The Rabin-Karp-Algorithm

The Rabin-Karp string matching algorithm calculates a hash value for the pattern, as well as for each M-character subsequences of text to be compared. If the hash values are unequal, the algorithm will determine the hash value for next M-character sequence. If the hash values are equal, the algorithm will analyze the pattern and the M-character sequence. In this way, there is only one comparison per text subsequence, and character matching is only required when the hash values match.

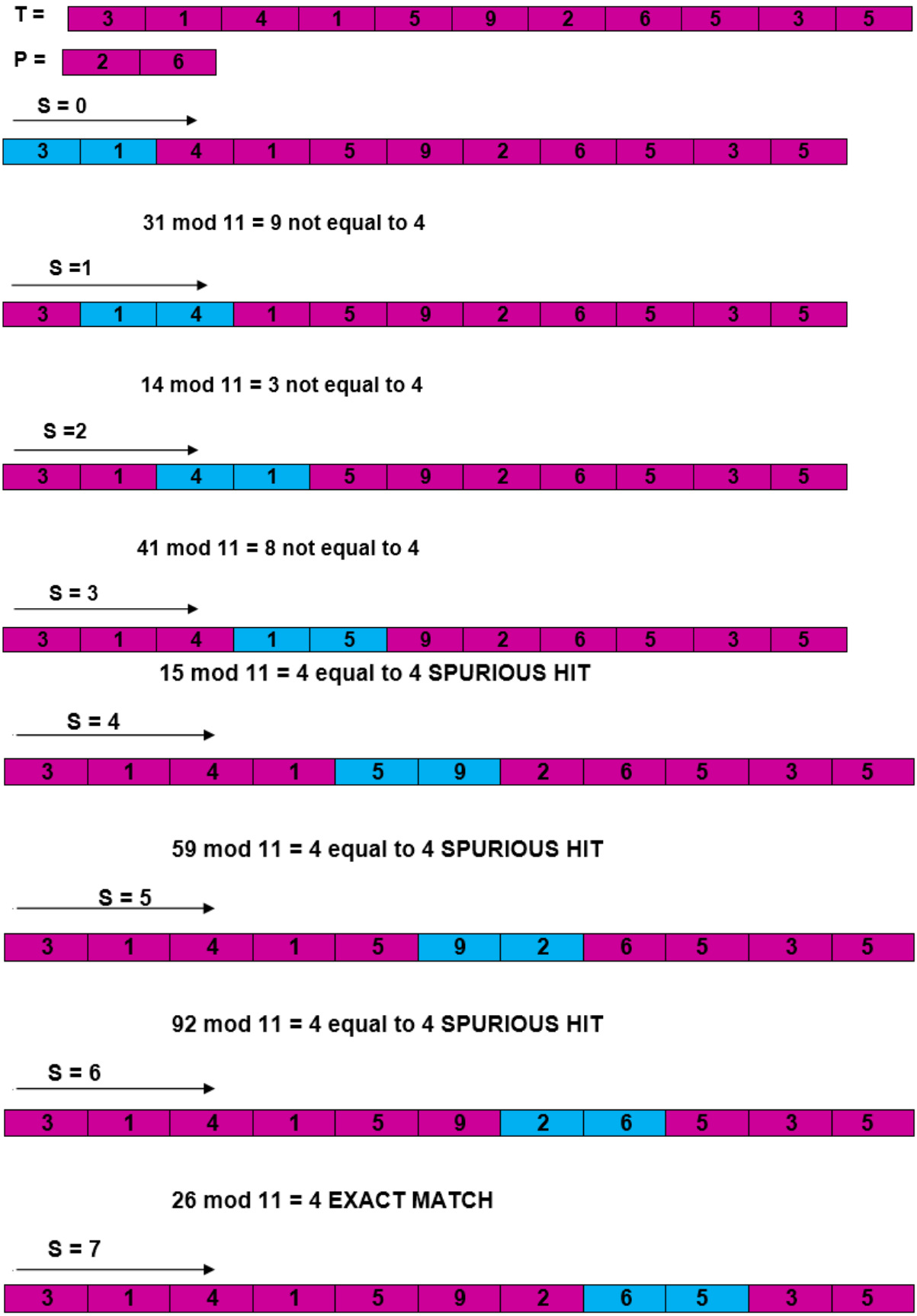
RABIN-KARP-MATCHER (T, P, d, q)

1. n ← length [T]
2. m ← length [P]
3. h ← d^{m-1} mod q
4. p ← 0
5. t_0 ← 0
6. for i ← 1 to m
7. do p ← (dp + P[i]) mod q
8. t_0 ← (dt_0+T [i]) mod q
9. for s ← 0 to n-m
10. do if p = t_s
11. then if P [1.....m] = T [s+1.....s + m]
12. then "Pattern occurs with shift" s
13. If s < n-m
14. then t_{s+1} ← (d (t_s-T [s+1]h)+T [s+m+1])mod q

Example: For string matching, working module q = 11, how many spurious hits does the Rabin-Karp matcher encounters in Text T = 31415926535.....

1. T = 31415926535.....
2. P = 26
3. Here T.Length =11 so Q = 11
4. And P mod Q = 26 mod 11 = 4
5. Now find the exact match of P mod Q...

Solution:


T =
3 1 4 1 5 9 2 6 5 3 5
P =
2 6
S = 0
3 1 4 1 5 9 2 6 5 3 5
31 mod 11 = 9 not equal to 4
S =1
3 1 4 1 5 9 2 6 5 3 5
14 mod 11 = 3 not equal to 4
S =2
3 1 4 1 5 9 2 6 5 3 5
41 mod 11 = 8 not equal to 4
S = 3
3 1 4 1 5 9 2 6 5 3 5
15 mod 11 = 4 equal to 4 SPURIOUS HIT
S = 4
3 1 4 1 5 9 2 6 5 3 5
59 mod 11 = 4 equal to 4 SPURIOUS HIT
S = 5
3 1 4 1 5 9 2 6 5 3 5
92 mod 11 = 4 equal to 4 SPURIOUS HIT
S = 6
3 1 4 1 5 9 2 6 5 3 5
26 mod 11 = 4 EXACT MATCH
S = 7
3 1 4 1 5 9 2 6 5 3 5

65 mod 11 = 10 not equal to 4

53 mod 11 = 9 not equal to 4

35 mod 11 = 2 not equal to 4

The Pattern occurs with shift 6.

Complexity:

The running time of **RABIN-KARP-MATCHER** in the worst case scenario **O ((n-m+1) m** but it has a good average case running time. If the expected number of strong shifts is small **O (1)** and prime q is chosen to be quite large, then the Rabin-Karp algorithm can be expected to run in time **O (n+m)** plus the time to require to process spurious hits.