DCSRP Award 2021

A Highly Accurate Query-Recovery Attack against Searchable Encryption using Non-Indexed Documents

Marc Damie*, Florian Hahn, Andreas Peter

May 19, 2022

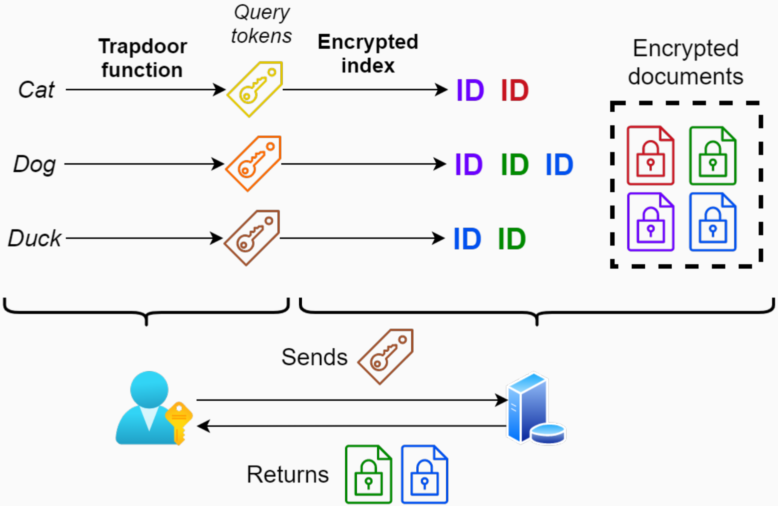
UNIVERSITY OF TWENTE.

1. Motivations

2. Score attack

3. Refined score attack

Searchable Symmetric Encryption (SSE)



UNIVERSITY OF TWENTE.

SSE attacks

- **Scope**: Passive query-recovery attacks against SSE
- SSE schemes leak the **access pattern** and the **search pattern**
- All these attacks exploit this leakage to compute a trapdoor co-occurrence and compare it to a keyword co-occurrence obtained using documents known by the attacker
- **Known-data attacks** (attacker-known documents are indexed) vs. **Similar-data attacks** (the documents are only similar $\iff$ non-indexed)

UNIVERSITY OF TWENTE.

State of the art

- Islam et al. (2012), Cash et al. (2015), Blackstone et al. (2020): only effective results as known-data attack / Pouliot and Wright (2016): low accuracy as a similar-data attack
- $\Rightarrow$ **No accurate similar-data attack**
- Known-data setup can be considered as a strong (unrealistic?) assumption

Our contributions

- A **scoring approach** to design efficient attacks with **interpretable** results
- Weakening of the attacker assumptions by proposing a highly effective similar-data attack achieving **recovery rates of up to 90%**
- A proper formalization of the concept of similarity for document sets
- Extensive analysis of our best attack: its qualities and its limitations

Attacker knowledge

- **Adversary**: honest-but-curious server
- **Similar document set**: documents similar but different to the indexed documents $\Rightarrow$ extract a vocabulary and a keyword co-occurrence matrix
- **Observed queries**: the attacker has observed some queries $\Rightarrow$ compute a trapdoor co-occurrence matrix
- **Known queries**: for a small part of the observed queries, knows the underlying keyword

1. Motivations

2. Score attack

3. Refined score attack

Creating a keyword/trapdoor vector

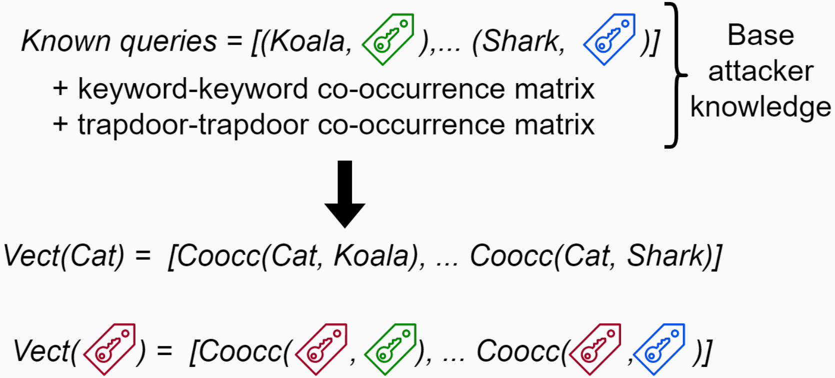


Figure: Attacker knowledge transformation

Scoring function

MatchingScore(Cat, ) = - ln(||Vect(Cat) - Vect()||)

- Using this vectorization, we can directly compare trapdoors to keywords
- The matching score is a logarithmic transformation of a distance between a keyword vector and a trapdoor vector
- Having a score provides a **result interpretability**: the higher a score is, the more likely a given prediction is

Attack algorithm

- Compute the matching score of each trapdoor-keyword pair and return the keyword providing the highest score for each trapdoor
- **Very fast**: few seconds
- **Exploitable prediction scores**: can be used to design improvement strategies (e.g. refinement and clustering presented in the paper)

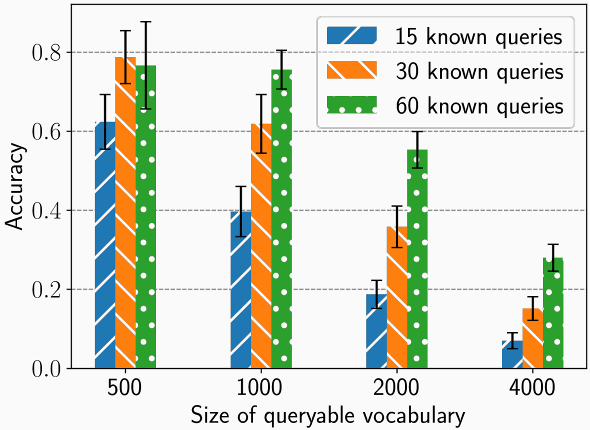
Experimental setup

- Dataset: **Enron** (i.e. documents = 30K emails)
- Attacker knowledge generation: indexed document set and attacker document set are **disjoint**
- The keyword universe (i.e. ”queryable” vocabulary) is composed of the *m* most frequent keywords of the indexed document set.

UNIVERSITY OF TWENTE.

Experimental results

Comment: improves the state-of-the-art but still impractical (no. of known queries needed too high)



UNIVERSITY OF TWENTE.

1. Motivations

2. Score attack

3. Refined score attack

Refinement strategy

Goal: reduce drastically the number of known queries needed.

We iteratively impute new known queries to refine our predictions:

1. Use the base score attack on the unknown queries.
2. Sort the predications based on their respective score.
3. Add the *k* most certain queries to the known query set. Go to step 1.

Experimental results

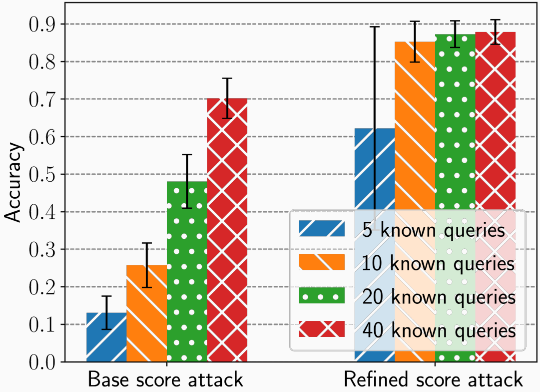
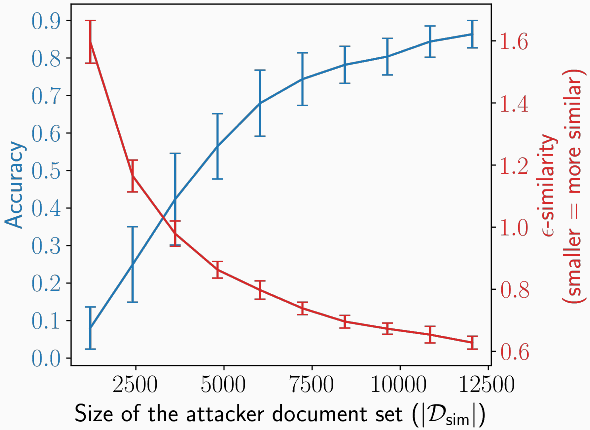


Figure: Score attack vs. Refined score attack

UNIVERSITY
OF TWENTE.

Similarity analysis

We propose a similarity metric ϵ to compare document sets. The attacker **assumes** that $\mathcal{D}_{\text{real}}$ and $\mathcal{D}_{\text{sim}}$ are ϵ-similar, with ϵ **sufficiently small**:



Refined score attack: Mitigation

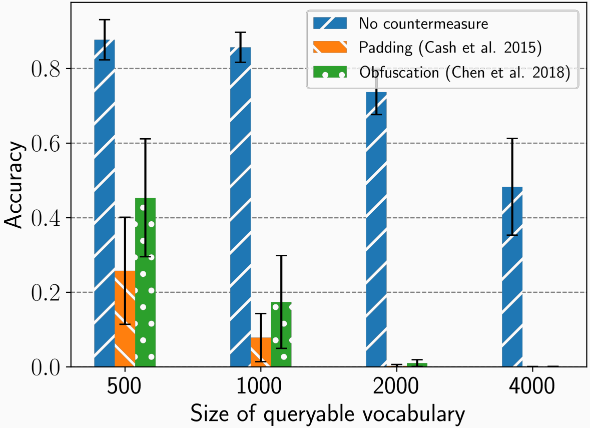


Figure: Comparison of the accuracy for two countermeasures.

Conclusion

- **Highly accurate attacks** using non-indexed documents are possible (i.e. Score and Refined Score attacks)… but can be mitigated

- Our attacks work under **weaker assumptions** on the attacker knowledge than previous attacks and move toward realistic attack situations

- **Future work**: understanding the real-life impact of SSE attacks

UNIVERSITY
OF TWENTE.

Thank you for your attention!

Source code: `https://github.com/MarcT0K/Refined-score-atk-SSE`

Feel free to contact me:

$\rightarrow$ marc.damie@inria.fr