IB761/KI610 Big Data Algorithms - Lab 3

Summer Term 2024 Prof. A. Siebert

1. k-Server

In a 2-Server problem, we have 5 cities named L, H, B, I, A. Their distances are given by this distance table:

	L	H	B	I	A
L	0	400	300	1900	2600
H	400	0	500	2400	2900
B	300	500	0	2300	2700
I	1900	2400	2300	0	800
A	2600	2900	2700	800	0

The cost of a service is equal to the distance from the current location of a server to the city to be served. In the beginning, the two servers are located at B and L.

The requested sequence is $\sigma =$ `HALALI`.

(a) What is the cost of service if we use the greedy algorithm, i.e. if we always choose the closest server?

(b) Figure out a better solution.

2. Bloom Filter (exam question, July 2023)

Joe Flower has implemented a Bloom filter for 40 000 000 elements. He uses a bit array B of size 200 000 000 and two hash functions.

a. What error probability can Joe expect?

b. How many hash functions should Joe use to minimize the error probability?

c. What fraction of B do we expect to be filled (i.e. set to 1) after the insertion of those 40 000 000 elements, with two hash functions?

3. `Frequent`

Determine the output, i.e. the state of the monitors, of the `Frequent` algorithm for k=4, given the input sequence
`L, K, M, P, L, K, K, M, P, K, K, P, M, L, P`

That is, fill the following table where you enter the state of the monitors after each processing of a new item:

Items	Monitor 1	Monitor 2	Monitor 3
	./.	./.	./.
L	(L, 1)	./.	./.
K			
M			
P			
L			
K			
K			
M			
P			
K			
K			
P			
M			
L			
P			

Verify that all elements whose fraction is greater than 1/k are found by `Frequent`.

4. Cuckoo Hashing (exam question, July 2023)

Use Cuckoo Hashing to insert the element P into the already partially filled table (size = 11). The hash functions h1, h2 are defined as follows:

element	A	B	C	D	E	X	Y	Z	P
h1(element)	2	10	6	5	8	6	9	6	3
h2(element)	5	9	8	3	10	3	2	4	9

The state of the Cuckoo Hashing table is as follows:

1			
2	Y		
3	D		
4	Z		
5	A		
6	X		
7			
8	C		
9	B		
10	E		
11			