Assignment 6

Streams Algorithms

(1) Given stream length = 75

cycles will repeat from 1 to 10

∴ 1 to 5 repeats 8 times

6 to 10 repeats 7 times.

∴ Surprise number = $5 \times 8^2 + 5 \times 7^2 = 565$

From AMS approaches, timestamp will be in the range of if $n \times (2m-1)$ value will be near to $n \times (2m-1)$ surprise number

(a) {4, 31, 72}

4 → $75 \times (2(8) - 1) = 1125$

31 → $75 \times (2(5) - 1) = 675$

72 → $75 \times (2(1) - 1) = 75$

Median = 675

(b) {25, 34, 47}

Median = 34

34 → $75 \times (2(5) - 1) = 675$

(c) {9, 50, 68}

Median = 50

50 → $75 \times (2(3) - 1) = 375$

(d) {3, 45, 72}

Median = 45

45 → $75 \times (2(4) - 1) = 525$ ← near to 565

∴ This timestamp is considered.

Scanned by CamScanner

(2) Given

End Time	100	98	95	92	87	80	65
Size	1	1	2	2	4	8	8

101 to 109 → it's appear in the stream

0	101	100	95	87	80	65
1 →	1	2	4	4	8	8

	102	101	100	95	87	80	65
1 →	1	1	2	4	4	8	8

	103	102	100	95	87	80	65
1 →	1	2	1	4	4	8	8

	104	103	102	100	95	87	80	65
1 →	1	1	2	2	4	4	8	8

	105	104	102	95	80
1 →	1	2	4	8	16

(3) Given hash function $h(x) = (3x + 7) \% 11$

x	h(x)	binary
1	10	1010
2	2	0010
3	5	0101
4	8	1000
5	0	0000
6	3	0011
7	6	0110
8	9	1001
9	1	0001
10	4	0100

Scanned by CamScanner

the given estimate is 4

$\therefore 4 = 2^2$ => only atmost 2 zeros should be present from right side upto the first 1.

$\therefore$ All the rows except 4 and 5 can be considered

$\{1, 6, 7, 10\} = 1 + 0 + 1 + 2 = 4$

$\{2, 5, 7, 10\} = 1 + 4 + 1 + 2 = 8$

$\{2, 3, 6, 9\} = 1 + 0 + 0 + 0 = 1$

$\{1, 5, 8, 9\} = 1 + 4 + 0 + 0 = 5$

(4) No of users $= 10^8$

According to the problem,

each record stores 100 bytes

$\therefore$ we can store $= \frac{10^{10}}{100} = 10^8$ records

Sample of data occupies 10^{10} bytes

Pick a subset of users and collect 10^{10} bytes

records of length 100 bytes.

Bucket number : 0 to 999999

No of emails generated by single user $= \frac{1}{10^8} = 10^{-8}$

When n emails are seen,

$10^8 \times p \times 10^{-8} \times n = pn$

Each record = 100 bytes

$= \frac{10^{10}}{100} = 10^8$

Scanned by CamScanner

$pn = 10^8$

$p = \frac{10^8}{n}$

If the threshold t, the fraction p users selected is

$$\frac{(t+1)}{1,000,000} = \frac{10^8}{n}$$

$$\therefore t = \frac{10^{14}}{n} - 1$$

(5) According to the problem,

$S \rightarrow 23$ numbers $\quad \therefore d = 23$

Bit array length $= 100$, $\quad t = 100$

Hash functions $= 1$

Fraction of 0's $= e^{-kd/t} = e^{-23/100}$

Fraction of 1's $= 1 - e^{-kd/t} = 1 - e^{-23/100}$

Scanned by CamScanner