Step 8: Generate the Matching Positions of Records

- MPRs are the Record Index with attribute values removed and positions added. Position values correspond to positions in the UTMs of the SI.
- UTMs are sent to Carol (in step 9), which are encrypted (independently by Alice and Bob) MPRs are sent to link unit. Carol (in step 9)

Step 13: Classify the candidate record pairs

- Attribute similarities are summed for each record pair into one final similarity score. These record pairs with a total similarity score ≥ 6 will be classified as alternatives.
- Alice and Bob can agree on the shared information about the matches, send the matched records back to Alice and Bob (step 14).

Experiments

- Data from Australian telephone directory: Surnames (13,049 values and postcodes: 2,832 values) We sampled data sets from 100, 10%, 1%.
- One series of Bob’s data sets were modified (one character deleted per attribute value)

Security Analysis

- We assume all parties follow the “honest but curious” behavior.
- Alice and Bob learn each others attribute values (not the content of individual records). Knowing a rare value can infer information about people with rare names (for example)
- Carol can compile frequency statistic about BKVs, size of blocks, matching positions, etc. If Alice and Carol collude they can learn everything about Bob’s values (and vice-versa)
Conclusions and future work:

- First approach to PPRL that is experimentally evaluated on more than 1 million records.
- Our approach is scalable in size of data sets, but has a quadratic computational complexity.
- Several security drawbacks are found and their origin.
- We plan to develop a two-party version of our protocol with no leakage unit.
- We will investigate how adding extra attribute values and records will improve security of our protocol.

Timing – Steps at Database Owner

Communication – Send Matching Positions of Records to Linkage Unit

Timing – Steps at Linkage Unit