Part 1 - The Cam Probe Sequence$^{\text{TM}}$®©

Cameron Farzaneh

April 13, 2018

The probing sequence in which I have invented is as such:

$$h(k, i) = (h'(k) + c_1 i^3 + c_2 i) \bmod m$$

Where $h'(k) = k \bmod m$

The probing sequences that currently exist, and that are used by many world wide, include Double, Quadratic and Linear. Although these sequences are very good (in that they all guarantee a unique permutation for every unique key), they face problems and strict conditions to make them work well and efficiently. For example Quadratic Probe, my competitor, faces the problem of **secondary clustering**. It requires a strict values of c_1, c_2 and m to make full use of the hash table.

This has motivated me to invent my own probing sequence, in which I attempt to minimize the effect of clustering, all while still probing the entire hash table. Like Quadratic Probe, the c_1 and c_2 for Cam Probe$^{\text{TM}}$ need to be chosen wisely in order to span the entire hash table. For a hash table size of 2^{20}, I have discovered that a value of $c_1 = 2$ and $c_2 = 3$ for works perfectly and hits every index of the hash table.

To understand why this probing sequence works, we must dive into the realm of *number theory*. Like Quadratic Probe, the initial position is offset by amounts that depend on the manner of the first value in the equation, in our case, $c_1 i^3$. Because $c_1 i^3$ grows faster than $c_1 i^2$, the next value that will be probed to has a greater distance from the previous probed index. This alone does not contribute to why each index is hit. The biggest part has to do with the second term in the equation, $c_2 i$. If we were only to have $c_1 i^3$ alone, it will follow a pattern, in which the function grows to. But it will never reach the integers in between the growth of $c_1 i^3$. This is where $c_2 i$ comes into play. It allows there to be an offset from $c_1 i^3$ to hit the numbers in between. And finally, the c_1 and c_2 are used to tweak the probing sequence for the appropriate hash table size.

I believe that my probing sequence mitigates the issue of **secondary clustering**. There maybe some form of clustering as the number of elements grows, but it will be much smaller than Quadratic Probe's clustering issue. It also does span all the hash table and I have showed this in my code.