Reality as Countermeasure?

Statistical Risk Assessment of Passive Attacks on Encrypted Keyword Search

Marc Damie, Jean-Benoist Leger, Florian Hahn, Andreas Peter

June 6, 2023 (ESSA3)

Who am I? 🧑‍🎓

- MSc. in CS from Univ. of Tech. of Compiègne 🇫🇷, specialized in **Data Mining**.
- **PhD student** between Inria 🇫🇷 and University of Twente 🇳🇱.
- Working on **privacy-preserving machine learning** under the supervision of Florian Hahn (UTwente), Andreas Peter (Uni. Oldenburg 🇩🇪), and Jan Ramon (Inria).
- Previously worked on **attacking searchable symmetric encryption**: Damie et al. (USENIX 2021), Dijkslag et al. (ACNS 2022).
- Still have a few **SSE-related ideas in mind**.

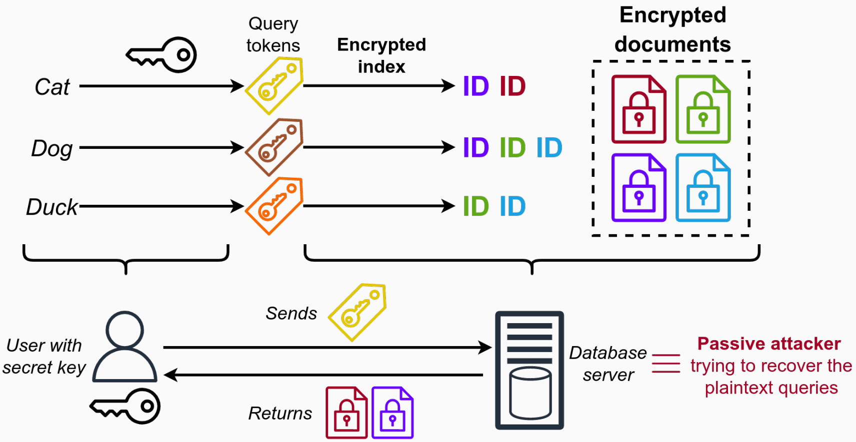
1. Introduction

2. The attacker knowledge, a noisy knowledge

3. Statistical risk assessment

4. Conclusion

Searchable Symmetric Encryption (SSE) 🔐



Attacks against SSE schemes ⚔️

- **Similar-data attacks** (based on co-occurrence information)
- Known-data attacks (based on co-occurrence information)
- Query-frequency attacks
- Active attacks
- Other attacks: against range queries, conjunctive-keyword search, etc.
- **Our focus**: similar-data attacks against static schemes with single-keyword search
- Our approaches **can be extended** to other settings.

What precisely does "similar" data mean? 🔍

- After our attack papers ⇒ **unsatisfied by the notion of “similar” data**.
- The ML literature is **more specific regarding data distribution assumptions**.
- We started exploring the **limits of this similarity assumption** using statistics.

From statistical exploration to concrete SSE problems 📚

Our statistical exploration reached novel conclusions for two main problems:

Practicality of SSE attacks

All the attack papers successively improved state-of-the-art, but the literature gives **no tool** to evaluate their **efficiency in real-world scenarios**.

SSE attack analysis

The **parameters influencing attack accuracy** are unclear, and attack papers often make **arbitrary choices** in the experiments (e.g., uniform document set splitting).

Our contributions 💡

- A robust statistical method to **assess the risk** of deploying an SSE scheme in concrete use cases.
- We show that the uniform dataset splitting used in all attack papers simulates an **advantageous scenario for the attacker** (i.e., the best source of similar doc.).
- An **attack analysis methodology** based on a similarity metric. We provide several novel conclusions about the parameters influencing attack accuracy.

📅 Paper under submission...

Our contributions 💡

- A robust statistical method to **assess the risk** of deploying an SSE scheme in concrete use cases. [**Focus of this presentation** 🎯]

- We show that the uniform dataset splitting used in all attack papers simulates an **advantageous scenario for the attacker** (i.e., the best source of similar doc.).

- An **attack analysis methodology** based on a similarity metric. We provide several novel conclusions about the parameters influencing attack accuracy.

📅 Paper under submission...

1. Introduction

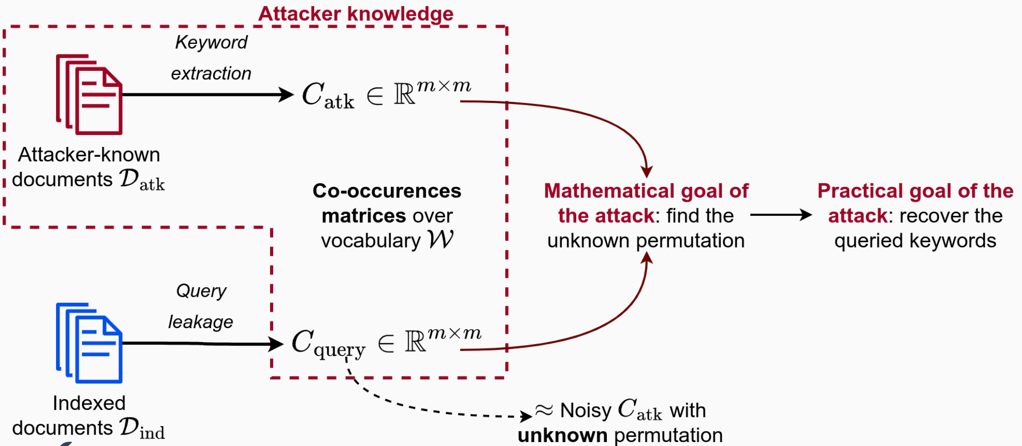
2. The attacker knowledge, a noisy knowledge

3. Statistical risk assessment

4. Conclusion

Simplified attacker knowledge: co-occurrence matrices 📝

Let n_{ind} (resp. n_{atk}) be the size of D_{ind} (resp. D_{atk}).



Revisiting the co-occurrence matrices 🔬

Our intuition

As in ML, we consider a **dataset as a sample of a random distribution.** We want to leverage the **randomness contained in the document sets**.

Co-occurence matrix distribution

The co-occurrence matrix is drawn from a random matrix distribution composed of (dependent) **Binomial variables.** Details in the paper.

NB: D_{ind} and D_{atk} can have different random distributions.

Towards a statistical hardness assumption 🛡️

⚠️ **Estimation vs. probabilities**: C_{query} and C_{atk} = **estimators** of unknown proba.

SSE attack as an estimation problem

- SSE attack problem $\approx$ **representative sampling for a survey**.
- $\Rightarrow$ attack success depends on the knowledge **size, quality and distribution**.

Statistical hardness assumption

- Classic crypto: **computationally expensive** cryptoanalysis $\Rightarrow$ sec. guarantee.
- Encrypted search: **unlikelihood** of having a precise estimation (i.e., a "similar enough" dataset) $\Rightarrow$ sec. guarantee.
- Risk assessment **quantifies the statistical hardness**.

1. Introduction

2. The attacker knowledge, a noisy knowledge

3. Statistical risk assessment

4. Conclusion

The gap in the literature

Concrete deployment problem

A company wants to deploy **encrypted mailboxes with SSE** for its employees.

Existing solutions to assess the risk?

- Consider the research results on **Enron and Apache datasets** $\Rightarrow$ Problem: Enron and Apache are not similar (i.e., **cannot represent all email use cases**)
- The company has a **dedicated sample dataset** $\Rightarrow$ Problem: the **dataset size limits the simulations** (e.g., cannot simulate attacks with large attacker knowledge).

What about a theoretical bound? 🤔

Problems with theoretical bounds

- SSE attack problem is **complex**: $\mathcal{NP}$-complete, **dependent** random variables.
- A theoretical bound could be **non-informative** (i.e., too loose).
- Any scheme modification (e.g., attack mitigation) **requires a new analysis**.

Benefits of empirical bounds

- **Consider the use case specificities** (*via a sample dataset*) to obtain tight bounds.
- **Support search scheme modifications**, such as attack countermeasures.

$\Rightarrow$ **Our objective**: a method to **bound the attack accuracy** for a given use case (i.e., based on a **sample dataset representative** of the use case).

Estimating an empirical bound 📉

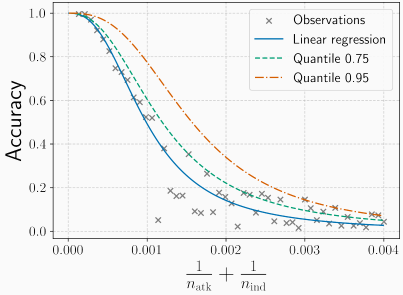


Figure: Accuracy upper bound of the IHOP attack (quantile: 0.95)

Conservative risk assessment

"Advantageous" simulation parameters: realistic attackers cannot benefit from better conditions.

Quantile regression

A quantile regression estimates (b, a) s.t. $Q_Y(\alpha) = b \cdot X + a^{\dagger} \Rightarrow$ ideal for a bound estimation.

Our upper bound function

$Q_{\text{Acc}}(\alpha; n_{\text{ind}}, n_{\text{atk}}) = \text{expit}(b \cdot \log(\frac{1}{n_{\text{ind}}} + \frac{1}{n_{\text{atk}}}) + a)$.
Detailed motivations in the paper.

$^{\dagger} Q_Y(\alpha)$: quantile α of data distribution Y.

Supporting real-world deployments 🛠️

Setting a maximum index size

Deduce n_{max} s.t. $\lim_{n_{\text{atk}} \to \infty} Q_{\text{Acc}}(\alpha; n_{\text{max}}, n_{\text{atk}}) < \text{negl}$

Security guarantee

If the size limit is respected, the **attack accuracy remains negligible** with high probability.

Limitation

The estimated upper bound holds for a **specific attack** on a **given use case**.

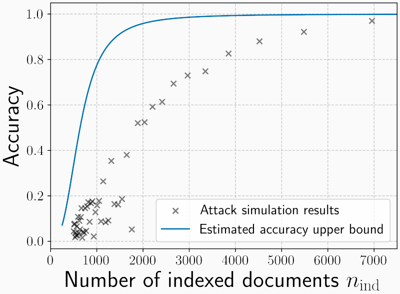


Figure: Accuracy upper bound of the IHOP attack (quantile: 0.95)

Risk assessment pipeline ⚙️

- **Find a sample dataset** representative of the use case.
- **Simulate attacks** using this dataset and the advantageous simulation parameters identified in the paper.
- Compute the **quantile regression** on the simulation results.
- Estimate a **maximum index size** and decide whether it is too low for the use case.

♻️ **Reproduce this protocol** if new attacks are released (or if the use case evolves).

Tuning the security of SSE deployments 🔧

- Maximum index size could be too small ⇒ **insecure use case** by default.
- Solution: **attack mitigation** techniques.
- Risk assessment helps choose parameters **minimizing the overhead**.
- Can also tune the secure index parameters (e.g., queryable vocabulary).

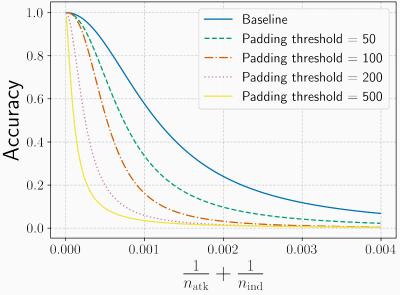


Figure: Accuracy upper bound with varying mitigation parameters

1. Introduction

2. The attacker knowledge, a noisy knowledge

3. Statistical risk assessment

4. Conclusion

Conclusion 📌

- The **stochastic model of co-occurrence matrices** provides a novel understanding of the SSE attack problem.
- We conceived a simple **risk assessment protocol based on robust statistical tools** to support real-world deployments.
- Some use cases can be deployed **securely without dedicated attack mitigation techniques**.
- We also provide various **novel insights about attack analysis methodology**.

What's next? 🧪

- 💡 A **unified security framework** for all privacy-preserving technologies with **statistical leakage** (including SSE and PPML)? **Bayes security** measure [CSF'23]?
- 💡 Formalizing the notion of **statistical hardness** assumption.
- 💡 **Building upon recent papers**? (Gui et al. [2023], Kornaropoulos et al. [2022])
- 💡 **Extending** the risk assessment and similarity analysis to **other settings**: range queries, active attacks, query-frequency attacks, etc.

✒️ **Contact me** if you want to collaborate on these topics: marc.damie@inria.fr

Thank you for your attention!

Additional slides

Uniform document set splitting, a favored attacker simulation

ϵ-similarity metric 📏

Let C_{ind} be the matrix C_{query} with the same rotation as C_{atk}.

Definition

The document sets D_{ind} and D_{atk} are ϵ-*similar* if:

$$\epsilon = \left\| \frac{C_{\text{ind}}}{n_{\text{ind}}} - \frac{C_{\text{atk}}}{n_{\text{atk}}} \right\|$$

Interpretation

The ϵ-similarity quantifies the divergence between two document sets.

Inria

Uniform document set splitting, a favored attacker simulation 🧑‍💻

All attack papers use uniform splitting (e.g., on the Enron email dataset) to generate the document sets in their experiments.

Goal of this contribution

Shows that uniform splitting $\Rightarrow$ best-case scenario for the simulated attacker.

Steps

- Uniform splitting contrary to other methods $\Rightarrow$ equal document set distributions
- Equal (document set) random distributions $\Rightarrow$ smaller ϵ-similarity
- Smaller ϵ-similarity $\Rightarrow$ higher accuracy [Done in a previous paper]

Uniform sampling ⇒ equal document set distributions

Let p_{ind} and p_{atk} parametrize the random distributions of C_{ind} and C_{atk}.

Statistical test

We conceived a **statistical test** for the hypothesis $p_{\mathrm{ind}} = p_{\mathrm{atk}}$ ($p_{\mathrm{ind}}, p_{\mathrm{atk}} \in [0,1]^{m \times m}$).

Experimental results

Tested the hypothesis with two sampling methods:

- Uniform sampling ⇒ Test not rejected (p-value always above 0.01).
- Year sampling ⇒ test strongly rejected (p-value below machine epsilon).

Equal random distributions $\Rightarrow$ smaller ϵ-similarity

Let $\mathcal{E}^{p_{\text{ind}}, p_{\text{atk}}}$ be the random distribution of the ϵ metric.

Stochastic Dominance

Let X, Y be two random distributions, $X \preccurlyeq Y \iff \forall z, \mathbb{P}(X \geq z) \leq \mathbb{P}(Y \geq z)$

Our result

We prove that asymptotically: $\mathcal{E}^{p_{\text{ind}}, p_{\text{ind}}} \preccurlyeq \mathcal{E}^{p_{\text{ind}}, p_{\text{atk}}}$.

Interpretation

Equal document set distributions stochastically produce smaller ϵ

Additional slides

Attack analysis based on a similarity metric

Attack analysis based on a similarity metric 📊

Goal of this contribution

Use a similarity metric to improve attack analysis and comparison.

Example novel insight

The document set similarity is not the only factor influencing attack success.

Attack comparison

ϵ-similarity + regression techniques $\Rightarrow$ consistent and interpretable results.

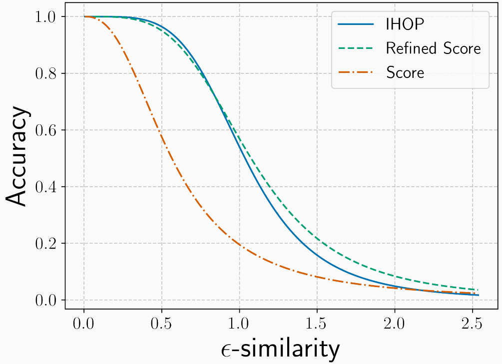


Figure: Comparison of the estimation accuracy functions for three attacks.

A few novel insights about SSE attacks

- Indexed and attacker document set sizes have a **symmetric influence on accuracy**.
- Document set similarity is **not the only factor influencing attack success**.
- Leakage does not need to be indistinguishable, **just noisy enough**.