IDM hw3
2012-11262
이승준

Q1.

1)

there are 2 ‘5’ on stream and length of stream is 10.
2/10 = 1/5

2)

sigma (1/i) where I in [2,24]
= 2.7759…
~ 3

Q2.

1)

1 * 2 + 2*2 + 4/2
= 8

2)

15 – 1)
= 7

3)

same to 1)
8

4)

[100100101] [11100001] 0 [101] [1]
1 + 2 + 4 + 4 /2
= 9

Q3.

1)

O(kn)

2)

(1 – (1 – 1/n)^km)^k
m = # of spam site, 5 G
n = # of bits, 8G
false positive = (1 – e^(-5/8k))^k

k = n/m ln2 = 8/5 ln2 = 1.1
~ 1

3)

(1 – e^(-1/8k))^k = 1/20
k = 1.965..

k >= 2

4)

let assume that we use
same # of hash functions,
same # of m

false positive of the original filter
= (1 – e^(-km/n))^(k)
= $\left(1 - e^{-\frac{km}{n}}\right)^k$

false positive of filter which has half hash function.
(1 – e^(-km/(n/2)))^(k)
= $\left(1 - e^{-\frac{2km}{n}}\right)^k$

And two filter
((1 – e^(-km/(n/2)))^(k))^2
= $\left(1 - e^{-\frac{2km}{n}}\right)^{2k}$

is $\left(1 - e^{-\frac{km}{n}}\right)^k - \left(1 - e^{-\frac{2km}{n}}\right)^{2k} > 0$?

it becomes a same Inequation question with having twice hash functions guarantees better accuracy.

And answer is not always. It depends on m, n.

Q4.

let split h to

permutation

mark : mapping [0,31] to [0,5]

least 0 bit

p1	p2	p3	
1	16	16	16
7	6	14	2
9	24	24	8
5	20	4	28
10	1	29	11
9	24	24	8
22	13	25	15
29	12	28	4
12	19	7	17
17	0	0	0

mark1	mark2	mark3	
1	0	0	0
7	0	0	0
9	0	0	0
5	0	0	0
10	1	1	2
9	0	0	0
22	1	1	4
29	0	0	0
12	2	3	1
17	0	0	0

k		3	2	3
b	2.6666666667			
estimation	8.2088198057			

where

mark				
0	0	2	4	6…
1	1	5	9	13…
2	3	11	19…	
3	7	23		
4	15			
5	31			