1. Supponete di effettuare una ricerca in una lista concatenata di lunghezza n, dove ogni elemento contiene una chiave k ed un valore hash $h(k)$. Ogni chiave é una lunga stringa di caratteri. Come potreste trarre vantaggio dai valori hash quando ricercate nella lista un elemento con una data chiave? Proporre una procedura che effettui la ricerca di una chiave in maniera efficiente.

Soluzione

Supponiamo di avere una lista concatenata in cui ogni elemento, x, contiene una chiave $key[x] = [0, 1, ..., n-1]$, consistente in una lunga stringa di caratteri ed una campo valore hash denominato $hash[x]$, associato alla chiave $key[x]$.

Il valore hash puó essere utilizzato per ricercare in modo efficiente una chiave k all'interno della lista concatenata. Dato che $x = y \Rightarrow hash[x] = hash[y]$ possiamo confrontare esclusivamente le stringhe degli elementi x per cui $hash[x] = hash[k]$.

```
HASH-LIST-SEARCH(L)
    h ← hash(k)
    x ← Head[L]
    while (x ≠ NULL) do
        if (hash[x] = h) then
            if key[x] = k then
                return x
        x ← next[x]
    return NULL
```

2. Considerate una variante del metodo della divisione i cui $h(k) = k \bmod m$, dove $m = 2^p - 1$ e k é una stringa di caratteri interpretata come un numero con base 2^p. Dimostrate che, se la stringa x puó essere ottenuta dalla stringa y permutando i suoi caratteri, allora le stringhe x e y sono indirizzate nella medesima cella. Indicate un'applicazione nella cui funzione hash non sarebbe desiderabile questa proprietá.

Soluzione
Si supponga che una chiave k sia una stringa di n caratteri interpretata come un numero in base 2^p. Generallizando possiamo porre $k = [c_1, c_2, c_3, ..., c_n]$. Il valore numerico, V_k, a cui la chiave k é associato viene dato dalla formula

$$\begin{aligned} V_k = \quad & c_1 \cdot (2^p)^{n-1} + c_2 \cdot (2^p)^{n-2} + c_3 \cdot (2^p)^{n-3} + ... + c_{n-1} \cdot (2^p) + c_n \\ & c_1 \cdot 2^{p(n-1)} + c_2 \cdot 2^{p(n-2)} + c_3 \cdot 2^{p(n-3)} + ... + c_{n-1} \cdot 2^p + c_n \end{aligned}$$

Si puó facilmente dimostrare che, $\forall i > 0$, $2^{pi} \bmod 2^p - 1 = 1$. Questo infatti si riduce a dimostrare che il valore $2^{ip} - 1$ é divisibile per $2^p - 1$. Applicando la proprietá delle serie geometriche si ottinene infatti

$$\frac{(2^p)^i - 1}{2^p - 1} = \sum_{i=0}^{i-1} (2^p)^i = 1 + 2^p + 2^{2p} + ... + 2^{p(i-1)}$$

che é chiaramente un valore intero.
Infine applicando le due seguenti proprietá del modulo

1) $(a + b) \bmod c = ((a \bmod c) + (b \bmod c)) \bmod c$
2) $a \bmod c = b \Rightarrow xa \bmod c = xb \bmod c$

si ottine la sueguente relazione che dimostra la tesi

$$\begin{aligned} & V_k \bmod m = \\ & ((c_1 \cdot 2^{p(n-1)} \bmod m) + ... + (c_{n-1} \cdot 2^p \bmod m) + (c_n \bmod m)) \bmod m = \\ & ((c_1 \bmod m) + (c_2 \bmod m) + ... + (c_{n-1} \bmod m) + (c_n \bmod m)) \bmod m = \\ & (c_1 + c_2 + ... + c_{n-1} + c_n) \bmod m \end{aligned}$$

3. Considerate una tavola hash di dimensione $m = 1000$ e una corrispondente funzione hash $h(k) = \lfloor m(kA \mod 1) \rfloor$ per $A = (\sqrt{5} - 1)/2$. Calcolate le celle in cui saranno mandate le chiavi $61, 62, 63, 64$ e 65.

Soluzione

$$(\sqrt{5} - 1)/2 = (2,236 - 1)/2 = 1,236/2 = 0,618$$

$$\begin{aligned}
h(61) &= \lfloor 1000(61 \cdot 0,618 \mod 1) \rfloor \\
&= \lfloor 1000(37,698 \mod 1) \rfloor \\
&= \lfloor 1000(0,698) \rfloor \\
&= 698
\end{aligned}$$

$$\begin{aligned}
h(62) &= \lfloor 1000(62 \cdot 0,618 \mod 1) \rfloor \\
&= \lfloor 1000(38,316 \mod 1) \rfloor \\
&= \lfloor 1000(0,316) \rfloor \\
&= 316
\end{aligned}$$

$$\begin{aligned}
h(63) &= \lfloor 1000(63 \cdot 0,618 \mod 1) \rfloor \\
&= \lfloor 1000(38,934 \mod 1) \rfloor \\
&= \lfloor 1000(0,934) \rfloor \\
&= 934
\end{aligned}$$

$$\begin{aligned}
h(64) &= \lfloor 1000(64 \cdot 0,618 \mod 1) \rfloor \\
&= \lfloor 1000(39,552 \mod 1) \rfloor \\
&= \lfloor 1000(0,552) \rfloor \\
&= 552
\end{aligned}$$

$$\begin{aligned}
h(65) &= \lfloor 1000(65 \cdot 0,618 \mod 1) \rfloor \\
&= \lfloor 1000(40,17 \mod 1) \rfloor \\
&= \lfloor 1000(0,17) \rfloor \\
&= 170
\end{aligned}$$

4. Supponete di inserire le chiavi $10, 22, 31, 4, 15, 28, 17, 88, 59$ in una tavola hash di lunghezza $m = 11$ utilizzando l'indirizzamento aperto con la funzione hash ausiliaria $h'(k) = k \bmod m$. Illustrate il risultato dell'inserimento di queste chiavi utilizzando l'ispezione lineare, l'ispezione quadratica con $c_1 = 1$ e $c_2 = 3$ e il doppio hashing con $h_2(k) = 1 + (k \bmod (m-1))$.

Soluzione

- Scansione lineare
  $h(10) \rightarrow [10]$
  $h(22) \rightarrow [0]$
  $h(31) \rightarrow [9]$
  $h(4) \rightarrow [4]$
  $h(15) \rightarrow [4] \rightarrow [5]$
  $h(28) \rightarrow [6]$
  $h(17) \rightarrow [6] \rightarrow [7]$
  $h(88) \rightarrow [0] \rightarrow [1]$
  $h(59) \rightarrow [4] \rightarrow [5] \rightarrow [6] \rightarrow [7] \rightarrow [8]$

- Scansione quadratica
  $h(10) \rightarrow [10]$
  $h(22) \rightarrow [0]$
  $h(31) \rightarrow [9]$
  $h(4) \rightarrow [4]$
  $h(15) \rightarrow [4] \rightarrow [8]$
  $h(28) \rightarrow [6]$
  $h(17) \rightarrow [6] \rightarrow [10] \rightarrow [9] \rightarrow [3]$
  $h(88) \rightarrow [0] \rightarrow [4] \rightarrow [3]$
  $h(59) \rightarrow [4] \rightarrow [8] \rightarrow [7]$

- Hashing doppio
  $h(10) \rightarrow [10]$
  $h(22) \rightarrow [0]$
  $h(31) \rightarrow [9]$
  $h(4) \rightarrow [4]$
  $h(15) \rightarrow [4] \rightarrow [10] \rightarrow [5]$
  $h(28) \rightarrow [6]$
  $h(17) \rightarrow [6] \rightarrow [3]$
  $h(88) \rightarrow [0] \rightarrow [9] \rightarrow [7]$
  $h(59) \rightarrow [4] \rightarrow [3] \rightarrow [2]$

5. Supponete di applicare la tecnica del doppio hashing per risolvere le collisioni, ovvero utilizzare una funzione hash $h(k, i) = (h_1(k) + ih_2(k)) \bmod m$. Dimostrate che se m ed $h_2(k)$ hanno il massimo comune divisore $d \geq 1$ per qualche chiave k, allora una ricerca senza successo della chiave k esamina soltanto una frazione $1/d$ della tavola hash prima di ritornare alla cella iniziale $h_1(k)$. Pertanto se $d = 1$, e quindi m e $h_2(k)$ sono numeri relativamente primi, la ricerca puó esaminare l'intera tavola hash.

Soluzione
L'aritmetica modulare si basa sul concetto di congruenza modulo n. Dati tre numeri interi a, b, n, con $n \neq 0$, diciamo che a e b sono congruenti modulo n se la loro differenza $(a - b)$ é un multiplo di n. In questo caso scriviamo

$$a \equiv b \bmod n$$

Per cui se ($a \bmod n = b$) allora ($a \equiv b \bmod n$).

Se $h_2(k)$ ed m fossero primi tra loro la scansione che utilizza la funzone hash $(h(k, i) = h_1(k) + ih_2(k)) \bmod m$ ispezionerebbe tutte le caselle della nostra tabella. Questo viene dimostrato facilmente nel seguente modo.

Siano i e j due distinti valori per cui si abbia $ih_2(k) \bmod m = jh_2(k) \bmod m = a$. Allora $(ih_2(k) \equiv a \bmod m)$ ed anche $(jh_2(k) \equiv a \bmod m)$. Ne segue per una proprietá delle congruenze che $ih_2(k) \equiv jh_2(k) \bmod m$, da cui $i \equiv j \bmod m$.
Questo implica che $(i-j)$ é un multiplo di m, per cui devono essere ispezionate almeno m caselle prima di ottenere una posizione uguale.

Se $h_2(k)$ e m avessero un massimo comune divisore $d \geq 1$ allora potremmo porre $h_2(k) = c' \cdot d$ ed $m = c'' \cdot d$. Da questo si avrebbe

$$i \cdot h_2(k) \bmod m = (i \cdot d \cdot c') \bmod (c'' \cdot d) = (i \cdot c') \bmod c''$$

e la distanza tra due valori uguali nella scanzione si ridurrebbe a $c'' = m/d$.

6. Supponete di applicare la tecnica della scansione quadratica risolvere le collisioni, ovvero utilizzare una funzione hash $h(k, i) = (h_1(k) + c_1 i + c_2 i^2) \bmod m$. Trovare i possibili valori di c_1, c_2 ed m in modo che la scansione controlli l'intera tabella.

Soluzione
Basandoci sul concetto di congruenza modulo n, utilizzata in uno degli esercizi precedenti, secondo la quale Dati tre numeri interi a, b, n, con $n \neq 0$, si ha che $a \equiv b \bmod n$ se la loro differenza $(a - b)$ é un multiplo di n, possiamo affermare che se $(a \bmod n = b)$ allora $(a \equiv b \bmod n)$.

Si puó preliminarmente osservare che la nostra sequenza di scansione, definita dalla funzione $(h_1(k) + c_1 i + c_2 i^2) \bmod m$, ispeziona tutta la tabella se e solo se la sequanza di scansione $(c_1 i + c_2 i^2) \bmod m$ ispeziona tutta la tabella. Questo perché nella scansione quadratica, fissati c_1 e c_2, la sequenza di scansione é definita dal valore iniziale $h_1(k)$.

Siano i e j due distinti valori per cui si abbia

$$(c_1 i + c_2 i^2) \bmod m = (c_1 j + c_2 j^2) \bmod m = a$$

Allora $((c_1 i + c_2 i^2) \equiv a \bmod m)$ ed anche $((c_1 j + c_2 j^2) \equiv a \bmod m)$. Ne segue per una proprietá delle congruenze che $(c_1 i + c_2 i^2) \equiv (c_1 j + c_2 j^2) \bmod m$, da cui $((c_1 i + c_2 i^2) - (c_1 j + c_2 j^2)) = mp$, per qualche intero p.

$$\begin{aligned} c_1 i + c_2 i^2 - c_1 j - c_2 j^2 &= c_1(i - j) + c_2(i^2 - j^2) \\ &= c_1(i - j) + c_2(i - j)(i + j) \\ &= (i - j)(c_1 + c_2(i + j)) \\ &= mp \end{aligned}$$

Ne segue che $(i - j)$ é un multiplo di m solo se $\forall n \in N$ i valori $(c_1 + c_2 n)$ e m sono primi tra loro, cioé $MCD((c_1 + c_2 n), m) = 1$.

Ad esempio possiamo scegliere i valori di c_1, c_2 e m nel seguente modo:

- scegliamo m in modo che sia una potenza di 2
- scegliamo come c_1 un numero dispari
- scegliamo come c_2 un numero pari

In questo modo avremo che $(c_1 + c_2 n)$ sará un numero dispari per ogni $n \in N$.

7. Supponete di inserire le chiavi $13, 2, 30, 42, 17, 8, 7, 67, 48$ in una tavola hash di lunghezza $m = 13$ utilizzando l'indirizzamento aperto con la funzione hash ausiliaria $h'(k) = k \bmod m$. Illustrate il risultato dell'inserimento di queste chiavi utilizzando l'ispezione lineare, l'ispezione quadratica con $c_1 = 3$ e $c_2 = 3$ e il doppio hashing con $h_2(k) = 1 + (k \bmod (m - 1))$.

8. Supponete di dover inserire l'insieme di chiavi $\{11, 21, 31, 3, 7, 83, 4, 22, 36, 25, 44\}$ in una tavola hash di lunghezza $m = 9$ utilizzando l'indirizzamento aperto. Si definisca una sequenza ottimale delle chiavi in input che minimizzi le celle ispezionate sapendo che l'inserimento avviene per ispezione quadratica con funzione hash ausiliaria $h'(k) = k \bmod m$ e con $c_1 = 3$ e $c_2 = 3$. Motivare la risposta.