

## Project2:

### 1. theoretical WC: $O(n) = \lfloor n/4 \rfloor + n + 1$

These are the cases usually happen at when  $n < 100$ ; when  $n$  is large enough, it is almost unlikely to get into this case. In this case, our algorithm will first call QCOUNT on every four elements in the array ( $\lfloor n/4 \rfloor$ ) and if no set with score 4 is found, then it will utilize the brute force technique to use  $n+1$  QCOUNT calls to generate the result.

### 2. theoretical expected WC for given number of runs:

a. For  $n > 100$ , for 100 runs, the theoretical expected WC may be close to  $\lfloor n/4 \rfloor + (\lceil \lfloor n/4 \rfloor / 8 \rceil - 1) / 2 + \lfloor n/4 \rfloor / 2 + \lceil \lfloor n/4 \rfloor / 8 \rceil + a$  ( $a = n \% 4$ ); with the increase of the number of runs, this WC can get bigger as in some extreme case, it will run into  $\lfloor n/4 \rfloor + 3(\lfloor n/4 \rfloor - 1)/4 + a$  ( $a = n \% 4$ ) (In this case, all 4 elements group get score 2 and all start with 01/10) query calls.

b. For  $n \leq 100$ , no matter the number of runs, the theoretical expected WC is always  $\lfloor n/4 \rfloor + n + 1$  query calls as it is a fixed number.

### 3. theoretical AVG:

let  $x = \lfloor n/4 \rfloor$

$AVG(n) = x + (\lceil x/8 \rceil - 1) / 2 + x/2 + \lceil x/8 \rceil$  when  $n \geq 100$

$AVG(2000) = 2000/4 + (\lceil 62.5 \rceil - 1)/2 + 250 + \lceil 500/8 \rceil$

$= 500 + 31 + 250 + 63 = 844$  (calls)

Assume the 0 and 1 in the array is evenly distributed, when  $n = 2000$ , the distribution will be similar to the following table:

1111 0000		Score 4 (62.5 groups)
<hr/>		
1110 0001		
1101 0010		Score 2 (250 groups)
1011 0100		
0111 1000		
<hr/>		
0011 0101		
1001 1010		Score 0 (187.5 groups)
1100 0110		

We first call QCOUNT on every four elements in the array, which takes  $\lfloor n/4 \rfloor$  query calls. Then we take  $(\lceil x/8 \rceil - 1) / 2$  calls to process the data in score 4 groups. After that, we combine score 4 and score 2 to generate a new myarray[],

and take  $x/2$  to process the data in score 2 groups. However, in half of the cases, we could not get the result right away, so we need one extra query call for these groups to generate the results, which takes  $\lceil x/8 \rceil$  query calls.

4. observed WC and AVG for given number of runs:

Observed AVG= 843.71

Observed WC = 894

The difference between observed output and theoretical output is because the data is not perfectly distributed and the actual distribution may vary.

```
yifanj6@andromeda-2 16:10:09 ~/CS165/project2
$ ./main
n= 20,    max= 26,    avg= 26.00
n= 200,   max= 251,   avg= 84.64
n= 2000,  max= 894,   avg= 843.74
n= 17,    max= 22,    avg= 22.00
n= 18,    max= 23,    avg= 23.00
n= 19,    max= 24,    avg= 24.00
```