Assignment 6 Writeup

Shirin Rokni
shrokni

November 21, 2022

1 Question 1

Question: How does the number of Bloom filter bits examined per miss vary (for the same input) as the Bloom filter varies in size?

Answer:

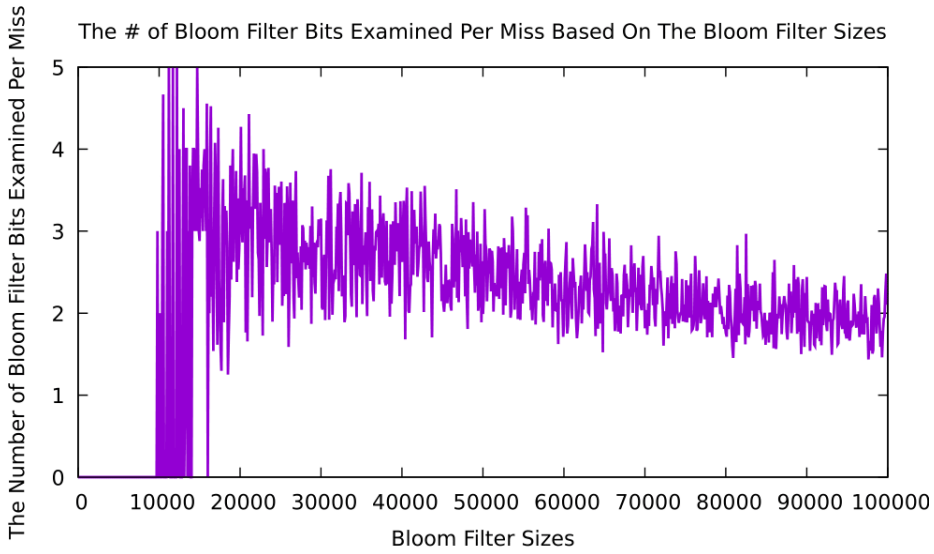

Figure 1: The number of Bloom filter bits examined per miss as the Bloom filter size varies with an array of [0...100000]. The text being parsed is the lyrics to 'Never Gonna Give You Up' by Rick Astly.

As the Bloom filter size increases, the number of Bloom filter examined per miss steadily **decrease**. Even though the Bloom filter size value goes up to 100,000, the trend of steadily decreasing would continue for a much larger range. The reason for not displaying a larger range is because my machine is unable to run that many data points in a reasonable amount of time. One notable observation is that the number of Bloom filter bits examined per miss is 0 for a few thousand points. This trend can be attributed to how, as the Bloom filter size increases, the number of the Bits examined per miss decrease.

2 Question 2

Question: How does changing the Bloom filter size affect the number of lookups performed in the hash table? This is related to the false positive rate of the Bloom filter.

Answer:

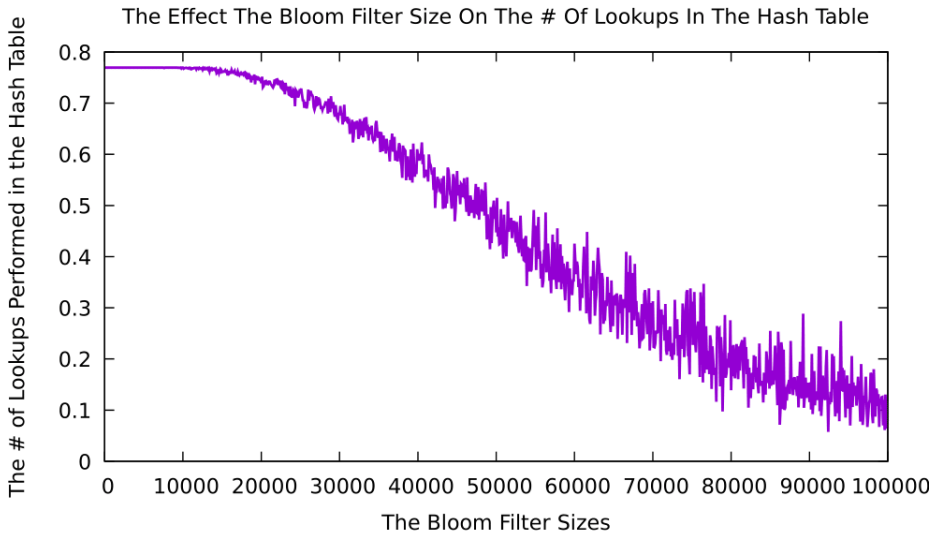

Figure 2: The number of lookups performed in the hash table as the Bloom filter size increases. The text being parsed is the bee movie script.

In order to find the number of lookups performed in the hash table, the false positive rate was used as the y-axis. As the graph displays, as the Bloom filter size increases, the number of lookups performed in the hash table **decreases**. With a larger size in the Bloom filter, the lookups approach a horizontal asymptote. Even though the graph only goes to 100,000, if the graph were to continue for a larger array size, the trend would approach a horizontal asymptote. This trend can be attributed to how a larger Bloom filter means that the number of lookups can be decreased since it optimizes space and makes it easier for each parsed word to be identified.

3 Question 3

Question: How does the number of links followed without the move-to-front rule compare to the number followed with the move-to-front rule?

Answer:

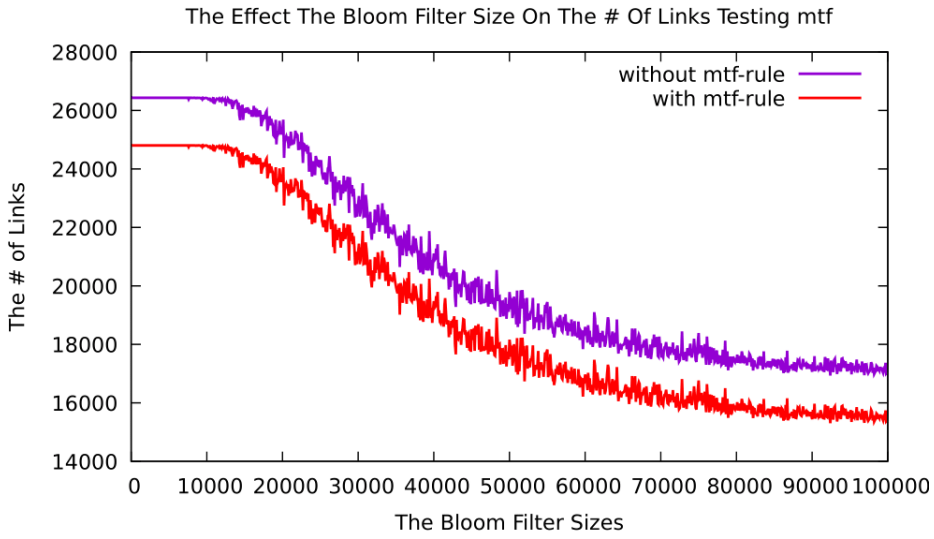

Figure 3: The number of links with the move-to-front rule enabled and disabled as the bloom filter size increases. The text being parsed is the bee movie script.

In order to display the relationship between the number of links with and without the move-to-front rule, the graph shows a changing Bloom filter size. This way, there can be multiple data points where the two cases can display their performance. The graph shows that as the size of the Bloom filter increases, the number of links decreases. However, it also shows that the move-to-front rule **consistently decreases** the number of links in comparison to the number of links without the move-to-front rule. This trend can be attributed to how the move-to-front decreases average seek length when there are a lot of words that are repeated in the text that is being parsed.

4 Question 4

Question: How does the number of links examined vary as the size of the hash table varies? What does this say about setting the size of the hash table when using a chained hash table?

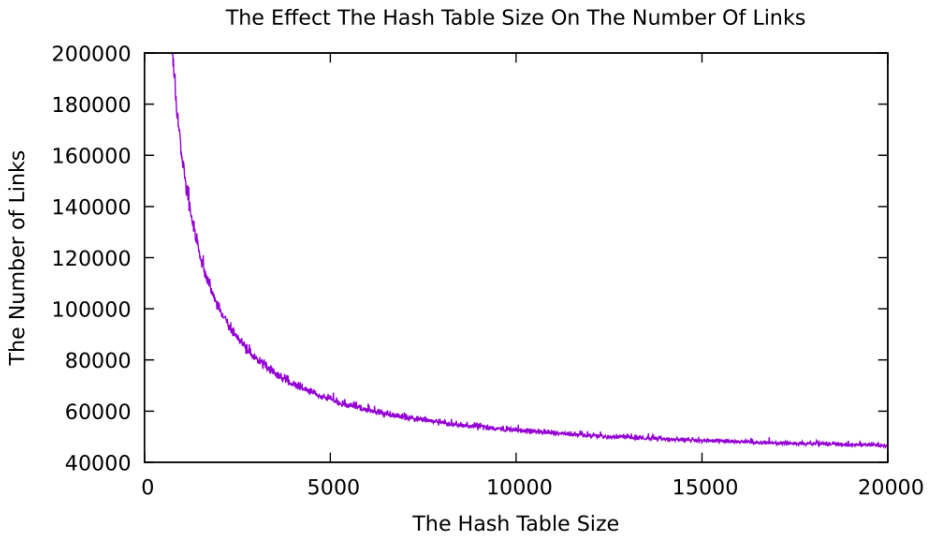

Figure 4: The number of links as the Hash table size increases. The text being parsed is the bee movie script.

As the graph depicts, as the size of Hash table increases, the number of links approach a horizontal asymptote. In other words, the number of links **consistently decrease**. This says that using a chained hash table makes it so that the larger the hash table is, the less often it needs to be accessed. This is because the chained hash table is able to avoid collisions by using linked lists, so having a larger hash table would make it so that the number of linked lists used decreases.