THE QUADRATIC HASH METHOD WHEN THE TABLE SIZE IS NOT A PRIME NUMBER

VLADIMIR BATAGELJ

"Jožef Stefan" Institute, University of Ljubljana
Ljubljana, Yugoslavia

Almost all the papers dealing with the quadratic hash methods have considered the case when the table size is a prime number. In this paper it is shown that, contrary to what is normally assumed, for the greatest part of tables whose size is not a prime number there exists a quadratic hash method whose period of search equals the table size.

KEY WORDS AND PHRASES: quadratic search, hash code, scatter storage, table size .

In the literature dominates the convinction that the period of the quadratic search, when the table size is not a prime number, is usually too small for effective use [1] . For this reason the papers deal mainly with the quadratic methods for the tables whose size is a prime number. The only exception, that I know, is the sequence

$$z_i \equiv z_0 + Ri + \frac{1}{2}i(i+1) \pmod{2^k}$$

due to Hopgood and Davenport [2] , which has the period of search $2^k - R$. The coefficients in this quadratic expression are not all integers; so it is still possible, that the quadratic search methods with integer coefficients for the tables whose size is not a prime number are really worse (the period of search is lesser) .

I tried to show this - I found out the opposite.

Consider the sequence

$$z_i \equiv z_o + ai + bi^2 \pmod{d} \qquad (1)$$

for $i = 0,1,2,\ldots$, where a and b are integer constants and d is the table size. The most important question is whether there exist indices i and j such that

$$z_i \equiv z_j \qquad \text{and} \qquad 0 \leq i < j < d \qquad (2)$$

That is equivalent to

$$z_j - z_i \equiv (j - i)(a + b(i + j)) \equiv 0 \pmod{d} \qquad (3)$$

When d is a prime number, the result is known [1] , [3] , [4] : The sequence (z_i) examines in the first d steps one half of the table (each entry twice) . This is due to the fact that the set of residues modulo a prime number is a field. In a field the equation

$$a + bx \equiv 0 \pmod{d}$$

has exactly one solution for any a and b $(b \not\equiv 0)$. If besides this

$$a \equiv 0 \pmod{d} \qquad \text{or} \qquad a \equiv b \pmod{d}$$

the sequence (z_i) examines a half of the table already in the first $(d+1)/2$ steps.

For the primes of the form $4k+3$ we can construct a "quadratic" sequence which examines the whole table in the first d steps [3] , [5] .

The existence of a solution of equation (3) is in our case actually undesirable. If we can find the coefficients a and b for which there do not exist integers i and j which at the same time satisfy the condition (2) and the equation (3) , then the corresponding sequence (z_i) has the period of search

d .

In many cases we really can find such coefficients. Let d take the form

$$d = \prod_{i \in I} p_i^{\alpha_i}$$

where p_i are prime numbers, and

$$\exists i \in I : \alpha_i > 1$$

If

$$B = \prod_{i \in I} p_i$$

and A satisfy the condition

$$(A , B) = 1$$

then the sequence (z_i) with $a = A$ and $b = BC$ examines the whole table in the first d steps.

The proof is trivial. Evidently

$$c = a + b(i + j) = A + BC(i + j)$$

is coprime with d

$$(c , d) = 1$$

For this reason we can divide the equation (3) by c . We get an equivalent equation

$$j - i \equiv 0 \pmod{d}$$

which has no solution under the condition (2) .

Coefficients A and C can be used to reduce secondary clustering [6] . If $BC \equiv 0 \pmod{d}$ this method reduces to the linear one proposed by Bell and Kaman [7] .

EXAMPLE 1: $d = 2^k$, $k > 1$

$$z_i \equiv z_o + (2Q + 1)i + 2Ri^2 \pmod{2^k}$$

for any integer Q and R .

EXAMPLE 2: $d \equiv 10^k$, $k > 1$

$$z_i \equiv z_o + Qi + 10Ri^2 \pmod{10^k}$$

Q and R are integers. Q must be an odd number which last figure is not equal 5 .

REFERENCES:

[1] W.D. MAURER: An Improved Hash Code for Scatter Storage; Comm.ACM 11,1 (Jan.1968), 35-38

[2] F.R.A. HOPGOOD, J. DAVENPORT: The Quadratic Hash Method When the Table Size is a Power of 2; The Computer Journal 15,4 (1972), 314-315

[3] CHARLES E. RADKE: The use of Quadratic Residue Research; Comm.ACM 13,2 (Feb.1970), 103-105

[4] LESLIE LAMPORT: Comment on Bell`s Quadratic Quotient Method for Hash Code Searching; Comm.ACM 13,9 (Sept.1970), 573-574

[5] A. COLIN DAY: Full Table Quadratic Searching for Scatter Storage; Comm.ACM 13,8 (Aug.1970), 481-482

[6] JAMES R. BELL: The Quadratic Quotient Method: A Hash Code Eliminating Secondary Clustering; Comm.ACM 13,2 (Feb.1970), 107-109

[7] JAMES R. BELL, CHARLES H. KAMAN: The Linear Quotient Hash Code; Comm.ACM 13,11 (Nov.1970), 675-677