

CS466 Project Proposal

Mingkun Ni, Yuanhao Zhang
{m8ni, y2384zha}@uwaterloo.ca
University of Waterloo
Waterloo, ON, Canada

July 8, 2019

Background

Hash has been a research question with a long history. The concept of hash is first used in a memo of Hans Peter Luhn in 1953 (Donald K., 2000). After the analysis of the linear probing algorithm by Donald Knuth, the analysis of hashing function becomes famous, and many talented researchers have put efforts into the development of related fields. Nowadays, hash has become not only a well-development research field but also a popular tool that is widely used in programming. Because of its wide utility, we consider it as a focus of our project.

After reading through some related research papers, we decide to look into one of them, titled *Ball and bins: smaller hash families and faster evaluation*. This paper introduces two new constructions that we could guarantee $O(\log n / \log \log n)$ maximum load when throw n balls into n bins, but they come with either a smaller description length or a faster calculation time. The special part of this research is that it innovatively describes the structure of two constructions as a multi-layer random graph. The research considers each layer as a different hash process with different input and output sizes, which is the number of bins in each layer. Thus, we want to look deep in the first construction of this research and elaborate more details.

Plan

Our research will contain 4 main sections. Section 1 will contains the introduction about the research problem including some background information and prior work contributed by other researchers. We will also introduce the construction briefly. Section 2 will list all tools that our formal proof in section 3 will needs. The formal proof of these tools will also be included. Section 3 will be the essential parts of this research: it will contain the formal introduction of our construction and a formal proof of why this construction works. In the end, Section 4 will introduce some extensional uses of this construction.

We plan to finish reading and collecting information about the construction and related tools before 24th July, and we will start writing our research paper right after that. We expect to finish writing the first draft before the end of 3rd Aug and the final draft before the end of 14th Aug.

Citation

A. Pagh and R. Pagh. Uniform hashing in constant time and optimal space. *SIAM Journal on Computing*, 38(1):8596, 2008.

J. Naor and M. Naor. Small-bias probability spaces: Efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838856, 1993.

L. E. Celis, O. Reingold, G. Segev and U. Wieder, "Balls and Bins: Smaller Hash Families and Faster Evaluation," 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, Palm Springs, CA, 2011, pp. 599-608, doi: 10.1109/FOCS.2011.49.

M. Dietzfelbinger and F. Meyer auf der Heide. A new universal class of hash functions and dynamic hashing in real time. In *Proceedings of the 17th International Colloquium on Automata, Languages and Programming*, pages 619, 1990.

N. Alon, M. Dietzfelbinger, P. B. Miltersen, E. Petrank, and G. Tardos. Linear hash functions. *Journal of the ACM*, 46(5):667683, 1999.

N. Alon, O. Goldreich, J. Hastad, and R. Peralta. Simple construction of almost kwise independent random variables. *Random Structures and Algorithms*, 3(3):289304, 1992.

O. Goldreich and A. Wigderson. Tiny families of functions with random properties: A quality-size trade-off for hashing. *Random Structures and Algorithms*, 11(4):315343, 1997.