Information Retrieval
17/1/22

Q1: Rsync

$F_{old} = \underbrace{0011}_{h=3}\ \underbrace{1010}_{h=2}$

$F_{new} = \underbrace{0011}_{h=3}\ \overbrace{\underbrace{0101}_{h=1}\,0}^{h=2}$

$<$ 3 hash, Obj, 2 hash $>$ ⟹ client

Q1: Zsync

$F_{new} = \underbrace{0011}_{h=3}\ \underbrace{1111}_{h=3}$ ⟶ Two blocks of different content get the same hash value, so that the execution is wrongfull.

In the case of $h(x) = x \bmod 7$ we have

$F_{new} = \underbrace{0011}_{h=3}\ \underbrace{1111}_{h=1}$ $\xrightarrow{3,1}$ $F_{old} = \overbrace{\underbrace{0011}_{3}\underbrace{1000}_{1}}^{0}\ldots$

blocks	h
0011	3
0111	0
1110	0
1100	5
1000	1

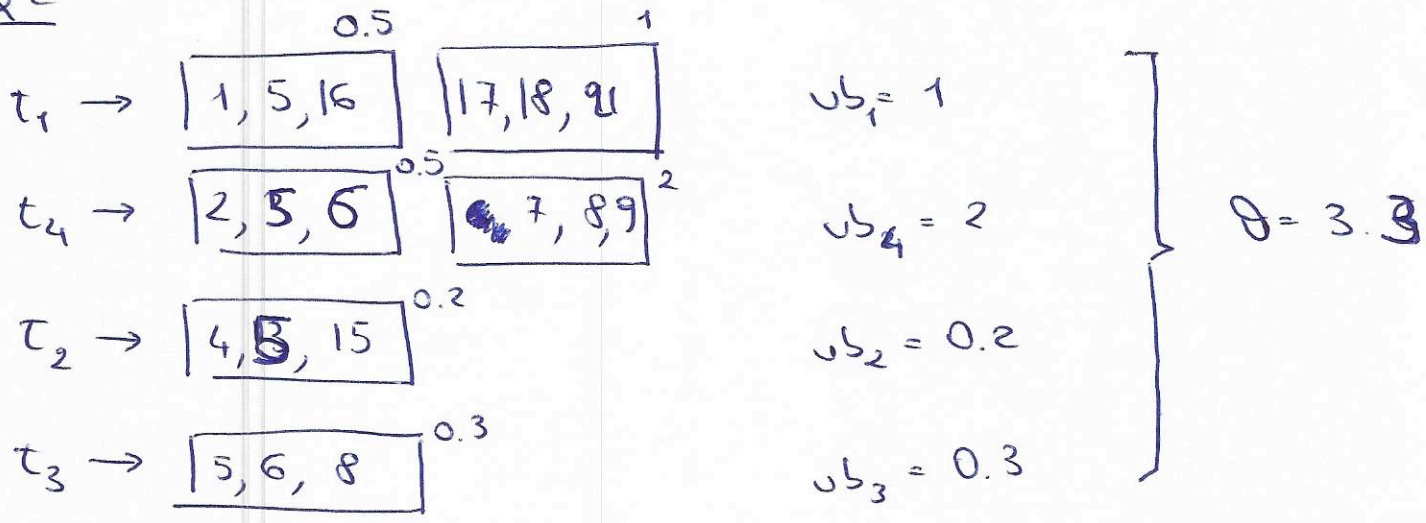
Also this execution is wrong because the blocks "1000" is recognized as "1111" so that the client would answer "11" to the server, so that the server would not send anything and the client would reconstruct $\underbrace{0011}_{3}\ \underbrace{1000}_{1}$.

Q4.

The graph may be (A) ⟶ (B)

The difference comes from the fact that $PPR_B(A) = 0$, because no edge goes out of B and so no "PR" flows outside B.
Moreover the teleportation step is centered in B.
Conversely, $PPR_A(B)$ has flow "PR" going out of A which is not well.

Q2



The candidate pivot is 5, and it occurs in all lists so that it is a potentially useful top-K doc.

So we deploy the local upper bounds which allow us to discover that those local blocks sum to $1.5 < 3.3$ and hence 5 is not a "good" doc!!.

This means that 5 is discarded and all iterators move to the right except the iterator of t_4 that discards entirely the first block which is the one having the leftmost right extreme.

Q3.

	$x \bmod 7$	$2x \bmod 7$	$3x \bmod 7$	$4x \bmod 7$
$A = \{1, 2, 3\}$	**1**, 2, 3	**2**, 4, 6	3, 6, **2**	4, **1**, 5
$B = \{2, 3, 5\}$	**2** 3, 5	4, 6, **3**	6, 2, **1**	**1**, 5, 6

$J(A, B) = \frac{2}{4} \neq \frac{1}{4}$ estimated by comparing $[1, 2, 2, 1]$ vs $[2, 3, 1, 1]$ (=)