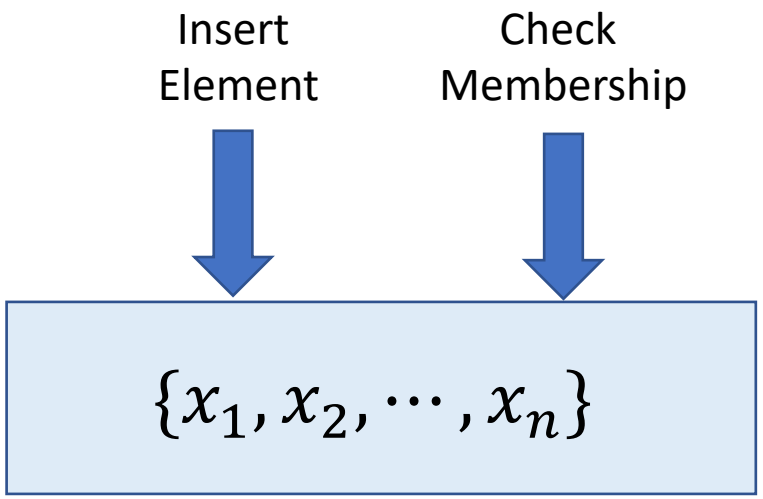
Bloom Filters

Sriram Sankaranarayanan

Data Structures and Algorithms

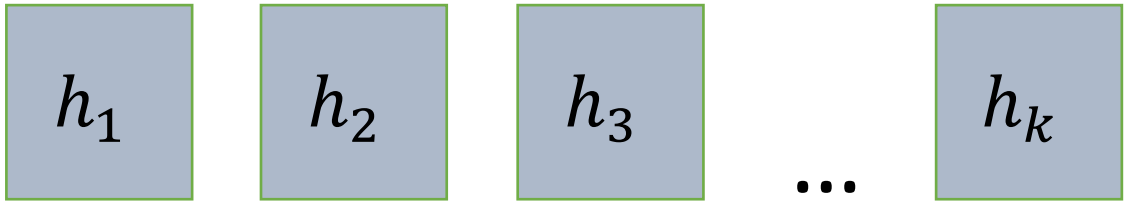
What is a Bloom Filter?

- A fast set data structure based on hashing.

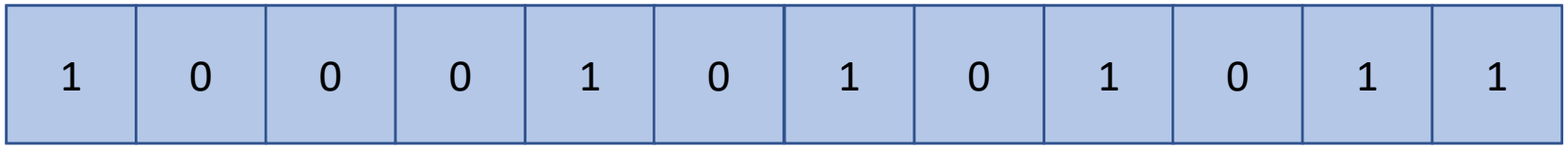


- Based on hash-tables.
- Approximate in nature: false positives possible.

Basic Idea



k randomly chosen hash functions



”Bit” String with m bits.

Insert element x: Set the bits $h_1(x), h_2(x), \ldots, h_k(x)$

Membership of element x: Are the bits $h_1(x), h_2(x), \ldots, h_k(x)$ all set to 1?

Bloom Filter: Properties

- Constant time insertion and membership check
  - More precisely $\Theta(k)$
- If element was inserted, membership query will return true.
- False positives possible.
  - Membership query may return true but element may not have been inserted.

Bloom Filter: Analysis

- Probability of false positive?

Bloom Filter By Numbers

- n = 5,000 strings (these could be long strings) inserted
- m = 25,000 bit vector size (5 bits/element)
- k = 3 hash functions.

- Probability of false positives is

$$\left(1 - e^{-\frac{kn}{m}}\right)^k = (1 - e^{-0.6})^3 = 0.09$$

Bloom Filters : Caching

*Maggs, Bruce M.; Sitaraman, Ramesh K. (July 2015), "Algorithmic nuggets in content delivery" (PDF), SIGCOMM Computer Communication Review, **45** (3): 52–66.*