Hash Functions

Fan-Hsun Tseng

Department of Computer Science and Information Engineering

National Cheng Kung University

Outline

- Division Method
- Mid-square Method
- Folding Method
- Digital Analysis Method

Division Method

- Representation with % or MOD
- Number of data is n
- Usually use a prime number M, where $M < n$
  - Collision happens often when M is not a prime number
- Hash address = key % M
- HomeBucket = hash(key)%M

Example

- Six keys: 26, 57, 8, 62, 77, 42
- $M = 13$

$$57\%13 = 5;\ 8\%13 = 8;\ 62\%13 = 10;$$
$$26\%13 = 0;\ 77\%13 = 12;\ 42\%13 = 3$$

26			42		57			8		62		77
0	1	2	3	4	5	6	7	8	9	10	11	12

Mid-square Method

- First phase: key squares as key^2
- Retrieve some specific numbers (*k* numbers)
- For example:
    - Retrieve thousand, hundred, tens, three numbers as hash address
    - key is 5762
        - key^2 = 33,200,644
        - Hash(5762) = 064
    - key = 2642,
        - key^2 = 6,980,164
        - Hash(2642) = 016

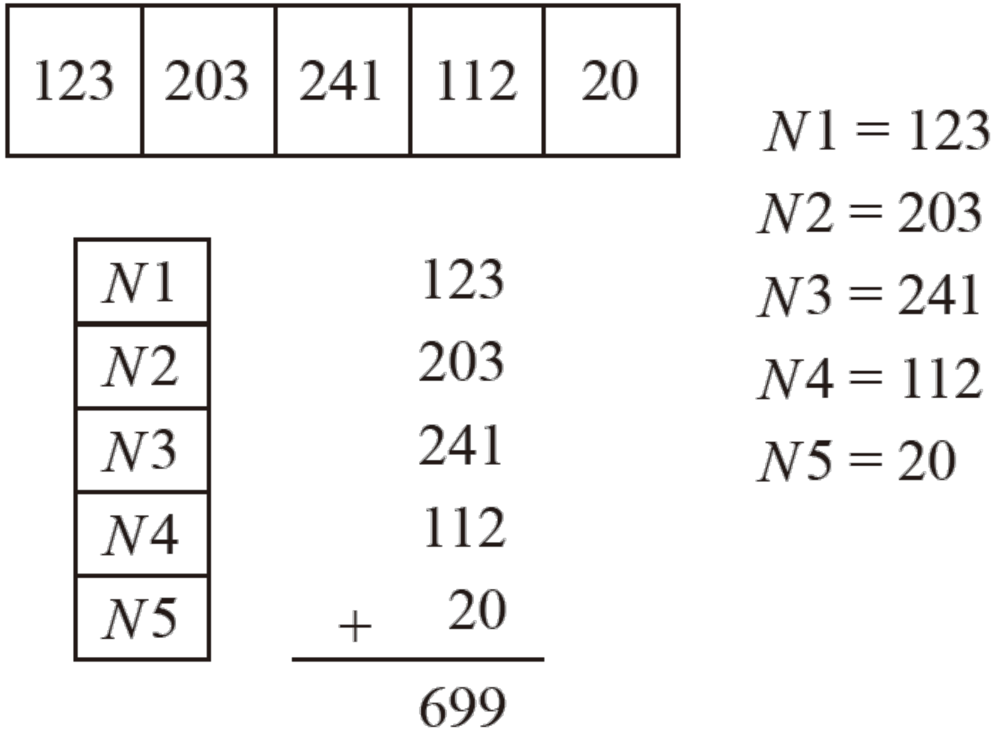
Folding Method

- Shift folding
  - Separate the key into several equal size segments
  - If the length of the last segment is less than *k*, then align to right which implies that pads 0 on the left until length equals to *k*

- Folding at the boundaries
  - Reverse the segments periodically

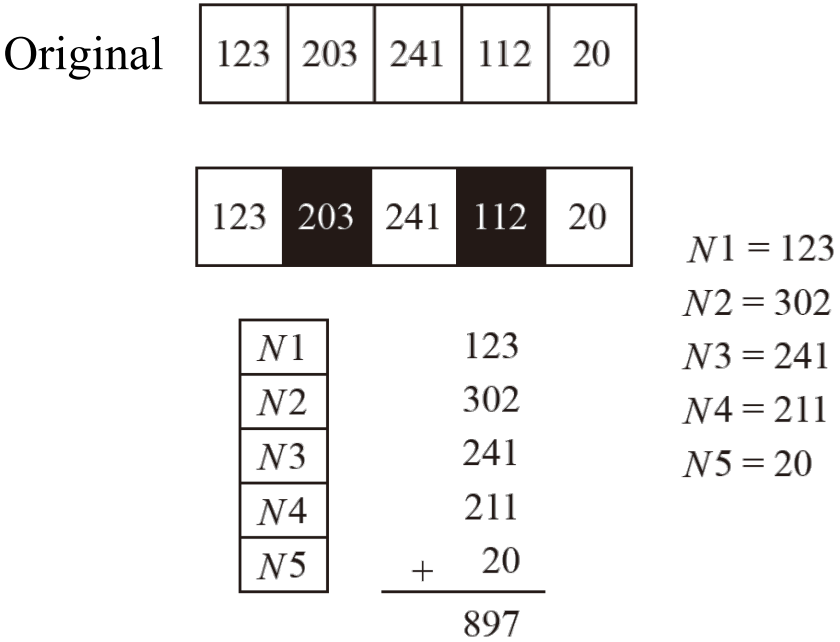
Shift Folding

- If N = 12,320,324,111,220, let k = 3, then the segments are
  - $N1$ = 123, $N2$ = 203, $N3$ = 241, $N4$ = 112, and $N5$ = 20
- Then align these segments to right and add them as the hash address, which is 699



Folding at the Boundaries

- *N* = 12,320,324,111,220, let *k* = 3, then the segments are
  - *N*1 = 123, *N*2 = 203, *N*3 = 241, *N*4 = 112, and *N*5 = 20
- Reverse N2 from 203 to 302, also reverse N4 from 112 to 211, then add all segments as hash address, which is 897



Original

123	203	241	112	20

123	203	241	112	20

*N*1	123
*N*2	302
*N*3	241
*N*4	211
*N*5	+ 20
	897

*N*1 = 123
*N*2 = 302
*N*3 = 241
*N*4 = 211
*N*5 = 20

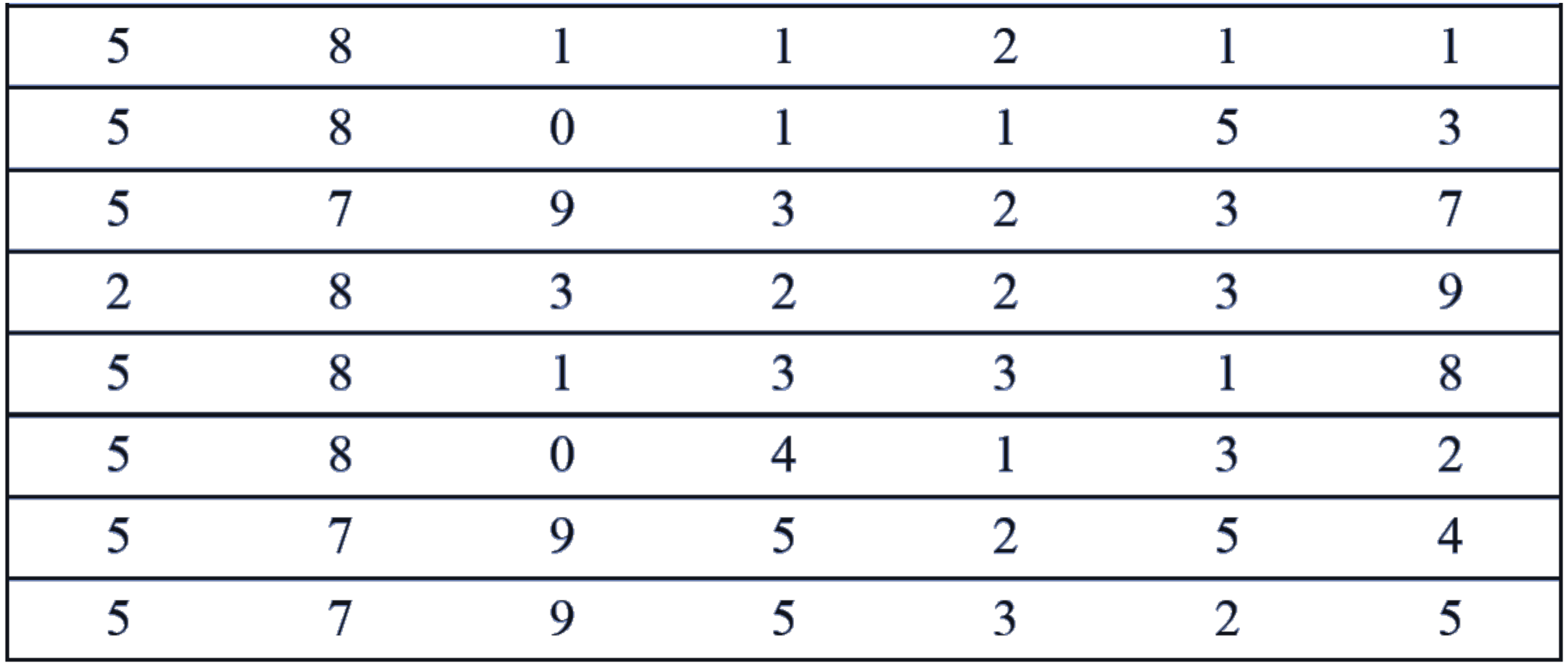
Digital Analysis Method

- Suitable for static (fixed) files
  - All (key, data) are known
  - Files are unchanged

- Analyze the distribution on every digit, selects the digits which is uniform

Digital Analysis Method (Cont’d)

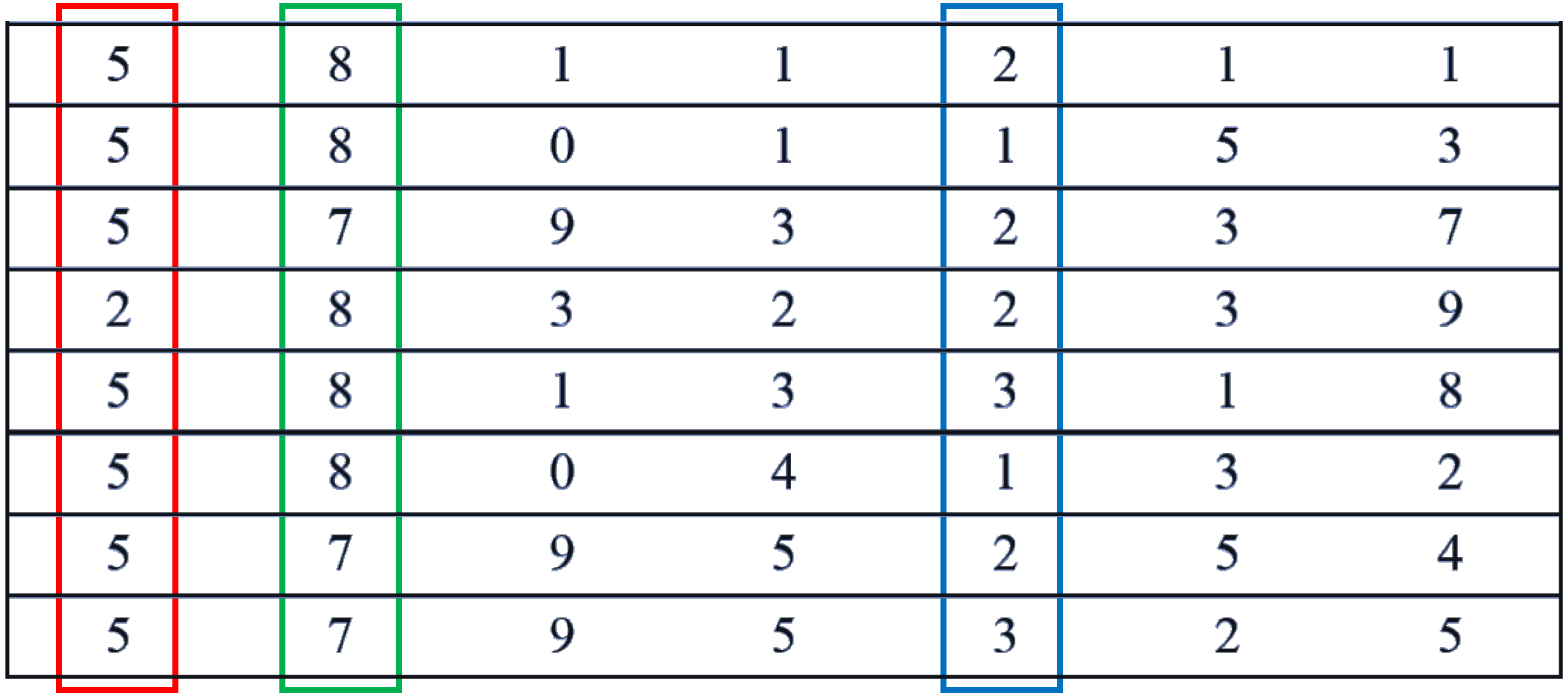
- A million number



5	8	1	1	2	1	1
5	8	0	1	1	5	3
5	7	9	3	2	3	7
2	8	3	2	2	3	9
5	8	1	3	3	1	8
5	8	0	4	1	3	2
5	7	9	5	2	5	4
5	7	9	5	3	2	5

Digital Analysis Method (Cont’d)

- Million, ten thousand, hundred digits are not uniform
- Select other digits, then do the MOD calculation



5	8	1	1	2	1	1
5	8	0	1	1	5	3
5	7	9	3	2	3	7
2	8	3	2	2	3	9
5	8	1	3	3	1	8
5	8	0	4	1	3	2
5	7	9	5	2	5	4
5	7	9	5	3	2	5

Digital Analysis Method (Cont’d)

- M = 101, then
  - key $5,8\underline{11},2\underline{11} \rightarrow 1111$, $1111\%101 = 0$
  - key $5,8\underline{01},1\underline{53} \rightarrow 0153$, $0153\%101 = 42$
  - key $5,7\underline{95},2\underline{54} \rightarrow 9554$, $9554\%101 = 60$
  - key $2,8\underline{32},2\underline{39} \rightarrow 3239$, $3239\%101 = 7$