

Remarks on Set Membership Filter Efficiency

Jonathan Burns*, Ionic Security

Set membership filters are probabilistic data structures used to succinctly encode a list of items at the cost of returning false positive results. Filter efficiency – a function of the number of items stored in a filter, the false positive rate of random membership tests, and the number of bits used to encode the filter – can be used to quantify how well a filter stores its item list. We use urn models to show that the expected peak efficiency of a standard Bloom filter increases with the number of items stored in the filter to an asymptotic upper bound of $\ln 2$, and the expected peak efficiency of a classic Bloom filter decreases as additional items are inserted into the filter to an asymptotic lower bound of $\ln 2$.

Keywords: Bloom filter, Optimization, Urn Models