PPRL Bloom Filter attack exercise

	Bloom Filter	Frequency
1	110100	111
2	000101	78
3	111001	49
4	111000	43
5	000111	19

Name	Frequency
Daniel	242
Carlos	184
Danilo	115
David	95
Carla	41

1. Suppose that the filters were built using 2-grams.

Name	2-grams
Daniel	{da, an, ni, ie, el}
Carlos	{ca, ar, rl, lo, os}
Danilo	{da, an, ni, il, lo}
David	{da, av, vi, id}
Carla	{ca, ar, rl, la}

2. Now build the possible candidate set of 2-grams for each position of the filter.

Position	Type	2-gram set	Candidate 2-grams
P_0	+	{da, an, ni, ie, el, il, lo, av, vi, id}	{da, an, ni, ie, el, il, lo, av, vi, id}
	-	{ca, ar, rl, lo, os, la}	
P_1	+	{da, an, ni, ie, el, il, lo, av, vi, id}	{da, an, ni, ie, el, il, lo, av, vi, id}
	-	{ca, ar, rl, lo, os, la}	
P_2	+	{da, an, ni, il, lo, av, vi, id}	{il, lo, av, vi, id}
	-	{da, an, ni, ie, el, ca, ar, rl, lo, os, la}	
P_3	+	{da, an, ni, ie, el, ca, ar, rl, lo, os, la}	{ie, el, ca, ar, rl, lo, os, la}
	-	{da, an, ni, il, lo, av, vi, id}	
P_4	+	{ca, ar, rl, la}	{la}
	-	{da, an, ni, ie, el, ca, ar, rl, lo, os, il, lo, av, vi, id, la}	
P_5	+	{ca, ar, rl, lo, os, da, an, ni, il, lo, la}	{ca, ar, rl, lo, os, il, lo, la}
	-	{da, an, ni, ie, el, av, vi, id}	

3. Analyze Bloom Filters to try to re-identify the names encoded in each one.

Filter	Positions to analyze (1)	Discarded matches (2)	Result (3)
1	0, 1, 3	Carlos -> 0	{Daniel}
		Danilo -> 3	
		David -> 3	
		Carla -> 0	
2	3, 5	Daniel -> 5	{Carlos, Carla}
		Danilo -> 3	
		David -> 3	
3	0, 1, 2, 5	Daniel -> 2	{Danilo}
		Carlos -> 0	
		David -> 5	
		Carla -> 0	
4	0, 1, 2	Daniel -> 2	{Danilo, David}
		Carlos -> 0	
		Carla -> 0	
5	3, 5	Daniel -> 0	{Carla}
		Danilo -> 3	
		David -> 3	
		Carlos -> 4	

(1) -> We just analyze those positions that are set to 1 in the given filter

(2) -> We discard those names whose q-grams are not in any of the activated positions. The position that caused the exclusion is indicated along with the name.

(3) -> Those names that are not discarded are the candidate ones. In some cases, more than a name can be paired with a single filter. It could also happen that no candidates are found for a given filter.