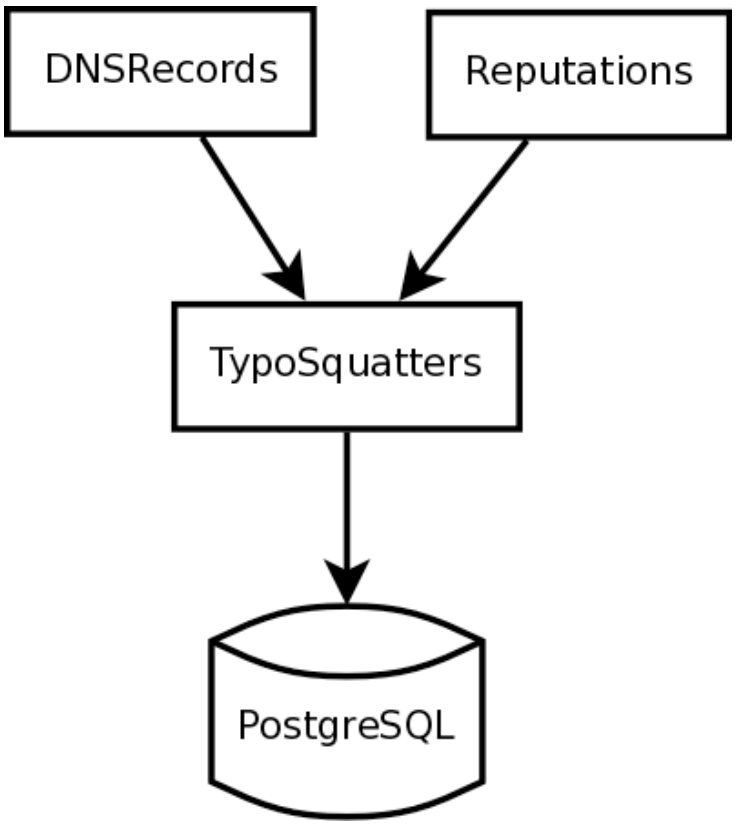
Typo Squatters:



The typo squatters module determines which domains that have been seen on the network are "typo squatters." A typo squatter is defined as a domain with bad reputation that has a low edit distance to a domain that does not have bad reputation. This module uses Levenshtein's algorithm for edit distance, but use of the Damerau–Levenshtein distance is under consideration.

Algorithm:

- Get set of domains with negative reputation from PostgreSQL
- For each domain A in the set of "bad" domains
    - For each domain B in the set of domains that were seen in DNS traffic
        - If | length(A) – length(B) | < threshold
            - Compute the Levenshtein distance ld between A and B
            - If ld < threshold AND B doesn't have bad reputation
                - A is a typo squatter