

## CS370 - Programming assignment 2 - Write up

- 1) What hash functions did you choose and why (Hint: Cryptographic or noncryptographic)? What is the output range of the hash functions? What is the size of the Bloom filter in each case?

I choose to use a non-cryptographic hash function called murmur. I choose this because it's much faster than cryptographic functions. The output range of this function is a signed 32-bit integer. This means the return value can be anywhere from -2,147,483,648 to 2,147,483,647. For an input of 3 hash functions the size is 897,919 and for 5 hash functions it is 1,074,935.

- 2) How long does it take for your Bloom Filter to check 1 password in each case (3 hashes vs 5 hashes)? Why does one perform better than the other if any?

To check one password with three hashes it takes me .585 seconds with 5 hashes it takes .883 seconds. Fewer hashes perform better because it takes longer for them to fill up the filter. Increasing the hash amount slows the bloom filter and it fills it up faster.

- 3) What is the probability of False Positive in your Bloom Filter in each case? What is the probability of False Negative in your Bloom Filter?

The probability of a false positive for 3 hashes is 10% and for 5 hashes it is 11%. We can calculate this by using the equations below. The chance of getting a false negative is 0% bloom filters never give a false negative

$$P = \left(1 - \left[1 - \frac{1}{m}\right]^{kn}\right)^k$$

$$\left(1 - \left(1 - \frac{1}{2988233}\right)^{5 \times 623518}\right)^5$$

$$\left(1 - \left(1 - \frac{1}{2988233}\right)^{3 \times 623518}\right)^3$$

- 4) How can you reduce the rate of False Positives?

We can reduce the amount of false positives by reducing the size of the bloom filter. Since the false positive percent depends on the size of the bit array we can just decrease that or we can also reduce the amount of hashes.

Hash3 time = .585  
Hash5 time = .883

.10

.11