



Semester : VII

Subject : Big Data Analytics

Academic Year: 2024 – 2025

## Module 4

### Bloom Filtering

Apply Bloom filter to insert 9 and 11 elements using 2 hash function.

1)  $h_1(x) = X \bmod 5$

2)  $h_2(x) = (2X+3) \bmod 5$

After inserting element test whether the element 15 and 16 is present.

Sol:

**Bloom Filter**

Ex:  $m = 5$   
 $h_1(x) = x \bmod 5$   
 $h_2(x) = (2x + 3) \bmod 5$

x	$h_1(x)$	$h_2(x)$	B										
Insert 9	4	1	<table border="1"><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr></table>	0	1	0	0	1	0	1	2	3	4
0	1	0	0	1									
0	1	2	3	4									
Insert 11	1	0	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>1</td></tr><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr></table>	1	1	0	0	1	0	1	2	3	4
1	1	0	0	1									
0	1	2	3	4									

Final Bloom Filter state:

1	1	0	0	1
0	1	2	3	4

Query 15  
 $h_1(15) = 0$      $h_2(15) = 3$   
SURELY 15 is not present

Query 16  
 $h_1(16) = 1$      $h_2(16) = 0$   
FALSE POSITIVE  
16 was probably present

2 answers  
NEGATIVE    FALSE POSITIVE

Apply Bloom filter to insert 10 and 7 elements using 2 hash function.

1)  $h_1(x) = X \bmod 5$

2)  $h_2(x) = (2X+6) \bmod 5$

After inserting element test whether the element 14 and 15 is present

**m (Size of Bloom filter) = 5**

**Hash Functions:**  
1.  $h_1(x) = x \bmod 5$   
2.  $h_2(x) = (2x + 6) \bmod 5$

**Insert elements:**  
• 10  
• 7

Element to insert	$h_1(x)$	$h_2(x)$	Bloom Filter										
10	$10 \bmod 5 = 0$	$(2(10) + 6) \bmod 5 = 26 \bmod 5 = 1$	<table border="1"><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr></table>	1	1	0	0	0	0	1	2	3	4
1	1	0	0	0									
0	1	2	3	4									
7	$7 \bmod 5 = 2$	$(2(7) + 6) \bmod 5 = 20 \bmod 5 = 0$	<table border="1"><tr><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td></tr><tr><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td></tr></table>	1	1	1	0	0	0	1	2	3	4
1	1	1	0	0									
0	1	2	3	4									

Yash Poddarwar

After inserting 10 and 7 , we had to check wheather14 and 15 is present?



Present elements:

- 10
- 7

Hash Functions:

- $h1(x) = x \bmod 5$
- $h2(x) = (2x + 6) \bmod 5$

Check for elements:

- 14
- 15

Bloom Filter:

1	1	1	0	0
0	1	2	3	4

Element	$h1(x)$	$h2(x)$	Presence
14	$14 \bmod 5$ 4	$(2(14) + 6) \bmod 5$ $34 \bmod 5$ 4	At index 4, the value is 0 which means that the element is not present in the set. (Accurately says absence)
15	$15 \bmod 5$ 0	$(2(15) + 6) \bmod 5$ $36 \bmod 5$ 1	The value is 1 at both the index 0 and 1, which depicts that element is present in the set. But in reality it is not.

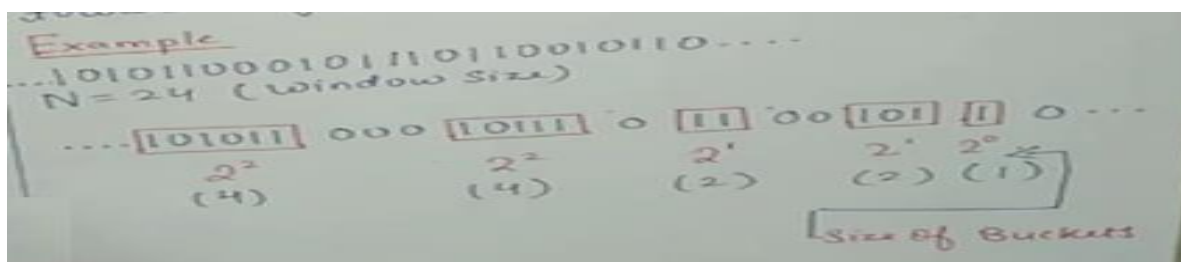
14 is not present, but 15 might be present in the data stream i.e false positive.

### Flajolet Matrin Numerical

Consider stream, $x = [1, 5, 10, 5, 15, 1]$ and hash function $h(x) = x \bmod 11$ Find the count of distinct/unique elements using the FM Algorithm					
$x$	$h(x)$	Binary	Count trailing 0	$R = \text{Max of count of trailing 0}$	Distinct elements count = $2^R$
1	$1 \bmod 11 = 1$	1	0	2	$2^2 = 4$
5	$5 \bmod 11 = 5$	101	0		
10	$10 \bmod 11 = 10$	1010	1		
5	$5 \bmod 11 = 5$	101	0		
15	$15 \bmod 11 = 4$	100	2		
1	$1 \bmod 11 = 1$	1	0		

DGIM:

Apply DGIM to count number of 1's in a given data stream  $S=10101101110110010110$  where the size of window is 24.



So, number of 1's is 13.



When we enter 0 and 1.

