algorithms

Supplementary Material A.2: Bloom filters at fifty: From probabilistic foundations to modern engineering and applications

Paul A. Gagniuc, Ionel-Bujorel Păvăloiu * and Maria-Iuliana Dascălu

Faculty of Engineering in Foreign Languages, National University of Science and Technology Politehnica Bucharest, RO-060042, Bucharest, Romania;

Glossary Entries

Approximate Membership Query (AMQ) - A probabilistic test that decides whether an element belongs to a set, with no false negatives and a small false-positive rate.

Bit Vector - A fixed-length sequence of bits that stores positions set by hash outputs during element insertion.

False-Positive Probability - The probability that the filter reports the presence of an element that was never inserted.

Hash Function - A deterministic function that maps an input element to an integer used as an index in the bit vector.

Double Hash Formula - A method that derives all required hash outputs from two base hash functions to reduce computational cost.

Hash Independence - The assumption that hash outputs do not correlate, which ensures uniform bit dispersion.

Bit Occupancy - The proportion of the bit vector that holds value 1 after multiple insertions.

Entropy Bound (m/n) - The ratio between filter size and element count, which determines space efficiency and minimal false-positive limits.

Optimal Hash Count - The value of k that minimizes the false-positive probability for fixed m and n.

Load Factor - A measure of saturation in probabilistic filters; high values raise false-positive rates.

Filter Saturation - A state in which additional insertions cause a rapid rise in false positives due to heavy bit occupancy.

Fallback Filter - A secondary Bloom filter that stores elements misclassified by a learned model to avoid false negatives.

Fingerprint - A compact hash remainder used in modern filter variants to represent elements with minimal memory cost.

Prefix Filter - A lightweight front-end filter that removes irrelevant inputs before deeper analysis.