

You have a set of m elements -

$$S = \{s_1, s_2, \dots, s_m\} \quad \begin{matrix} \{s_1, s_2\} \\ 11000\dots \end{matrix}$$

Bloom filter gives us a way to represent this using an n -bit binary vector - A .

Given s_i . Compute $h_1(s_i), h_2(s_i) \dots h_k(s_i)$.

Set $A[h_1(s_i)], A[h_2(s_i)] \dots A[h_k(s_i)] = 1$.

To check if s_i is there in S , you query all the above bits of A . If any one of them is zero, we know $s_i \notin S$. If all are 1, then s_i may be in S .

→ There could be false positives.

The prob of false positives can be reduced by adjusting parameters k and n/m . The best ... is achieved when roughly half

performance is achieved when roughly half
the array k is 0 and half is 1.