Hierarchical Bloom Filters: A Page Fault Reducing Bloom Filter

Adam Zawierucha (adz2)

Rice University
zawie@rice.edu

Abstract

Bloom filters [1] are used ubiquitously due to their speed and memory efficiency in theory and in practice. However, the default implementation of sufficiently large bloom filters sets and reads bits appearing in different physical pages of memory, which results in an increase in page faults. In general, increase memory usage increases page faults which in turn slows down programs [2]. In this paper we propose a bloom filter implementation that guarantees one page access per insert or query, minimizing page faults. The implementation we propose implements the abstract bloom filter operations on a collection of baby bloom filters with a size equal to the system's page size. We will show theoretically and empirically that this implementation is expected to be faster than the default implementation.

Keywords: probabilistic data structures, bloom filters, memory hierarchy, implementation, page fault analysis

References

[1] Burton H. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 1970.

[2] Y.C. Tay and Min Zou. A page fault equation for modeling the effect of memory size. *Performance Evaluation*, 63(2):99–130, 2006. ISSN 0166-5316. doi:https://doi.org/10.1016/j.peva.2005.01.007. URL https://www.sciencedirect.com/science/article/pii/S0166531605000180.