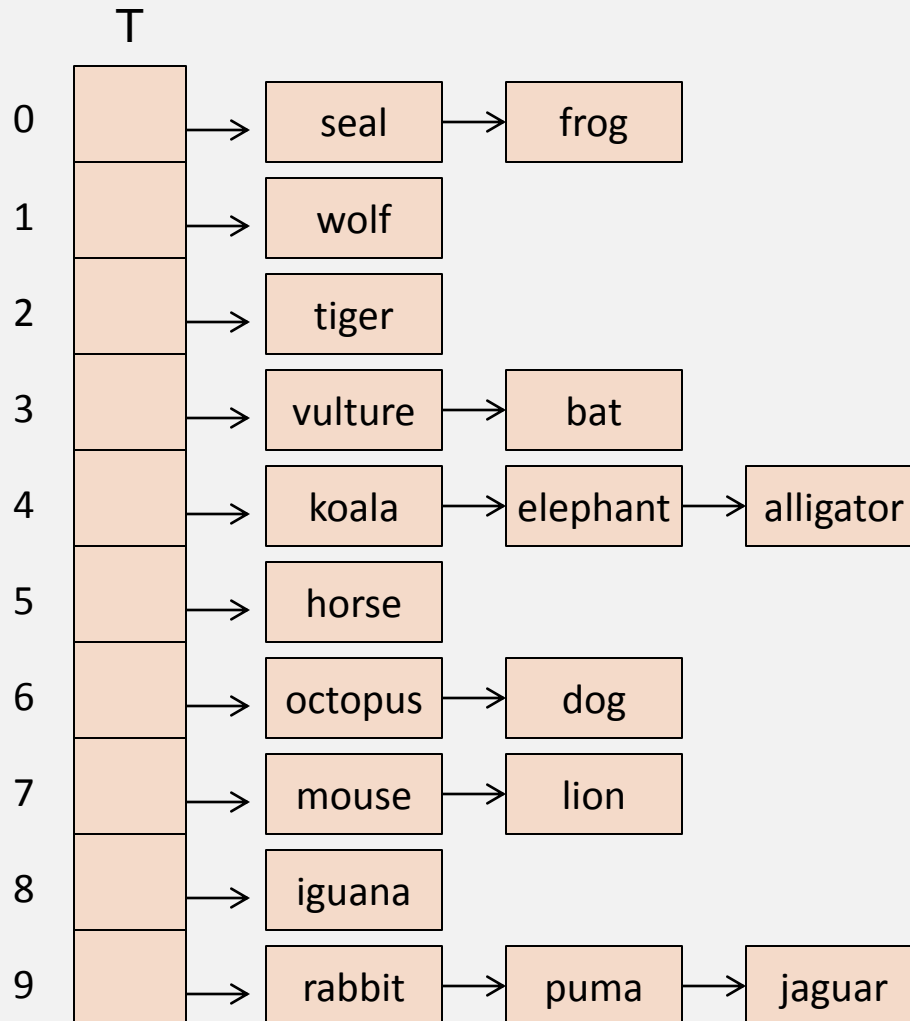
διαγραφή(T,x): διαγραφή του x από την αλυσίδα του $T[h(\kappa\lambda\epsilon\iota\delta\acute{\iota}[x])]$

Η διαγραφή παίρνει $O(1)$ χρόνο αν η θέση του x είναι γνωστή. (Δηλαδή έχουμε δείκτη στον κόμβο του x .) Διαφορετικά πρέπει να προηγηθεί αναζήτηση του x .

Ο αναμενόμενος χρόνος αναζήτησης εξαρτάται από το αναμενόμενο μήκος μιας αλυσίδας.

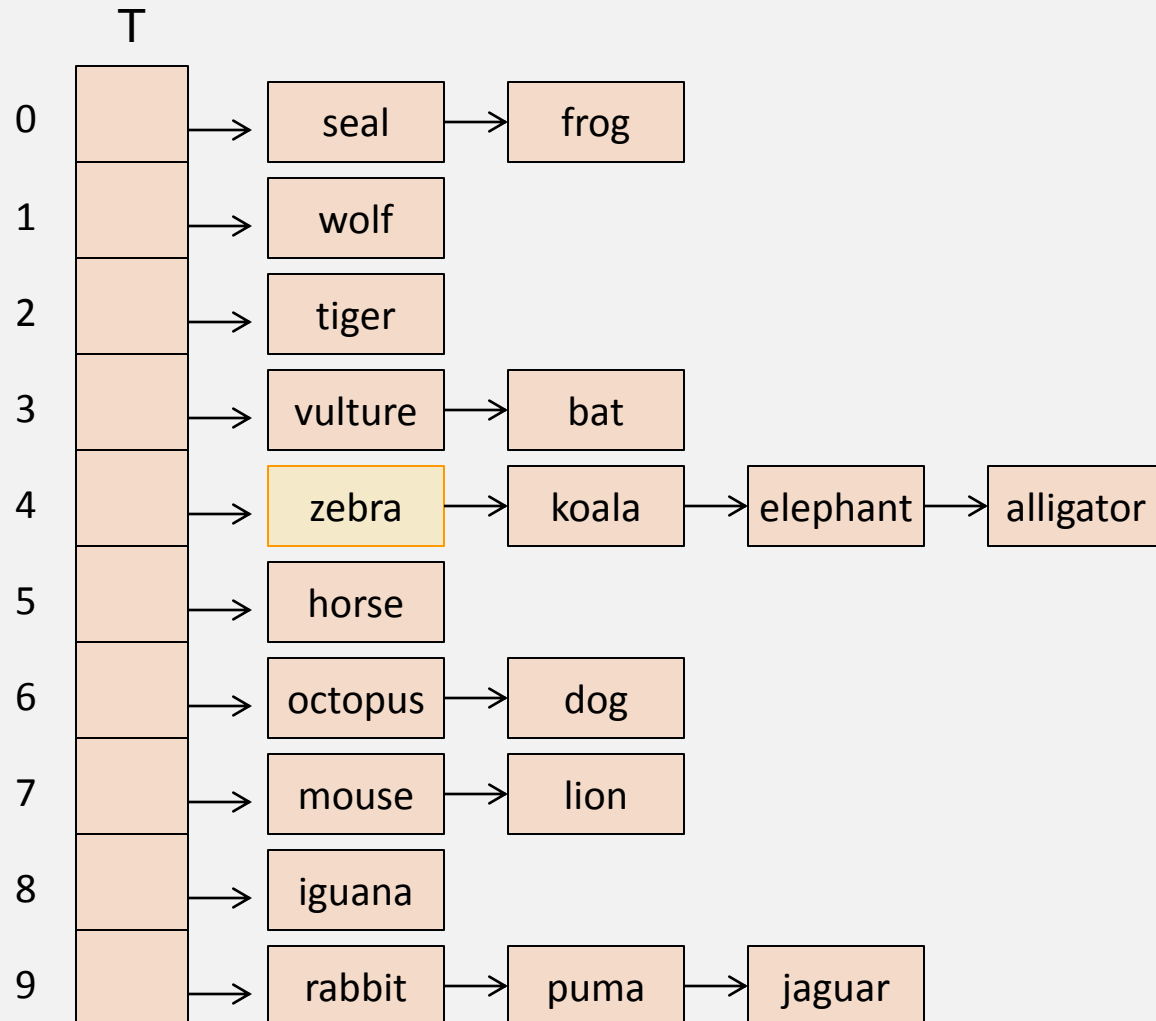
Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

Εισαγωγή “zebra” $\Rightarrow$ key = 115776262 $\Rightarrow$ hash = 4



Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

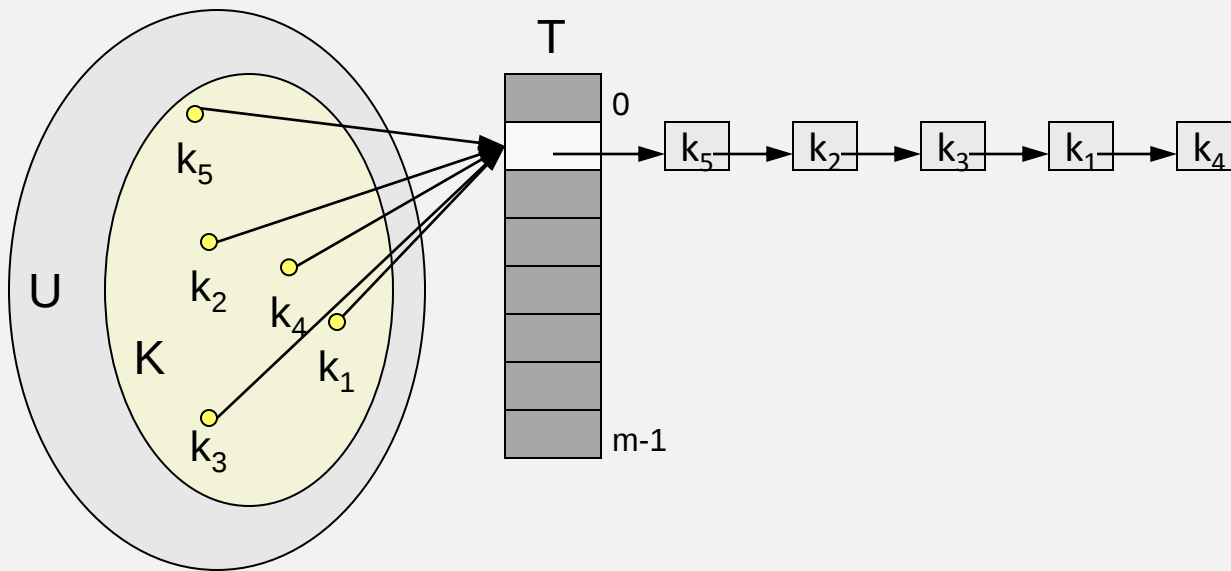
Εισαγωγή “zebra” $\Rightarrow$ key = 115776262 $\Rightarrow$ hash = 4



Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

Συντελεστής πληρότητας $\alpha = n/m$ $n = \text{πλήθος των ενεργών κλειδιών}$

Στη χειρότερη περίπτωση η διασπορά $h(k)$ είναι ίδια για όλα τα $k \in K$
 $\Rightarrow \Theta(n)$ χρόνος αναζήτησης στη χειρότερη περίπτωση.



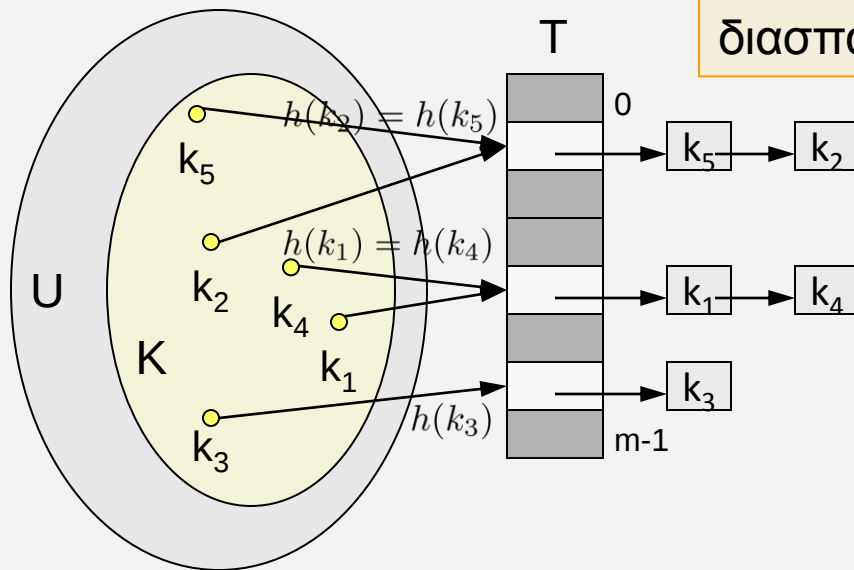
Πίνακες Διασποράς Αλυσιδωτής Σύνδεσης

Συντελεστής πληρότητας $\alpha = n/m$ n = πλήθος των ενεργών κλειδιών

Έστω n_i το μήκος της αλυσίδας του $T[i]$.

Έχουμε $n = n_0 + n_1 + \dots + n_{m-1}$ και επομένως το μέσο μήκος μιας αλυσίδας είναι α .

Υπόθεση απλής ομοιόμορφης διασποράς:
κάθε **νέο** στοιχείο έχει ίση πιθανότητα να διασπαρεί σε οποιαδήποτε από τις m θέσεις.



Αναμενόμενος χρόνος
ανεπιτυχούς αναζήτησης: $U \approx \alpha$

Αναμενόμενος χρόνος
επιτυχούς αναζήτησης: $S \approx 1 + \frac{\alpha}{2}$

Μια επιτυχής αναζήτηση διασχίζει κατά μέσο όρο τους μισούς κόμβους αλυσίδας που δεν είναι κενή.

Συναρτήσεις Διασποράς

Πότε είναι μία συνάρτηση διασποράς καλή;

Επιθυμητή ιδιότητα: να ικανοποιεί την υπόθεση της **απλής ομοιόμορφης διασποράς** (δηλ. κάθε νέο στοιχείο έχει ίση πιθανότητα να διασπαρεί σε οποιαδήποτε από τις m θέσεις).

Παράδειγμα: Τα κλειδιά είναι ανεξάρτητοι και ομοιόμορφα κατανεμημένοι αριθμοί στο διάστημα $[0,1]$. Επιλέγοντας $h(k) = \lfloor km \rfloor$ ικανοποιούμε την υπόθεση της απλής ομοιόμορφης διασποράς.

Η παραπάνω ιδιότητα απαιτεί γνώση της κατανομής των κλειδιών, που συνήθως δεν είναι γνωστή.

Συναρτήσεις Διασποράς

Πότε είναι μία συνάρτηση διασποράς καλή;

Επιθυμητή ιδιότητα: να ικανοποιεί την υπόθεση της **απλής ομοιόμορφης διασποράς** (δηλ. κάθε νέο στοιχείο έχει ίση πιθανότητα να διασπαρεί σε οποιαδήποτε από τις m θέσεις).

Διαιρετική μέθοδος $h(k) = k \bmod m$

κακή επιλογή : $m = \text{δύναμη του } 2$

καλή επιλογή : $m = \text{πρώτος αριθμός}$

Πολλαπλασιαστική μέθοδος $h(k) = \lfloor m(kA \bmod 1) \rfloor, \quad 0 < A < 1$

$$kA \bmod 1 = kA - \lfloor kA \rfloor$$

Πολλαπλασίασε πρόσθεσε και διαίρεσε (Multiply Add and Divide)

$$h(k) = [(ak + b) \bmod p] \bmod m$$

p πρώτος αριθμός $> m$

a, b ακέραιοι στο διάστημα $[0, p - 1]$ με $a > 0$

Καθολική Διασπορά

Για κάθε καθορισμένη συνάρτηση διασποράς μπορούμε να επιλέξουμε κλειδιά που θα δίνουν τη χειρότερη δυνατή επίδοση.

Προκειμένου να βελτιώσουμε την κατάσταση μπορούμε να βασιστούμε σε **τυχαιοκρατικούς αλγόριθμους**:

Ορίζουμε μία οικογένεια συναρτήσεων και κατά τη διάρκεια της εκτέλεσης επιλέγουμε τυχαία μία από αυτές τις συναρτήσεις ως συνάρτηση διασποράς.

⇒ Περιορίζουμε την πιθανότητα εμφάνισης παθολογικών περιπτώσεων.

Καθολική Διασπορά

$\mathcal{H}$ πεπερασμένη συλλογή συναρτήσεων

Η συλλογή $\mathcal{H}$ είναι καθολική αν για κάθε ζεύγος διαφορετικών κλειδιών $k, \ell \in U$ υπάρχουν το πολύ $|\mathcal{H}|/m$ συναρτήσεις $h \in \mathcal{H}$ με $h(k) = h(\ell)$

Καθολική Διασπορά

$\mathcal{H}$ πεπερασμένη συλλογή συναρτήσεων

Η συλλογή $\mathcal{H}$ είναι καθολική αν για κάθε ζεύγος διαφορετικών κλειδιών $k, \ell \in U$ υπάρχουν το πολύ $|\mathcal{H}|/m$ συναρτήσεις $h \in \mathcal{H}$ με $h(k) = h(\ell)$

Συνεπώς $\Pr_{h \in \mathcal{H}}[h(k) = h(\ell)] \leq 1/m$

Καθολική Διασπορά

$\mathcal{H}$ πεπερασμένη συλλογή συναρτήσεων

Η συλλογή $\mathcal{H}$ είναι καθολική αν για κάθε ζεύγος διαφορετικών κλειδιών $k, \ell \in U$ υπάρχουν το πολύ $|\mathcal{H}|/m$ συναρτήσεις $h \in \mathcal{H}$ με $h(k) = h(\ell)$

Συνεπώς $\Pr_{h \in \mathcal{H}}[h(k) = h(\ell)] \leq 1/m$

↑
Πιθανότητα σύμπτωσης όταν $h(k)$ και $h(\ell)$ επιλέγονται **τυχαία** και **ανεξάρτητα**.

Καθολική Διασπορά

$\mathcal{H}$ = καθολική συλλογή συναρτήσεων διασποράς που αντιστοιχίζει το σύμπαν U στο σύνολο $\{0, 1, \dots, m-1\}$

K = αυθαίρετο υποσύνολο του U με $|K| = n \leq m$

u = αυθαίρετο στοιχείο του U

h = τυχαία επιλεγμένη συνάρτηση της $\mathcal{H}$

X = αριθμός στοιχείων $k \in K$ με $h(k) = h(u)$ (τυχαία μεταβλητή)

Για κάθε $k \in K$ έχουμε δείκτρια τυχαία μεταβλητή $X_k = \begin{cases} 0, & h(k) \neq h(u) \\ 1, & h(k) = h(u) \end{cases}$

Έχουμε $X = \sum_{k \in K} X_k$

Άρα $E[X] = \sum_{k \in K} E[X_k] = \sum_{k \in K} \Pr[X_k = 1] \leq \sum_{k \in K} \frac{1}{m} = \frac{n}{m} \leq 1$

Καθολική Διασπορά

Θεώρημα Έστω ότι επιλέγουμε τυχαία $h \in \mathcal{H}$ και κάνουμε εισαγωγή n κλειδιών σε πίνακα T μεγέθους m . Έστω $i = h(k)$. Τότε

$$\mathbb{E}[n_i] \leq \begin{cases} \alpha, & k \notin T \\ 1 + \alpha, & k \in T \end{cases}$$

Πόρισμα Ο αναμενόμενος χρόνος εκτέλεσης μίας ακολουθίας N πράξεων με $O(m)$ πράξεις εισαγωγής είναι $O(N)$

Συνεπάγεται από το παραπάνω θεώρημα και την παρατήρηση ότι

$$n = O(m) \Rightarrow \alpha = O(1)$$

Μια Καθολική Οικογένεια Συναρτήσεων Διασποράς

p πρώτος αριθμός, μεγαλύτερος από κάθε κλειδί ($p > m$)

$$\mathbf{Z}_p = \{0, 1, \dots, p-1\}$$

$$\mathbf{Z}_p^* = \{1, 2, \dots, p-1\}$$

Για $a \in \mathbf{Z}_p^*$ και $b \in \mathbf{Z}_p$

$$h_{a,b}(k) = [(ak + b) \bmod p] \bmod m$$

Παράδειγμα $p = 17, m = 6$

$$h_{3,4}(8) = [(3 \cdot 8 + 4) \bmod 17] \bmod 6 = (28 \bmod 17) \bmod 6 = 5$$

Μια Καθολική Οικογένεια Συναρτήσεων Διασποράς

p πρώτος αριθμός, μεγαλύτερος από κάθε κλειδί ($p > m$)

$$\mathbf{Z}_p = \{0, 1, \dots, p-1\}$$

$$\mathbf{Z}_p^* = \{1, 2, \dots, p-1\}$$

Για $a \in \mathbf{Z}_p^*$ και $b \in \mathbf{Z}_p$

$$h_{a,b}(k) = [(ak + b) \bmod p] \bmod m$$

$$\mathcal{H}_{p,m} = \{h_{a,b} : a \in \mathbf{Z}_p^*, b \in \mathbf{Z}_p\}$$

Περιλαμβάνει $p(p-1)$ συναρτήσεις

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \longrightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m - 1) \rangle$ μετάθεση του $\langle 0, 1, \dots, m - 1 \rangle$

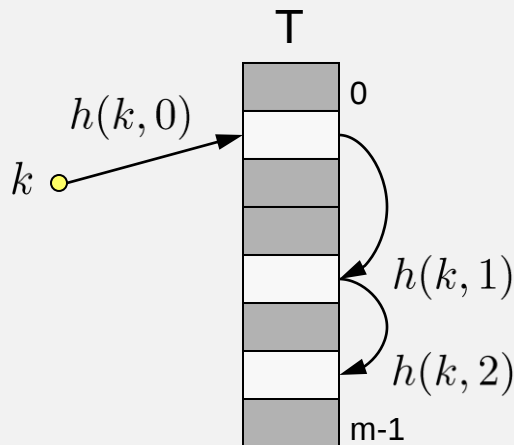
Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \rightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m-1\} \rightarrow \{0, 1, \dots, m-1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m-1) \rangle$ μετάθεση του $\langle 0, 1, \dots, m-1 \rangle$



αναζήτηση (T, k): βολιδοσκοπεί τις θέσεις $T[h(k, i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει το αντικείμενο με κλειδί k ή μία κενή θέση.

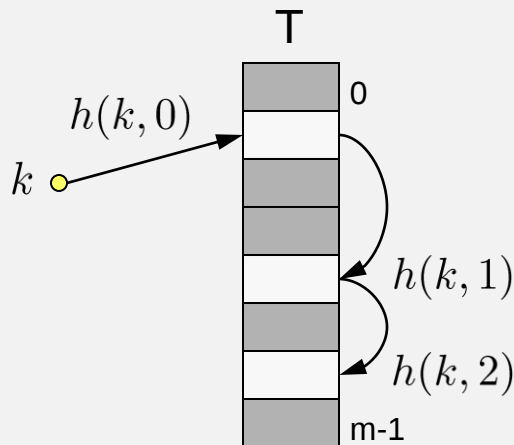
Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \rightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m-1\} \rightarrow \{0, 1, \dots, m-1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m-1) \rangle$ μετάθεση του $\langle 0, 1, \dots, m-1 \rangle$



αναζήτηση (T, k): βολιδοσκοπεί τις θέσεις $T[h(k, i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει το αντικείμενο με κλειδί k ή μία κενή θέση.

εισαγωγή (T, k): βολιδοσκοπεί τις θέσεις $T[h(\text{κλειδί}[x], i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει την πρώτη κενή θέση και τοποθετεί το αντικείμενο εκεί.

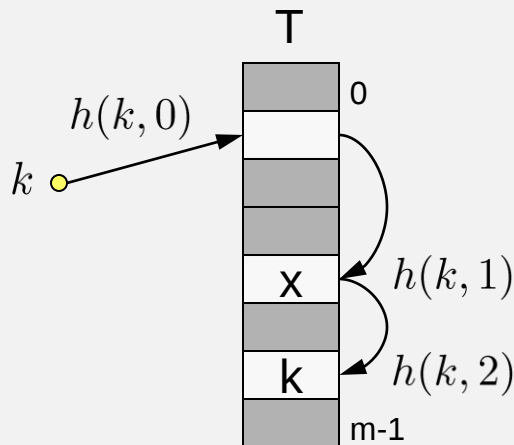
Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \rightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m-1\} \rightarrow \{0, 1, \dots, m-1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m-1) \rangle$ μετάθεση του $\langle 0, 1, \dots, m-1 \rangle$



αναζήτηση (T, k): βολιδοσκοπεί τις θέσεις $T[h(k, i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει το αντικείμενο με κλειδί k ή μία κενή θέση.

εισαγωγή (T, k): βολιδοσκοπεί τις θέσεις $T[h(\text{κλειδί}[x], i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει την πρώτη κενή θέση και τοποθετεί το αντικείμενο εκεί.

διαγραφή (T, x): ?

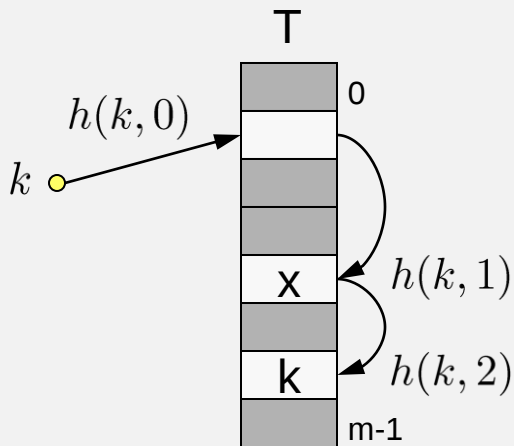
Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \rightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m - 1) \rangle$ μετάθεση του $\langle 0, 1, \dots, m - 1 \rangle$



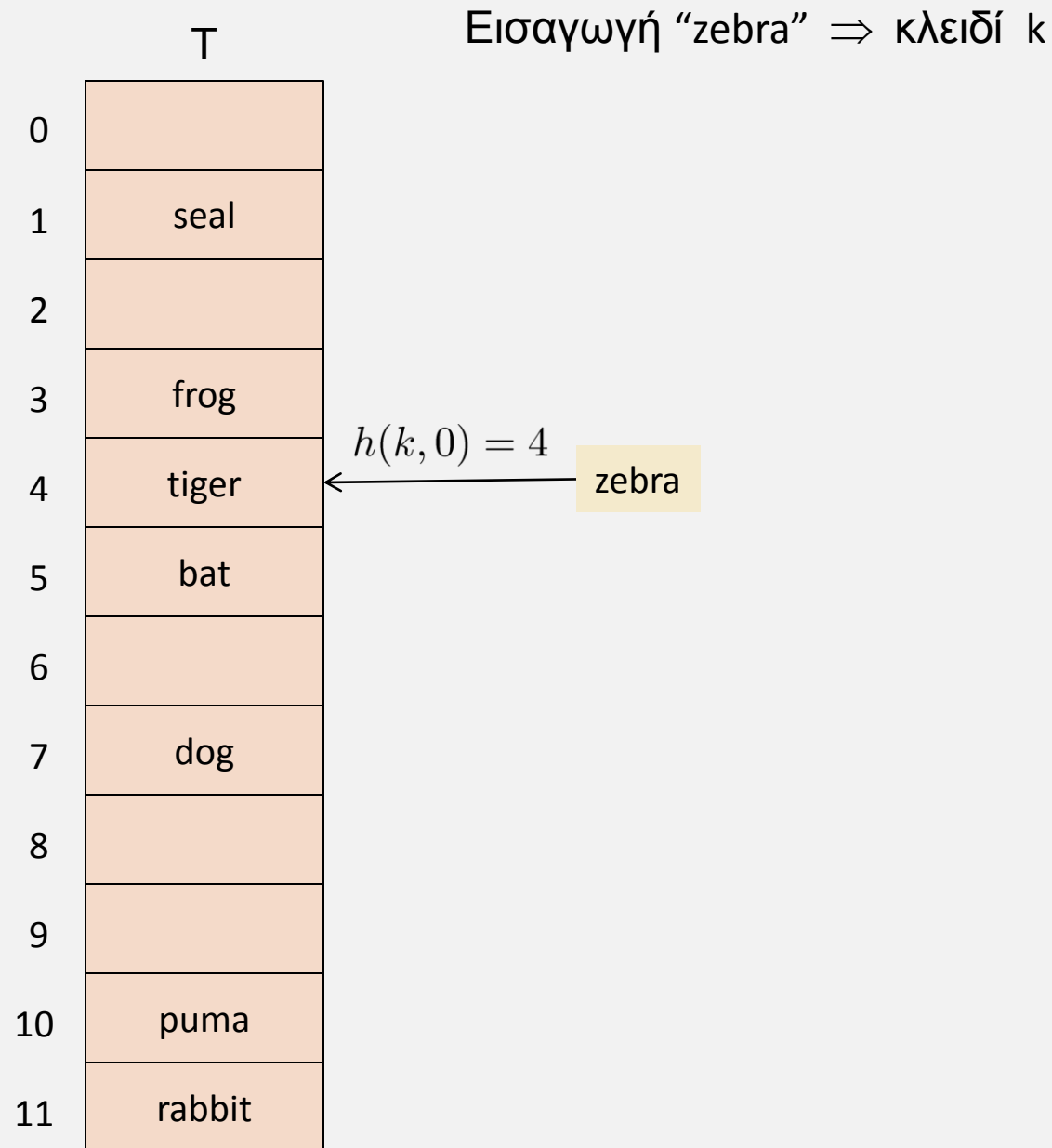
αναζήτηση (T, k): βολιδοσκοπεί τις θέσεις $T[h(k, i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει το αντικείμενο με κλειδί k ή μία κενή θέση.

εισαγωγή (T, k): βολιδοσκοπεί τις θέσεις $T[h(\text{κλειδί}[x], i)]$ για $i=0, 1, \dots, m-1$ μέχρι να βρει την πρώτη κενή θέση και τοποθετεί το αντικείμενο εκεί.

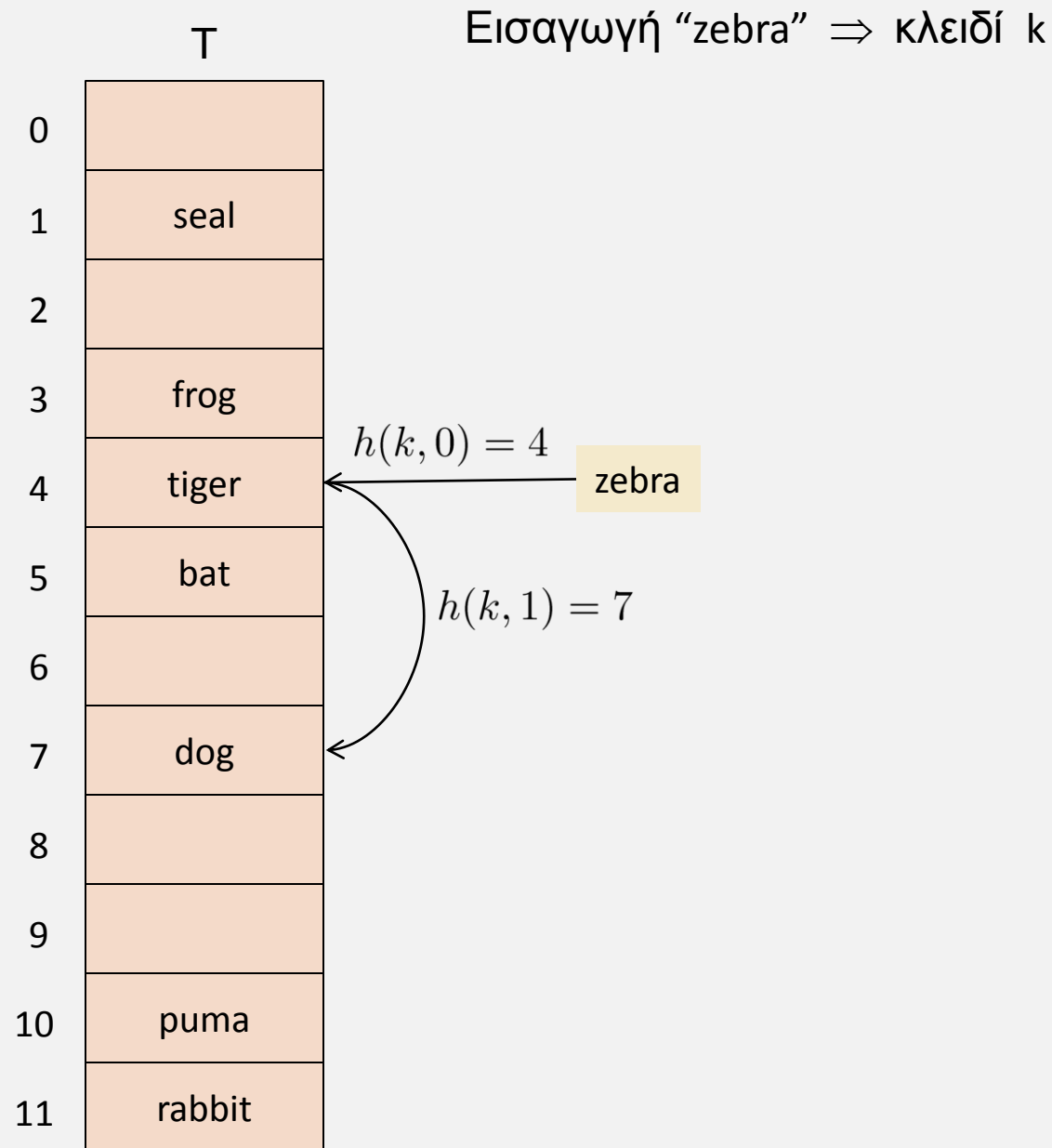
διαγραφή (T, x): σήμανση της θέσης του x ως διαγεγραμμένης. Η αναζήτηση αντιπαράρχεται τέτοιες θέσεις.

Ο χρόνος εκτέλεσης δεν εξαρτάται από τον συντελεστή πληρότητας!

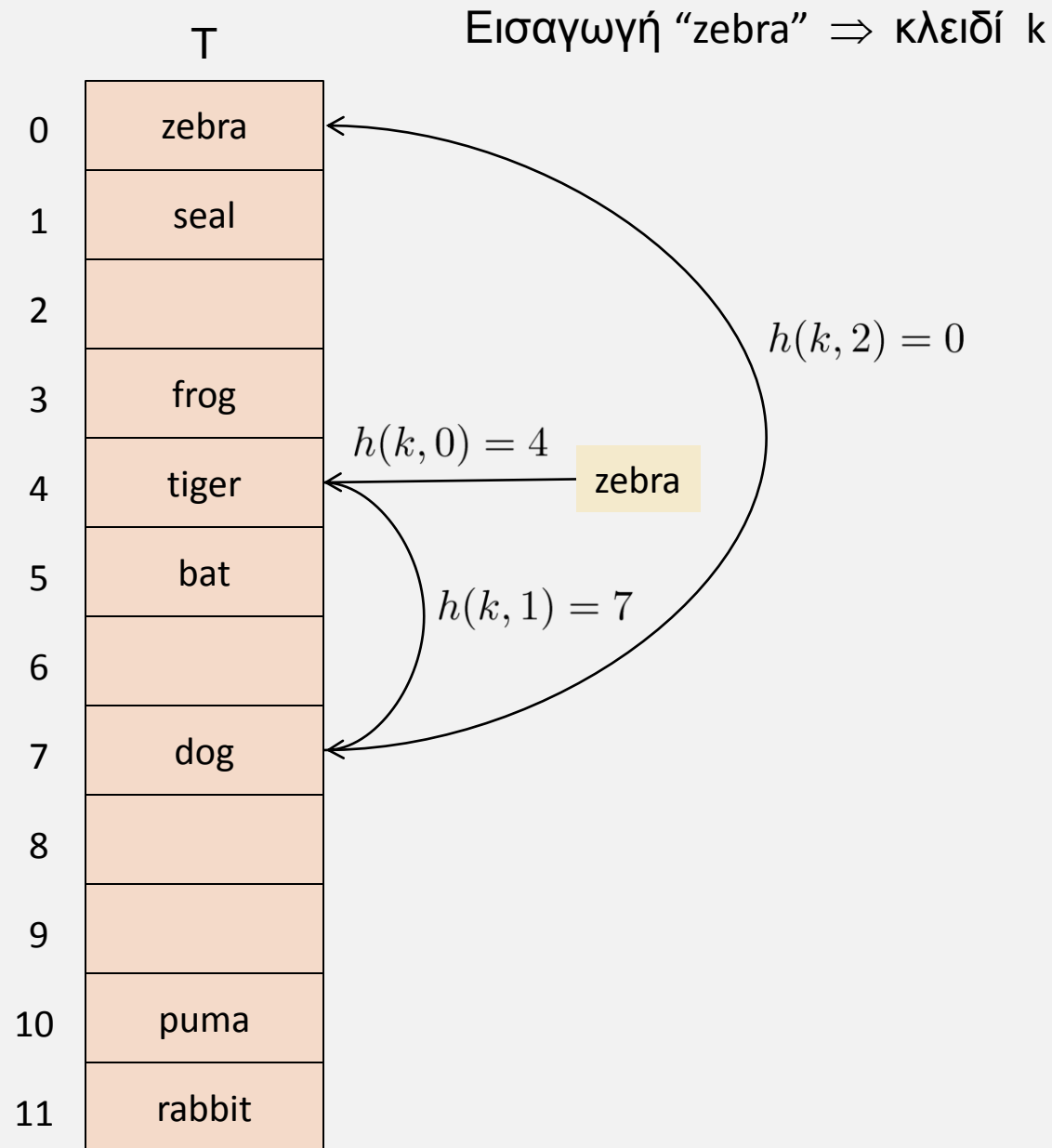
Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)



Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)



Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)



Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \rightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m - 1) \rangle$ μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$

Υπόθεση της **ομοιόμορφης διασποράς** : η βολιδοσκοπική ακολουθία κάθε κλειδιού είναι μία τυχαία μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Η συνάρτηση διασποράς αντιστοιχεί σε κάθε κλειδί μία ακολουθία από διευθύνσεις του πίνακα $T \rightarrow$ **βολιδοσκοπική ακολουθία**

Συνάρτηση Διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

$\langle h(k, 0), h(k, 1), \dots, h(k, m - 1) \rangle$ μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$

Υπόθεση της **ομοιόμορφης διασποράς** : η βολιδοσκοπική ακολουθία κάθε κλειδιού είναι μία τυχαία μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$

Δύσκολο να υλοποιηθεί!

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Γραμμική Βολιδοσκοπήση

$$h(k, i) = (h'(k) + i) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητική συνάρτηση διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Γραμμική Βολιδοσκοπήση

$$h(k, i) = (h'(k) + i) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητική συνάρτηση διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

T	
0	zebra
1	seal
2	
3	frog
4	tiger
5	bat
6	zebra
7	dog
8	
9	
10	puma
11	rabbit

Diagram illustrating the linear probing method. A table T shows the mapping of keys to slots. The keys are: zebra (slot 0), seal (slot 1), frog (slot 3), tiger (slot 4), bat (slot 5), zebra (slot 6), dog (slot 7), puma (slot 10), and rabbit (slot 11). The diagram shows the insertion of the key 'zebra' at slot 6, with arrows indicating the mapping: $h(k, 0) = 4$ (tiger), $h(k, 1) = 5$ (bat), and $h(k, 2) = 6$ (zebra).

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Γραμμική Βολιδοσκοπήση

$$h(k, i) = (h'(k) + i) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητική συνάρτηση διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

- Δίνει μόνο m ακολουθίες
- Αν i διαδοχικές θέσεις είναι κατειλημμένες τότε η πιθανότητα να συμπληρωθεί η επόμενη είναι $(i + 1)/m$

➡ **Πρωτεύουσα ομαδοποίηση** : οι μακριές αλληλουχίες τείνουν να επεκτείνονται

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Τετραγωνική Βολιδοσκοπήση

$$h(k, i) = (h'(k) + c_1 i + c_2 i^2) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητική συνάρτηση διασποράς

$$h : U \times \{0, 1, \dots, m - 1\} \rightarrow \{0, 1, \dots, m - 1\}$$

σταθερές $c_1, c_2 \neq 0$

- Δίνει μόνο m ακολουθίες
- **Δευτερεύουσα ομαδοποίηση** : πιο ήπιας μορφής ομαδοποίηση

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Διπλή Διασπορά

$$h(k, i) = (h_1(k) + ih_2(k)) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητικές συναρτήσεις διασποράς h_1, h_2

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Διπλή Διασπορά

$$h(k, i) = (h_1(k) + ih_2(k)) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητικές συναρτήσεις διασποράς h_1, h_2

$\langle h(k, 0), h(k, 1), \dots, h(k, m - 1) \rangle$ μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$



Πρέπει $h_2(k)$ και m να είναι αμοιβαία πρώτοι ($\gcd(h_2(k), m) = 1$)

π.χ. μπορούμε να έχουμε m ίσο με μία δύναμη του 2 και $h_2(k)$ περιττό

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Διπλή Διασπορά

$$h(k, i) = (h_1(k) + ih_2(k)) \bmod m, \quad i = 0, 1, \dots, m - 1$$

βοηθητικές συναρτήσεις διασποράς h_1, h_2

$\langle h(k, 0), h(k, 1), \dots, h(k, m - 1) \rangle$ μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$



Πρέπει $h_2(k)$ και m να είναι αμοιβαία πρώτοι ($\gcd(h_2(k), m) = 1$)

π.χ. μπορούμε να έχουμε m ίσο με μία δύναμη του 2 και $h_2(k)$ περιττό

- Δίνει m^2 ακολουθίες
- Καλή προσέγγιση της ομοιόμορφης διασποράς

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Υπόθεση της **ομοιόμορφης διασποράς** : η βολιδοσκοπική ακολουθία κάθε κλειδιού είναι μία τυχαία μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$

Συντελεστής πληρότητας $\alpha = n/m$. Για τη μέθοδο των μεταβλητών διευθύνσεων έχουμε $\alpha \leq 1$

Θεώρημα Το αναμενόμενο πλήθος βολιδοσκοπήσεων είναι το πολύ :

$$U = \frac{1}{1 - \alpha} \quad \text{σε μία ανεπιτυχή αναζήτηση}$$

$$S = \frac{1}{\alpha} \ln \frac{1}{1 - \alpha} \quad \text{σε μία επιτυχή αναζήτηση}$$

- Π.χ.
- Για $\alpha=0.5$ έχουμε $U=2$ και $S=1.386$
 - Για $\alpha=0.9$ έχουμε $U=10$ και $S=2.558$

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Υπόθεση της **ομοιόμορφης διασποράς** : η βολιδοσκοπική ακολουθία κάθε κλειδιού είναι μία τυχαία μετάθεση της $\langle 0, 1, \dots, m - 1 \rangle$

Αν ισχύει η υπόθεση ομοιόμορφης διασποράς, τότε η πιθανότητα η j -οστή βολιδοσκόπηση να βρει κατειλημμένη θέση δεδομένου ότι οι πρώτες $(j-1)$ πράξεις βολιδοσκόπησης βρήκαν κατειλημμένες θέσεις είναι $\frac{n - (j - 1)}{m - (j - 1)}$

Άρα η πιθανότητα να χρειαστούν τουλάχιστον i βολιδοσκοπήσεις είναι

$$\frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \frac{n-2}{m-2} \cdots \frac{n-i+2}{m-i+2} \leq \left(\frac{n}{m}\right)^{i-1} = \alpha^{i-1}$$

Επομένως, το αναμενόμενο πλήθος βολιδοσκοπήσεων σε ανεπιτυχή αναζήτηση είναι

$$U \leq \sum_{i=1}^{\infty} \alpha^{i-1} = 1 + \alpha + \alpha^2 + \alpha^3 + \dots = \frac{1}{1 - \alpha}$$

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Υπόθεση της **ομοιόμορφης διασποράς** : η βολιδοσκοπική ακολουθία κάθε κλειδιού είναι μία τυχαία μετάθεση της $\langle 0, 1, \dots, m-1 \rangle$

Το αναμενόμενο πλήθος προσπελάσεων σε επιτυχημένη αναζήτηση είναι ίσο με το αναμενόμενο πλήθος προσπελάσεων για την εισαγωγή κάθε ενός από τα n κλειδιά.

Το αναμενόμενο πλήθος προσπελάσεων για την εισαγωγή του i -οστού κλειδιού είναι ίσο με το αναμενόμενο πλήθος προσπελάσεων σε μη επιτυχημένη αναζήτηση όταν έχουν ήδη εισαχθεί $(i-1)$ κλειδιά. Επομένως:

$$S \leq \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{1 - i/m} = \frac{m}{n} \sum_{i=0}^{n-1} \frac{1}{m - i} = \frac{m}{n} (H_m - H_{m-n})$$

$$H_N = 1 + 1/2 + 1/3 + \dots + 1/N \approx \ln N \quad \text{N-οστός αρμονικός αριθμός}$$

Μέθοδος Μεταβλητών Διευθύνσεων (Ανοιχτή Διευθυνσιοδότηση)

Υπόθεση της **ομοιόμορφης διασποράς** : η βολιδοσκοπική ακολουθία κάθε κλειδιού είναι μία τυχαία μετάθεση της $\langle 0, 1, \dots, m-1 \rangle$

Το αναμενόμενο πλήθος προσπελάσεων σε επιτυχημένη αναζήτηση είναι ίσο με το αναμενόμενο πλήθος προσπελάσεων για την εισαγωγή κάθε ενός από τα n κλειδιά.

Το αναμενόμενο πλήθος προσπελάσεων για την εισαγωγή του i -οστού κλειδιού είναι ίσο με το αναμενόμενο πλήθος προσπελάσεων σε μη επιτυχημένη αναζήτηση όταν έχουν ήδη εισαχθεί $(i-1)$ κλειδιά. Επομένως:

$$\begin{aligned} S &\leq \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{1 - i/m} = \frac{m}{n} \sum_{i=0}^{n-1} \frac{1}{m - i} = \frac{m}{n} (H_m - H_{m-n}) \\ &\approx \frac{m}{n} (\ln(m) - \ln(m-n)) = \frac{m}{n} \ln \frac{m}{m-n} = \frac{1}{\alpha} \ln \frac{1}{1-\alpha} \end{aligned}$$

Ταξινομημένοι Πίνακες Διασποράς

Αν τα κλειδιά είναι ταξινομημένα έχουμε μείωση του χρόνου για αποτυχημένες αναζητήσεις: Αν ένα μεγαλύτερο κλειδί από το ζητούμενο k προσπελαστεί τερματίζει η αναζήτηση.

Μέθοδος με Αλυσίδες

Διατηρούμε τα κλειδιά στις αλυσίδες ταξινομημένα.

Μέθοδος Ανοικτής Διευθυνσιοδότησης

Τα κλειδιά πρέπει να εισαχθούν έτσι ώστε τα κλειδιά που προηγούνται στην ακολουθία αναζήτησης από το k , θα πρέπει να είναι μικρότερα από το k .

Ιδέα

Αν στην ακολουθία αναζήτησης για το κλειδί k δούμε κλειδί $k' > k$, τότε αντικαθιστούμε το k' με το k και συνεχίζουμε με την εισαγωγή του k' βάσει της ακολουθίας αναζήτησης του k' .

Ταξινομημένοι Πίνακες Διασποράς

Η τελική μορφή του πίνακα διασποράς μετά τις εισαγωγές των κλειδιών είναι η ίδια ανεξάρτητα από τη σειρά εισαγωγής των κλειδιών.

Υπόθεση

Η πιθανότητα να επιλεγεί ένα συγκεκριμένο κλειδί από τον χώρο κλειδιών είναι η ίδια για όλα τα κλειδιά του χώρου.

Τότε:

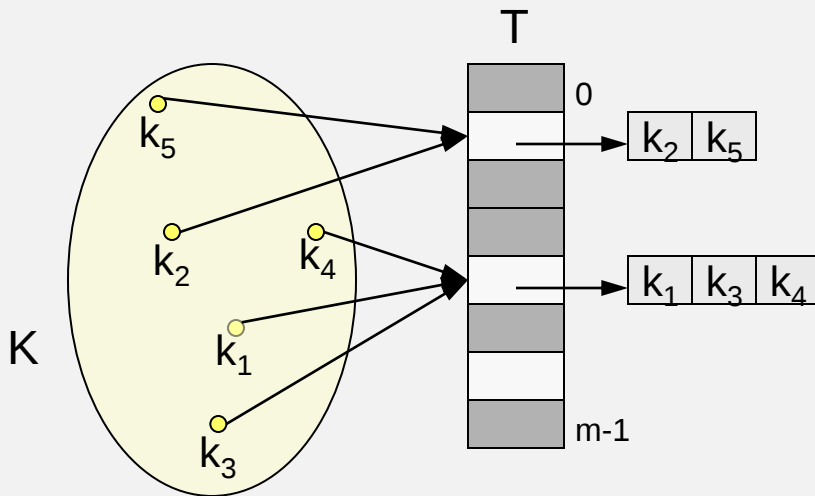
- Ο αναμενόμενος χρόνος για επιτυχημένη αναζήτηση δεν αλλάζει.
- Ο αναμενόμενος χρόνος για μη επιτυχημένη αναζήτηση γίνεται περίπου ίδιος με το χρόνο επιτυχημένης αναζήτησης.

Πλήρης Διασπορά

Όταν το σύνολο των κλειδιών είναι **στατικό** μπορούμε να πετύχουμε άριστη επίδοση : $O(1)$ χρόνο χειρότερης περίπτωσης ανά πράξη

Πλήρης Διασπορά

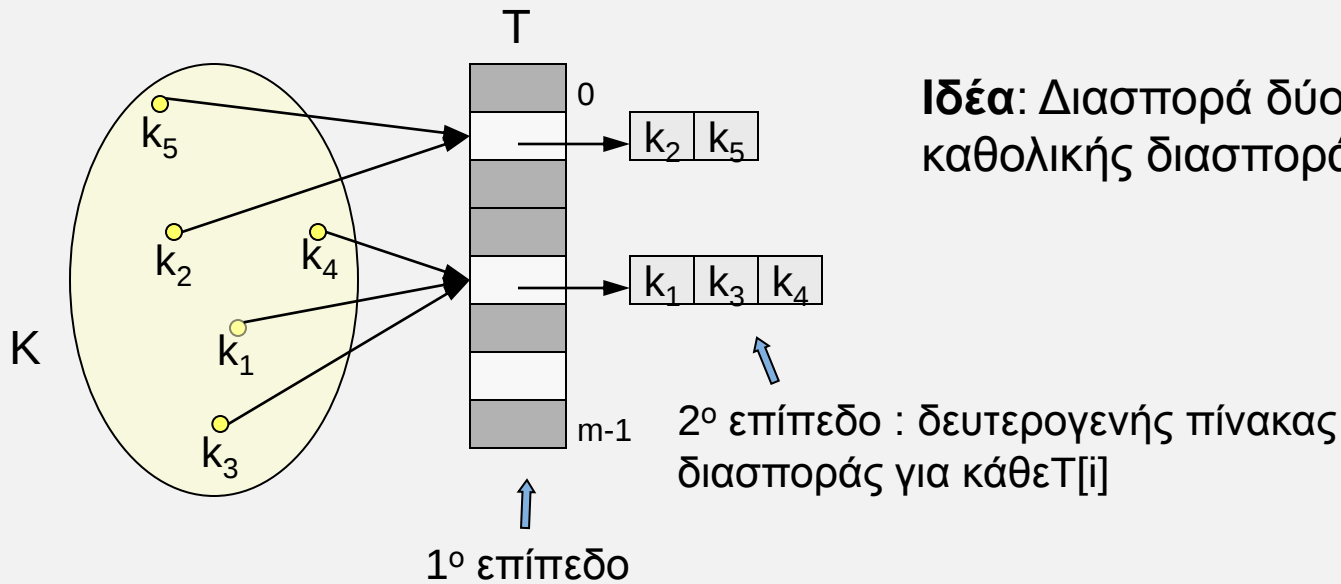
Όταν το σύνολο των κλειδιών είναι **στατικό** μπορούμε να πετύχουμε άριστη επίδοση : $O(1)$ χρόνο χειρότερης περίπτωσης ανά πράξη



Ιδέα: Διασπορά δύο επιπέδων με χρήση καθολικής διασποράς ανά επίπεδο

Πλήρης Διασπορά

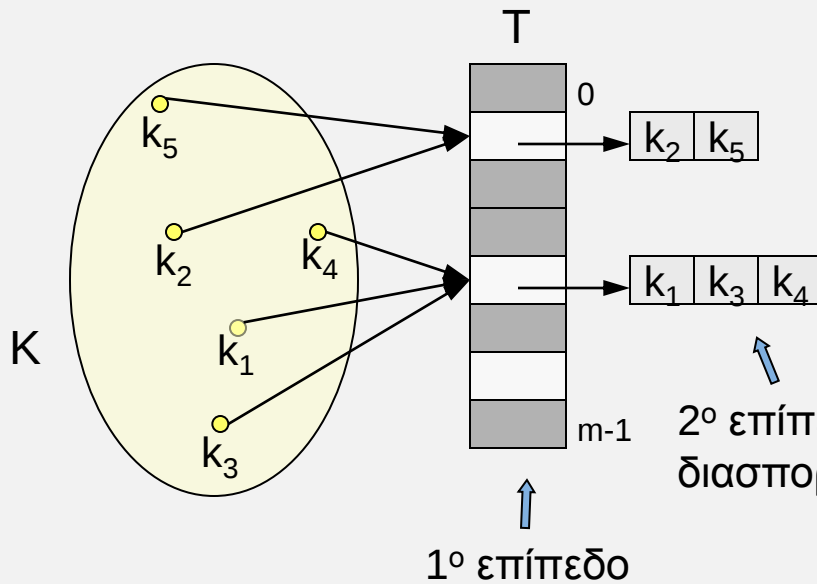
Όταν το σύνολο των κλειδιών είναι **στατικό** μπορούμε να πετύχουμε άριστη επίδοση : $O(1)$ χρόνο χειρότερης περίπτωσης ανά πράξη



Ιδέα: Διασπορά δύο επιπέδων με χρήση καθολικής διασποράς ανά επίπεδο

Πλήρης Διασπορά

Όταν το σύνολο των κλειδιών είναι **στατικό** μπορούμε να πετύχουμε άριστη επίδοση : $O(1)$ χρόνο χειρότερης περίπτωσης ανά πράξη



Ιδέα: Διασπορά δύο επιπέδων με χρήση καθολικής διασποράς ανά επίπεδο

Στο 2^ο επίπεδο αποφεύγουμε τις συμπτώσεις

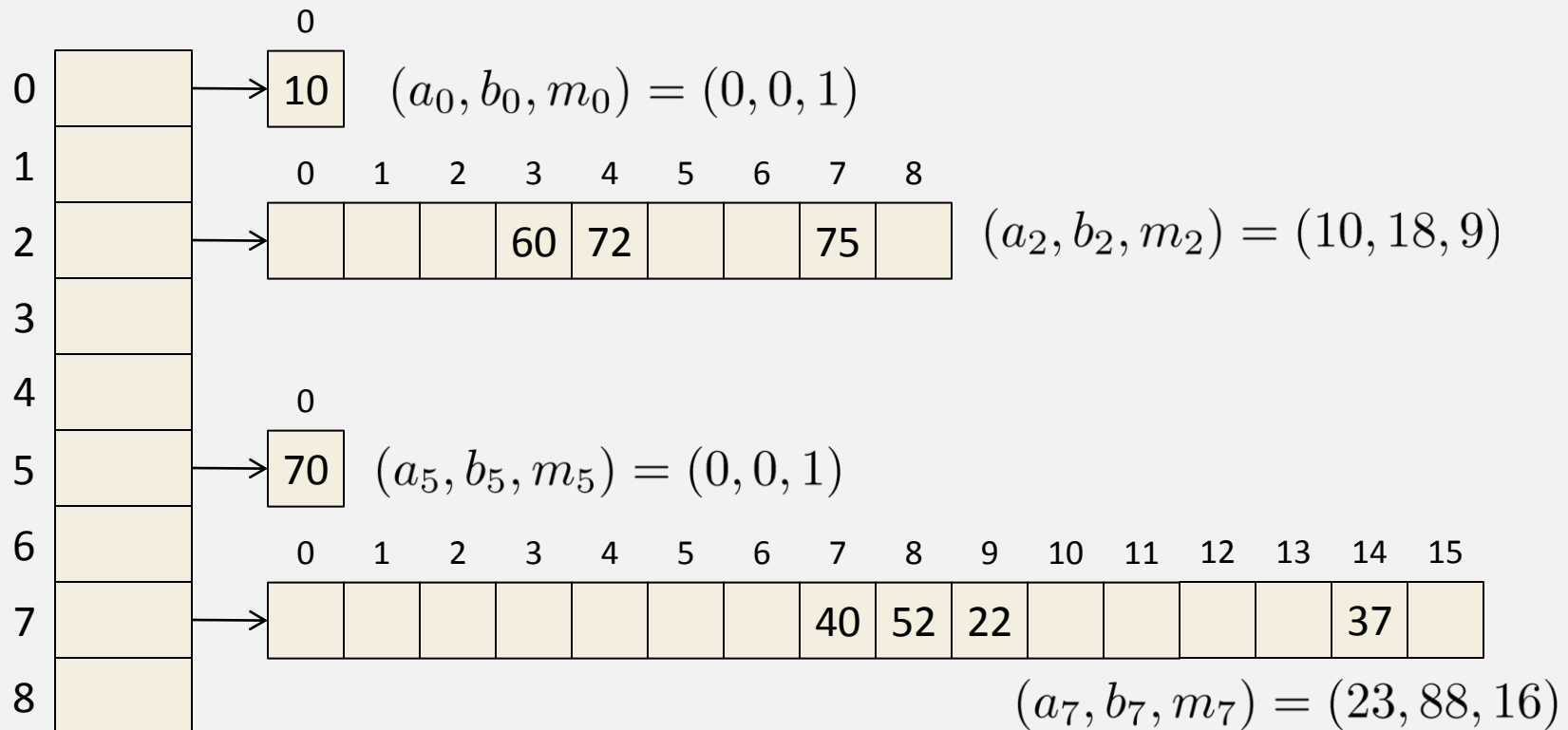
Πλήρης Διασπορά

Παράδειγμα $K = \{10, 22, 37, 40, 52, 60, 70, 72, 75\}$

Συνάρτηση διασποράς 1ου επιπέδου : $h(k) = [(a \cdot k + b) \bmod p] \bmod m$

$$(a, b, p, m) = (3, 42, 101, 9)$$

Συνάρτηση διασποράς 2ου επιπέδου : $h_j(k) = [(a_j \cdot k + b_j) \bmod p] \bmod m_j$



Πλήρης Διασπορά

$T =$ πρωτογενής πίνακας διασποράς με m θέσεις και συνάρτηση διασποράς h

$S_j =$ δευτερογενής πίνακας διασποράς με m_j θέσεις και συνάρτηση διασποράς h_j
 $j = 0, 1, \dots, m$

Έστω n ο συνολικός αριθμός κλειδιών και n_j ο αριθμός των κλειδιών που αποθηκεύονται στον S_j . $(\sum_{j=0}^{m-1} n_j = n)$

Για να αποφύγουμε τις συγκρούσεις στο δεύτερο επίπεδο θα επιλέξουμε $m_j = n_j^2$.

Θα δείξουμε ότι για κατάλληλα επιλεγμένη h ο αναμενόμενος χώρος που απαιτείται για τους δευτερογενείς πίνακες είναι

$$\mathbf{E}\left[\sum_{j=0}^{m-1} m_j\right] = O(n)$$

Πλήρης Διασπορά

Επιλέγουμε την πρωτογενή συνάρτηση διασποράς h από την καθολική οικογένεια $\mathcal{H}_{p,m}$.

Επιλέγουμε κάθε δευτερογενή συνάρτηση διασποράς h_j από την καθολική οικογένεια $\mathcal{H}_{p,m_j}$.

Πλήρης Διασπορά

Θεώρημα Έστω ότι επιλέγουμε τυχαία μια συνάρτηση διασποράς h από καθολική οικογένεια συναρτήσεων διασποράς. Αν αποθηκεύσουμε n κλειδιά και το μέγεθος του πίνακα διασποράς είναι $m = n^2$ τότε η πιθανότητα να υπάρχει σύγκρουση είναι $< 1/2$.

Ο αριθμός των πιθανών συγκρούσεων είναι $\binom{n}{2}$, όσα τα διαφορετικά ζεύγη κλειδιών.

Από τον τρόπο επιλογής της h προκύπτει ότι η πιθανότητα σύγκρουσης ενός ζεύγους κλειδιών είναι το πολύ $1/m$.

Άρα ο αναμενόμενος αριθμός συγκρούσεων X είναι

$$\mathbf{E}[X] \leq \frac{\binom{n}{2}}{m} = \frac{n(n-1)}{2n^2} < \frac{1}{2}$$

Από την ανισότητα του Markov έχουμε

$$\mathbf{Pr}[X \geq t] \leq \frac{\mathbf{E}[X]}{t} \Rightarrow \mathbf{Pr}[X \geq 1] < \frac{1}{2}$$

Πλήρης Διασπορά

Θεώρημα Έστω ότι επιλέγουμε τυχαία μια συνάρτηση διασποράς h από καθολική οικογένεια συναρτήσεων διασποράς. Αν αποθηκεύσουμε n κλειδιά και το μέγεθος του πίνακα διασποράς είναι $m = n^2$ τότε η πιθανότητα να υπάρχει σύγκρουση είναι $< 1/2$.

Το παραπάνω θεώρημα συνεπάγεται ότι για $m_j = n_j^2$ μπορούμε να βρούμε με λίγες δοκιμές μια συνάρτηση διασποράς h_j που να μην έχει συγκρούσεις για το σύνολο των κλειδιών που αποθηκεύεται στον S_j .

Πλήρης Διασπορά

Απομένει να επιλέξουμε μια πρωτογενή συνάρτηση διασποράς h έτσι ώστε

$$\mathbf{E}\left[\sum_{j=0}^{m-1} m_j\right] = \mathbf{E}\left[\sum_{j=0}^{m-1} n_j^2\right] = O(n)$$

Θεώρημα Έστω ότι επιλέγουμε τυχαία μια συνάρτηση διασποράς h από καθολική οικογένεια συναρτήσεων διασποράς. Αν αποθηκεύσουμε n κλειδιά και το μέγεθος του πίνακα διασποράς είναι $m = n$ τότε

$$\mathbf{E}\left[\sum_{j=0}^{m-1} n_j^2\right] < 2n$$

Πλήρης Διασπορά

Θεώρημα Έστω ότι επιλέγουμε τυχαία μια συνάρτηση διασποράς h από καθολική οικογένεια συναρτήσεων διασποράς. Αν αποθηκεύσουμε n κλειδιά και το μέγεθος του πίνακα διασποράς είναι $m = n$ τότε

$$\mathbf{E}\left[\sum_{j=0}^{m-1} n_j^2\right] < 2n$$

Για κάθε μη αρνητικό ακέραιο a ισχύει $a^2 = a + 2\binom{a}{2}$ άρα έχουμε

$$\mathbf{E}\left[\sum_{j=0}^{m-1} n_j^2\right] = \mathbf{E}\left[\sum_{j=0}^{m-1} \left(n_j + 2\binom{n_j}{2}\right)\right] = \mathbf{E}\left[\sum_{j=0}^{m-1} n_j\right] + 2 \cdot \mathbf{E}\left[\sum_{j=0}^{m-1} \binom{n_j}{2}\right]$$

Καθώς $\mathbf{E}\left[\sum_{j=0}^{m-1} n_j\right] = n$ απομένει να δείξουμε ότι $\mathbf{E}\left[\sum_{j=0}^{m-1} \binom{n_j}{2}\right] < \frac{n}{2}$

Πλήρης Διασπορά

Θεώρημα Έστω ότι επιλέγουμε τυχαία μια συνάρτηση διασποράς h από καθολική οικογένεια συναρτήσεων διασποράς. Αν αποθηκεύσουμε n κλειδιά και το μέγεθος του πίνακα διασποράς είναι $m = n$ τότε

$$\mathbf{E}\left[\sum_{j=0}^{m-1} n_j^2\right] < 2n$$

Παρατηρούμε ότι $\mathbf{E}\left[\sum_{j=0}^{m-1} \binom{n_j}{2}\right] =$ αναμενόμενο πλήθος συγκρούσεων, άρα από

την επιλογή της h έχουμε $\mathbf{E}\left[\sum_{j=0}^{m-1} \binom{n_j}{2}\right] \leq \frac{\binom{n}{2}}{m} = \frac{n(n-1)}{2n} < \frac{n}{2}$

Πλήρης Διασπορά

Θεώρημα Έστω ότι επιλέγουμε τυχαία μια συνάρτηση διασποράς h από καθολική οικογένεια συναρτήσεων διασποράς. Αν αποθηκεύσουμε n κλειδιά και το μέγεθος του πίνακα διασποράς είναι $m = n$ τότε

$$\mathbf{E}\left[\sum_{j=0}^{m-1} n_j^2\right] < 2n$$

Άρα ο αναμενόμενος χώρος που απαιτείται για την αποθήκευση όλων των δευτερογενών πινάκων είναι

$$\mathbf{E}\left[\sum_{j=0}^{m-1} m_j\right] < 2n$$

Από την ανισότητα του Markov έχουμε

$$\mathbf{Pr}\left[\sum_{j=0}^{m-1} m_j \geq 4n\right] \leq \frac{\mathbf{E}\left[\sum_{j=0}^{m-1} m_j\right]}{4n} < \frac{1}{2}$$

Δηλαδή μπορούμε να βρούμε με λίγες δοκιμές μια κατάλληλη πρωτογενή συνάρτηση διασποράς.

Δυναμικοί Πίνακες Διασποράς

Συντελεστής πληρότητας $\alpha = n/m$

Η απόδοση του κατακερματισμού επιδεινώνεται όσο αυξάνει ο συντελεστής α

Επιπλέον στην μέθοδο των μεταβλητών διευθύνσεων πρέπει $\alpha < 1$

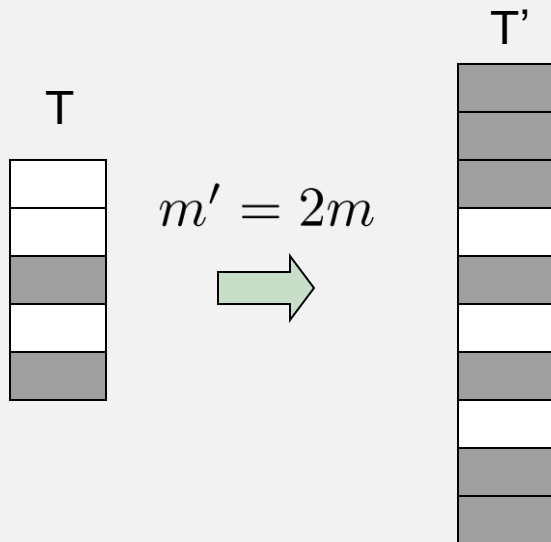
Δυναμικοί Πίνακες Διασποράς

Συντελεστής πληρότητας $\alpha = n/m$

Η απόδοση του κατακερματισμού επιδεινώνεται όσο αυξάνει ο συντελεστής α

Επιπλέον στην μέθοδο των μεταβλητών διευθύνσεων πρέπει $\alpha < 1$

Για να αντιμετωπίσουμε το παραπάνω ζήτημα μπορούμε να **διπλασιάσουμε** το μέγεθος του πίνακα διασποράς κάθε φορά που ο συντελεστής α φθάνει μία συγκεκριμένη τιμή (π.χ. 50% για τη μέθοδο των μεταβλητών διευθύνσεων)



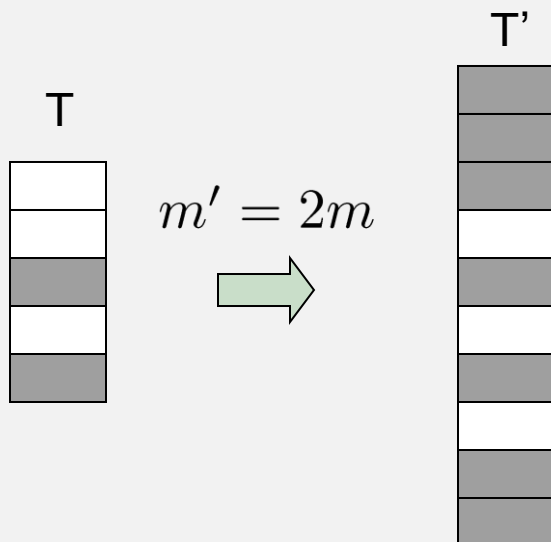
Δυναμικοί Πίνακες Διασποράς

Συντελεστής πληρότητας $\alpha = n/m$

Η απόδοση του κατακερματισμού επιδεινώνεται όσο αυξάνει ο συντελεστής α

Επιπλέον στην μέθοδο των μεταβλητών διευθύνσεων πρέπει $\alpha < 1$

Για να αντιμετωπίσουμε το παραπάνω ζήτημα μπορούμε να **διπλασιάσουμε** το μέγεθος του πίνακα διασποράς κάθε φορά που ο συντελεστής α φθάνει μία συγκεκριμένη τιμή (π.χ. 50% για τη μέθοδο των μεταβλητών διευθύνσεων)



Παίρνει $O(n)$ (αναμενόμενο) χρόνο γιατί όλα τα στοιχεία πρέπει να εισαχθούν στο νέο πίνακα.

Ευτυχώς ο διπλασιασμός δε συμβαίνει συχνά,
⇒ αποτελεί καλή λύση όταν μας ενδιαφέρει να είναι χαμηλό το **μέσο κόστος ανά πράξη** (σε αντίθεση με το κόστος χειρότερης περίπτωσης)