

Trả lời câu hỏi bài 3

Nguyễn Duy Vũ - VuHSO

Ngày 20 tháng 11 năm 2024

Câu 1

Đề bài: Với bảng băm có kích thước $m = 10000$ và hàm băm:

$$h(K) = \lfloor m \cdot (K \cdot A \bmod 1) \rfloor$$

với $A = \frac{\sqrt{5}-1}{2} \approx 0.6180339887$, tìm vị trí ánh xạ của $K = 123456$.

Giải thích: Hàm băm sử dụng phương pháp nhân:

$$h(K) = \lfloor m \cdot (K \cdot A \bmod 1) \rfloor$$

Với $m = 10000$, $A \approx 0.6180339887$, và $K = 123456$:

$$K \cdot A = 123456 \cdot 0.6180339887 \approx 76204.939460992$$

Phần thập phân của $K \cdot A$ là:

$$K \cdot A \bmod 1 = 0.939460992$$

Nhân phần thập phân với m :

$$h(K) = \lfloor 10000 \cdot 0.939460992 \rfloor = 9394$$

Đáp án: 9394.

Câu 2

Đề bài: Một bảng băm có 13 phần tử sử dụng hàm:

$$f(\text{key}) = \text{key} \bmod 13$$

và thăm dò tuyến tính. Các khóa 661, 182, 24, 103 được chèn theo thứ tự đó. Tìm vị trí của khóa 103.

Giải thích: Áp dụng $f(\text{key}) = \text{key} \bmod 13$:

- $661 \bmod 13 = 11$: Chèn vào vị trí 11.
- $182 \bmod 13 = 0$: Chèn vào vị trí 0.
- $24 \bmod 13 = 11$: Va chạm tại 11, thăm dò tuyến tính, chèn vào 12.
- $103 \bmod 13 = 12$: Va chạm tại 12, tiếp tục thăm dò tuyến tính, chèn vào 1.

Đáp án: 1.

Câu 3

Đề bài: Với văn bản thuần túy **ISRO**, hàm mã hóa:

$$C_k(M) = (kM + 13) \mod 26$$

với $k = 7$, tìm văn bản mã hóa.

Giải thích: Mã ASCII của các ký tự:

$$I = 8, \quad S = 18, \quad R = 17, \quad O = 14$$

Áp dụng hàm mã hóa $C_k(M)$ cho từng ký tự:

$$\begin{aligned} C_k(8) &= (7 \cdot 8 + 13) \mod 26 = 21 = V, \\ C_k(18) &= (7 \cdot 18 + 13) \mod 26 = 3 = D, \\ C_k(17) &= (7 \cdot 17 + 13) \mod 26 = 20 = U, \\ C_k(14) &= (7 \cdot 14 + 13) \mod 26 = 21 = V. \end{aligned}$$

Văn bản mã hóa là:

VDUV

Đáp án: VDUV.

Câu 4

Đề bài: Với bảng băm có kích thước $m = 100$ và hàm băm:

$$h(K) = \lfloor m \cdot (K \cdot A \mod 1) \rfloor$$

với $A = \frac{\sqrt{5}-1}{2}$, tìm vị trí ánh xạ của $K = 123456$.

Giải thích: Tương tự Câu 1, nhưng $m = 100$.

Tính $K \cdot A \mod 1$:

$$K \cdot A \mod 1 = 0.939460992$$

Nhân phần thập phân với m :

$$h(K) = \lfloor 100 \cdot 0.939460992 \rfloor = 93$$

Đáp án: 93.

Câu 5

Đề bài: Hàm băm

$$h(\text{key}) = \text{key} \mod 7$$

sử dụng thăm dò tuyến tính để chèn các khóa 44, 45, 79, 55, 91, 18, 63. Tìm vị trí của khóa 18.

Giải thích: Áp dụng $h(\text{key}) = \text{key} \mod 7$ và thăm dò tuyến tính:

- $44 \bmod 7 = 2$: Chèn vào vị trí 2.
- $45 \bmod 7 = 3$: Chèn vào vị trí 3.
- $79 \bmod 7 = 2$: Va chạm tại 2, thăm dò tuyến tính, chèn vào 4.
- $55 \bmod 7 = 6$: Chèn vào vị trí 6.
- $91 \bmod 7 = 0$: Chèn vào vị trí 0.
- $18 \bmod 7 = 4$: Va chạm tại 4, tiếp tục thăm dò tuyến tính, chèn vào 5.

Đáp án: 5.

Câu 6

Đề bài: Với lược đồ băm kép:

$$h(k, i) = (h_1(k) + i \cdot h_2(k)) \bmod m,$$

trong đó:

$$h_1(k) = k \bmod 23, \quad h_2(k) = 1 + (k \bmod 19),$$

và $k = 90$. Tìm địa chỉ trả về tại đầu dò thứ nhất.

Giải thích: Tính toán các giá trị:

$$h_1(90) = 90 \bmod 23 = 21,$$

$$h_2(90) = 1 + (90 \bmod 19) = 1 + 14 = 15.$$

Đầu dò thứ nhất:

$$h(k, 1) = (h_1(k) + 1 \cdot h_2(k)) \bmod 23 = (21 + 15) \bmod 23 = 36 \bmod 23 = 13.$$

Đáp án: 13.

Câu 7

Đề bài: Với phương pháp băm động cho khóa số nguyên 4 bit, tìm chuỗi khóa phù hợp để gây ra trạng thái được mô tả.

Giải thích: Phân tích trạng thái bảng băm từ đề bài, dựa trên các bit ít quan trọng nhất của khóa để phân chia. Cách duy nhất để chọn chuỗi khóa là kiểm tra tất cả các trạng thái cây được chia tách theo các bit tương ứng.

Đáp án: 9, 5, 13, 6, 10, 14.

Câu 8

Đề bài: Với băm kép:

$$h(k, i) = (h_1(k) + i \cdot h_2(k)) \mod m,$$

trong đó:

$$h_1(k) = k \mod m, \quad h_2(k) = 1 + (k \mod n),$$

với $m = 701$, $n = m - 1 = 700$, $k = 123456$. Tìm sự khác biệt giữa đầu dò thứ nhất và thứ hai.

Giải thích: Tính các giá trị:

$$h_1(123456) = 123456 \mod 701 = 255,$$

$$h_2(123456) = 1 + (123456 \mod 700) = 1 + 256 = 257.$$

Sự khác biệt giữa đầu dò thứ nhất và thứ hai là $h_2(k)$.

Đáp án: 257.

Câu 9

Đề bài: Một lợi thế của bảng băm liên kết (chaining) so với địa chỉ mở (open addressing) là gì?

Giải thích: Với phương pháp chaining, độ phức tạp trường hợp xấu nhất của tìm kiếm là $O(n)$, nhưng việc xóa dễ dàng hơn. Trong khi đó, open addressing phức tạp hơn khi xóa do phụ thuộc vào trạng thái của các phần tử khác trong bảng.

Đáp án: Xóa dễ hơn.

Câu 10

Đề bài: Trong các hàm băm sau đây, hàm nào có khả năng gây ra hiện tượng nhóm trong bảng băm nhiều nhất?

$$h(k) = k \% m$$

$$h(k) = \lfloor m \cdot (kA \mod 1) \rfloor$$

$$h(k) = k$$

$$h(k) = \left(\left(\frac{k}{m} \right) + k \cdot m \right) + k \% m$$

Giải thích:

- $h(k) = k \% m$: Hàm này có thể gây ra hiện tượng nhóm nếu m được chọn không phù hợp (ví dụ: m là số chẵn hoặc chia hết cho các yếu tố phổ biến của khóa k).

- $h(k) = \lfloor m \cdot (kA \bmod 1) \rfloor$: Đây là hàm băm nhân (multiplicative hashing), có xu hướng giảm hiện tượng nhóm nếu hằng số A được chọn cẩn thận.
- $h(k) = k$: Không phải là hàm băm thực sự, dễ gây ra nhóm vì nó phụ thuộc hoàn toàn vào giá trị của k .
- $h(k) = \left(\left\lfloor \frac{k}{m} \right\rfloor + k \cdot m\right) + k \% m$: Hàm này phức tạp, nhưng không tối ưu để tránh nhóm vì không sử dụng kỹ thuật băm hiệu quả.

Đáp án: $h(k) = k$.

Câu 11

Đề bài: Xem xét một bảng băm có 100 ô. Va chạm được giải quyết bằng cách sử dụng chuỗi. Giả sử băm đồng nhất đơn giản, xác suất 3 ô đầu tiên không được điền sau 3 lần chèn đầu tiên là bao nhiêu?

$$\frac{97 \times 97 \times 97}{100^3}, \quad \frac{99 \times 98 \times 97}{100^3}, \quad \frac{97 \times 96 \times 95}{100^3}, \quad \frac{97 \times 96 \times 95}{3! \times 100^3}$$

Giải thích: - Mỗi lần chèn khóa, xác suất chọn một ô không phải là 3 ô đầu tiên là $\frac{97}{100}$. - Sau 3 lần chèn, xác suất cả 3 lần đều không chọn 3 ô đầu tiên là $\left(\frac{97}{100}\right)^3$. - Công thức cụ thể cho xác suất này là $\frac{97 \times 97 \times 97}{100^3}$.

Đáp án: $\frac{97 \times 97 \times 97}{100^3}$.

Câu 12

Đề bài: Trong các hàm băm sau đây trên số nguyên, hàm nào sẽ phân phối khóa đồng đều nhất trên 10 nhóm $[0, 9]$, với i trong khoảng từ 0 đến 2020?

$$h(i) = (12 \cdot i) \bmod 10$$

$$h(i) = (11 \cdot i^2) \bmod 10$$

$$h(i) = i^3 \bmod 10$$

$$h(i) = i^2 \bmod 10$$

Giải thích:

- $h(i) = (12 \cdot i) \bmod 10$: Không phân phối đồng đều do tính chu kỳ của phép nhân với 12.
- $h(i) = (11 \cdot i^2) \bmod 10$: Tốt hơn, nhưng vẫn có thiên vị nhất định.
- $h(i) = i^3 \bmod 10$: Phân phối đồng đều tốt hơn các hàm khác nhờ đặc tính bậc cao.
- $h(i) = i^2 \bmod 10$: Phân phối kém đồng đều hơn vì tính chất bậc chẵn.

Đáp án: $h(i) = i^3 \bmod 10$.

Câu 13

Đề bài: Câu nào sau đây là đúng?

1. Hàm băm sẽ lấy một thông điệp có độ dài tùy ý và tạo ra một mã có độ dài cố định. 2. Hàm băm lấy một thông điệp có độ dài cố định và tạo ra một mã có độ dài thay đổi. 3. Hàm băm có thể cung cấp cùng một giá trị băm cho các thông điệp khác nhau.

Giải thích:

- 1: Đúng, vì hàm băm luôn tạo ra mã có độ dài cố định (ví dụ: SHA-256).
- 2: Sai, hàm băm không tạo ra mã có độ dài thay đổi.
- 3: Đúng, vì hiện tượng *collision* có thể xảy ra (hai thông điệp khác nhau có cùng giá trị băm).

Đáp án: Chỉ I và III.

Câu 14

Đề bài: Với bảng băm kích thước 20, xác suất bất kỳ khóa mới nào bị xung đột vượt quá 0.5 xảy ra sau bao nhiêu lần chèn?

Giải thích: Dựa trên bài toán ngày sinh (birthday problem), xác suất vượt quá 0.5 xảy ra khi số khóa n đạt khoảng $\sqrt{2 \cdot m}$, với $m = 20$.

Tính: $n \approx \sqrt{2 \cdot 20} = 6.3$.

Đáp án: 6.

Câu 15

Đề bài: Xác suất xảy ra va chạm khi băm n khóa vào bảng băm có kích thước m là bao nhiêu?

Giải thích: Xác suất va chạm:

$$1 - \prod_{i=0}^{n-1} \frac{m-i}{m}.$$

Với $n \ll m$, xác suất gần đúng là $O(n/m)$.

Đáp án: $O(n/m)$.