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

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

Previous work on quadratic hash methods is limited mainly to the case when the table size is a prime number. Here, certain results are derived for composite numbers. It is shown that all composite numbers containing at least the square of one of the component primes have full- period integer-coefficient quadratic hash functions.

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

CR Categories: 3.74, 4.10

From the literature one gains the impression that the period of quadratic search is usually too small for effective use when the table size is not a prime number [1]. For this reason, most authors limit themselves to quadratic methods for tables whose size is a prime number or a special prime power. For example, the sequence

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

due to Hopgood and Davenport [2] has the period of search $2^k - R$. The coefficient $\frac{1}{2}$ is not an integer which underlies the belief that integer-coefficient full-period quadratic hash functions may be rare or nonexistent. However, this belief, as we shall now show, is false.

Consider the sequence

$$z_i \equiv z_0 + ai + bi^2 \pmod{d} \qquad (1)$$

for $i=0,1,2,\ldots$ where a and b are integer constants and d is the table size. We reduce the problem of the period of search to the question whether there exist indices i and j such that

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

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

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

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

For 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 what we wish to avoid. 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.

Let d take the form

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

where p_i are prime numbers, and for some $i \in I : \alpha_i > 1$.

If

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

and A satisfies the condition

$$(A, B) = 1$$

then the sequence (z_i) with $a=A$ and $b=BC$ (C any integer) 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; we write

$$(c, d) = 1$$

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

$$j - i \equiv 0 \qquad (\mathrm{mod}\ d)$$

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: $d = 10^k$, $k > 1$

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

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

This can be extended to polynomials of higher degrees straightforwardly, i.e., let

$$z_i \equiv z_o + \sum_{k=1}^{n} a_k i^k \pmod d ,$$

$(a_1, B)=1$, and for every k , $2 \le k \le n$: $a_k \equiv Bb_k$. Then the sequence (z_i) has period d .

Lately, Franc Dacar of the Ljubljana University Computing Center has found necessary and sufficient conditions for integer-coefficient quadratic hash function to have full period [8]. He has found that the hash function has period d also in the case when the following conditions hold:

(1) $d=2d_1$, where d_1 is odd

(2) $a=2a_1$, $(a_1, d_1)=1$

(3) b is odd and all primes dividing d_1 also divide b .

If d_1 is a product of different primes, then $\Delta z_i \equiv d_1 + a$

(mod d) and consequently the sequence (z_i) is linear.

The quadratic method is usually realized by the following difference shema modulo d :

$$\Delta^2 z_i \equiv b \quad , \quad \Delta z_0 \equiv a \quad , \quad \Delta z_{i+1} \equiv \Delta z_i + \Delta^2 z_i \quad , \quad z_{i+1} \equiv z_i + \Delta z_i$$

that determines the sequence

$$z_i \equiv z_0 + ai + b\binom{i}{2} \pmod{d}$$

If b is even, the hash function has integer coefficients. For odd b it turns out that only sequences of the form

$$z_i \equiv z_0 + b\binom{i+1}{2} \pmod{2^k}$$

have the full period.

A detailed theory of quadratic hash functions (based on [8]) will be described in a separate paper.

ACKNOWLEDGMENTS: I am grateful to Glenn K. Manacher, editor for Programming techniques and to dr. Boštjan Vilfan from "Jožef Stefan" Institute for suggesting some improvements in exposition.

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(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

[8] F. DACAR: Range and Structure of Quadratic Sequences, (unpublished), Ljubljana University Computing Center, Ljubljana 1974