

Bloom filters vs. space-efficient perfect hash

Recall that classic Bloom filters use roughly $1.44 \log_2(1/f)n$ bits per key, as seen in class (where $f = (1 - p)^k$ is the failure probability minimized for $p \approx e^{-\frac{kn}{m}} = 1/2$). The problem asks to extend the implementation required in Problem 10 by employing an additional random universal hash function $s : U \rightarrow [m]$ with $m = \lceil \frac{1}{f} \rceil$, called signature, so that $s(x)$ is also stored (in place of x , which is discarded). The resulting space-efficient perfect hash table T has now a one-side error with failure probability of roughly f , as in Bloom filters: say why. Design a space-efficient efficient implementation of T , and compare the number of bits per key required by T with that required by Bloom filters.

SOLUTION