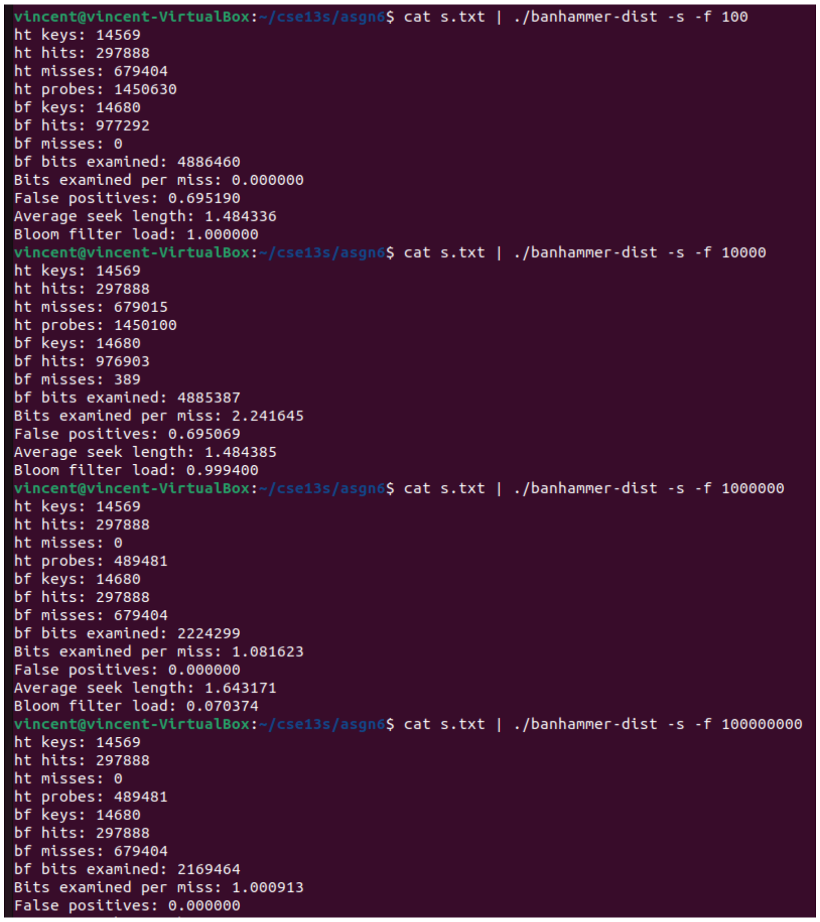
Vincent Siu
11/20/22
Vsiu2
asgn6

ASSIGNMENT 6 WRITEUP

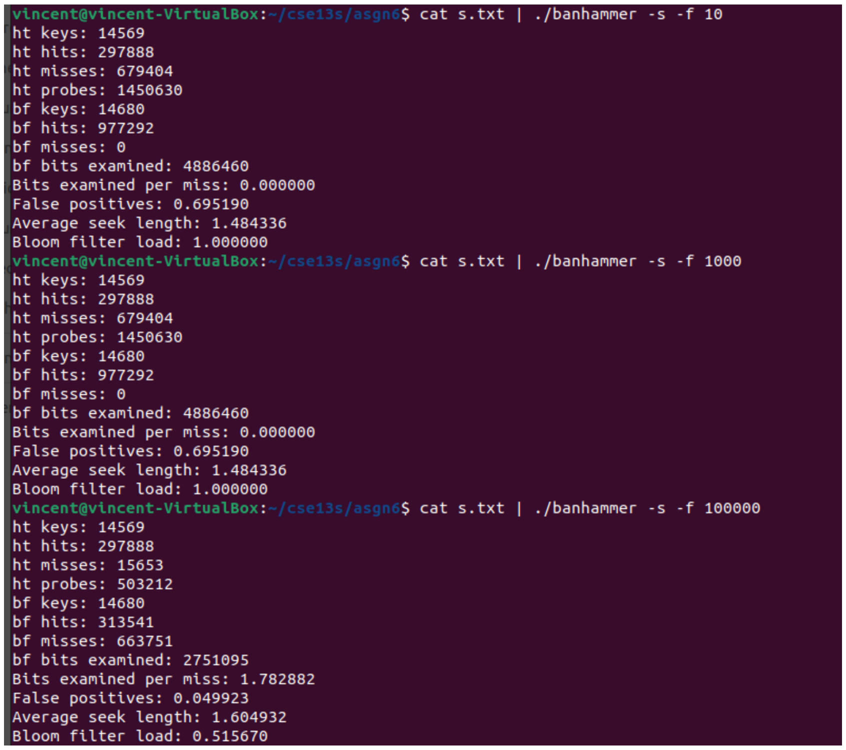
(a) How does the number of Bloom filter bits examined per miss vary (for the same input) as the Bloomfilter varies in size?
I think it really depends on the number of bits in the bloom filter. However, I do notice that, if it where to be a graph, there is a certain value that has the most examined per miss, to which a graph would form a bell curve. See the figure. We noticed that for a size of 10000, the examined/ miss is the highest aside from all the values we see tested here.



```
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -f 100
ht keys: 14569
ht hits: 297888
ht misses: 679404
ht probes: 1450630
bf keys: 14680
bf hits: 977292
bf misses: 0
bf bits examined: 4886460
Bits examined per miss: 0.000000
False positives: 0.695190
Average seek length: 1.484336
Bloom filter load: 1.000000
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -f 10000
ht keys: 14569
ht hits: 297888
ht misses: 679015
ht probes: 1450100
bf keys: 14680
bf hits: 976903
bf misses: 389
bf bits examined: 4885387
Bits examined per miss: 2.241645
False positives: 0.695069
Average seek length: 1.484385
Bloom filter load: 0.999400
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -f 1000000
ht keys: 14569
ht hits: 297888
ht misses: 0
ht probes: 489481
bf keys: 14680
bf hits: 297888
bf misses: 679404
bf bits examined: 2224299
Bits examined per miss: 1.081623
False positives: 0.000000
Average seek length: 1.643171
Bloom filter load: 0.070374
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -f 100000000
ht keys: 14569
ht hits: 297888
ht misses: 0
ht probes: 489481
bf keys: 14680
bf hits: 297888
bf misses: 679404
bf bits examined: 2169464
Bits examined per miss: 1.000913
False positives: 0.000000
```

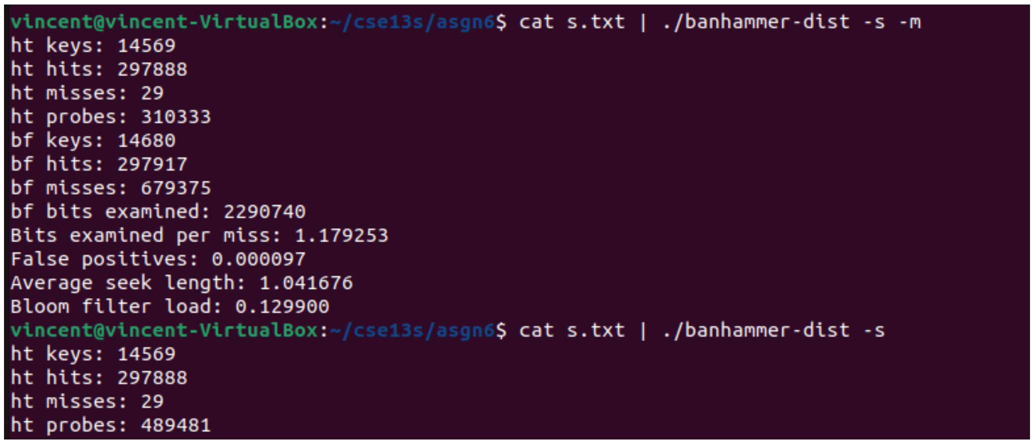
(b) How does changing the Bloom filter size affect the number of lookups performed in the hash table?
In the program, when I run several different inputs using the same text I noticed that the smaller the size of the bloom filter, the false positives are higher, meaning their relationship varies inversely. Notice in the figure below where it says “-f <size> and “false positives.” This is due to the fact that since bloom filters are space efficient, it uses less than what it needs to store a value. With a limited set of hashes we can use, there can be collisions, especially with a very limited number of bits in a bit vector, which more likely gives us false positives.

Vincent Siu
11/20/22
Vsiu2
asgn6



```
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer -s -f 10
ht keys: 14569
ht hits: 297888
ht misses: 679404
ht probes: 1450630
bf keys: 14680
bf hits: 977292
bf misses: 0
bf bits examined: 4886460
Bits examined per miss: 0.000000
False positives: 0.695190
Average seek length: 1.484336
Bloom filter load: 1.000000
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer -s -f 1000
ht keys: 14569
ht hits: 297888
ht misses: 679404
ht probes: 1450630
bf keys: 14680
bf hits: 977292
bf misses: 0
bf bits examined: 4886460
Bits examined per miss: 0.000000
False positives: 0.695190
Average seek length: 1.484336
Bloom filter load: 1.000000
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer -s -f 100000
ht keys: 14569
ht hits: 297888
ht misses: 15653
ht probes: 503212
bf keys: 14680
bf hits: 313541
bf misses: 663751
bf bits examined: 2751095
Bits examined per miss: 1.782882
False positives: 0.049923
Average seek length: 1.604932
Bloom filter load: 0.515670
```

(c) How does the number of links followed without the move-to-front rule compare to the number followed with the move-to-front rule?
When I ran several tests of this, I noticed that when mtf is enabled, the number of links decreases. This is because when an item is moved to the front after being found, and it is to be looked up again, we know right on the spot that that item is at the front. It decreases the number of links we have to traverse and the lookup times.

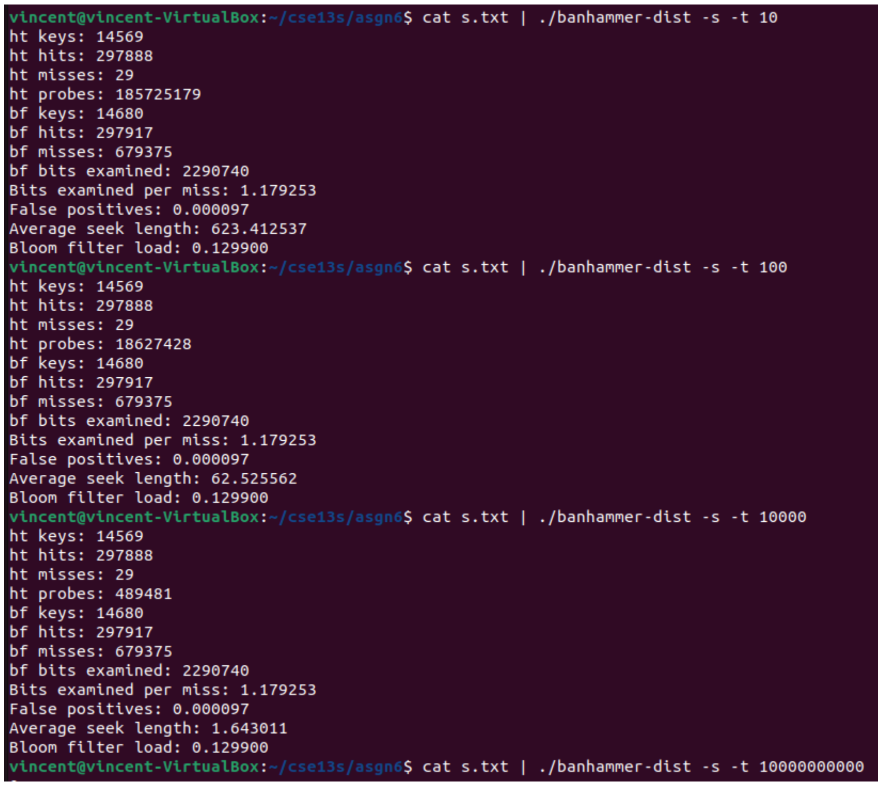


```
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -m
ht keys: 14569
ht hits: 297888
ht misses: 29
ht probes: 310333
bf keys: 14680
bf hits: 297917
bf misses: 679375
bf bits examined: 2290740
Bits examined per miss: 1.179253
False positives: 0.000097
Average seek length: 1.041676
Bloom filter load: 0.129900
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s
ht keys: 14569
ht hits: 297888
ht misses: 29
ht probes: 489481
```

(d) 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?

Vincent Siu
11/20/22
Vsiu2
asgn6

The links examined vary inversely with the size of the hash table. See the figure. This is because, since the size of the hash table is limited, each index of the hash table will have an increasing number of elements in each of the doubly linked lists. Instead of having to transverse through the hash table index, you rather have to transverse through thousands of links. We notices that ht_probes (being the number of links) decrease as we allocate more space in the hash table.



```
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -t 10
ht keys: 14569
ht hits: 297888
ht misses: 29
ht probes: 185725179
bf keys: 14680
bf hits: 297917
bf misses: 679375
bf bits examined: 2290740
Bits examined per miss: 1.179253
False positives: 0.000097
Average seek length: 623.412537
Bloom filter load: 0.129900
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -t 100
ht keys: 14569
ht hits: 297888
ht misses: 29
ht probes: 18627428
bf keys: 14680
bf hits: 297917
bf misses: 679375
bf bits examined: 2290740
Bits examined per miss: 1.179253
False positives: 0.000097
Average seek length: 62.525562
Bloom filter load: 0.129900
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -t 10000
ht keys: 14569
ht hits: 297888
ht misses: 29
ht probes: 489481
bf keys: 14680
bf hits: 297917
bf misses: 679375
bf bits examined: 2290740
Bits examined per miss: 1.179253
False positives: 0.000097
Average seek length: 1.643011
Bloom filter load: 0.129900
vincent@vincent-VirtualBox:~/cse13s/asgn6$ cat s.txt | ./banhammer-dist -s -t 10000000000
```