Hiding the access pattern is not enough:

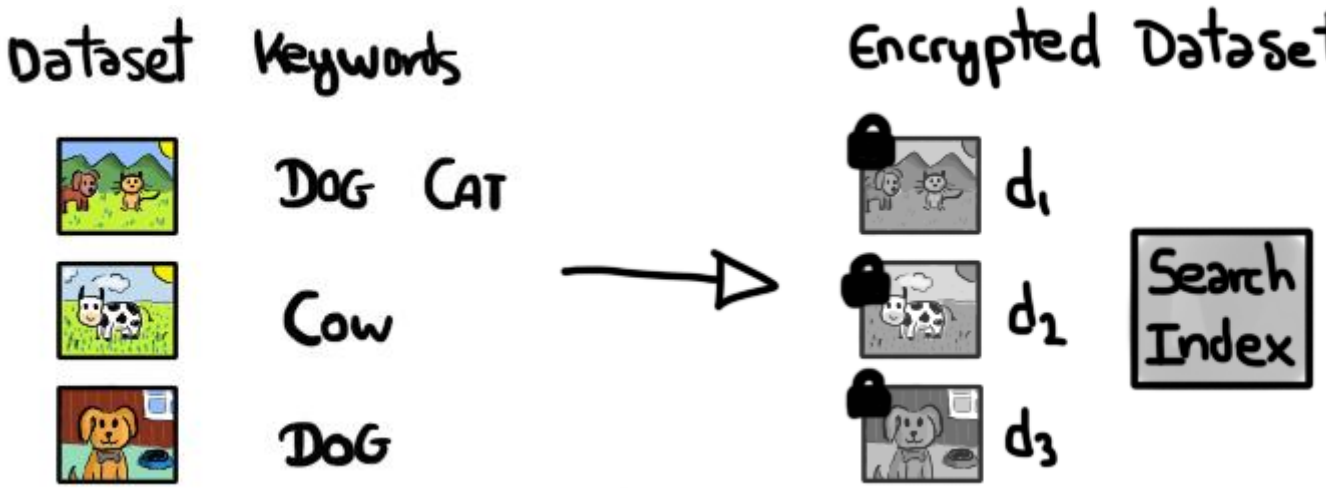
Exploiting search pattern leakage in Searchable Encryption

Simon Oya Florian Kerschbaum

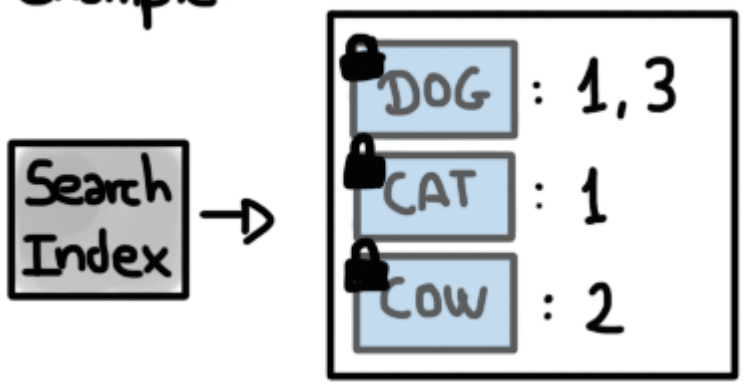
University of Waterloo CrySP

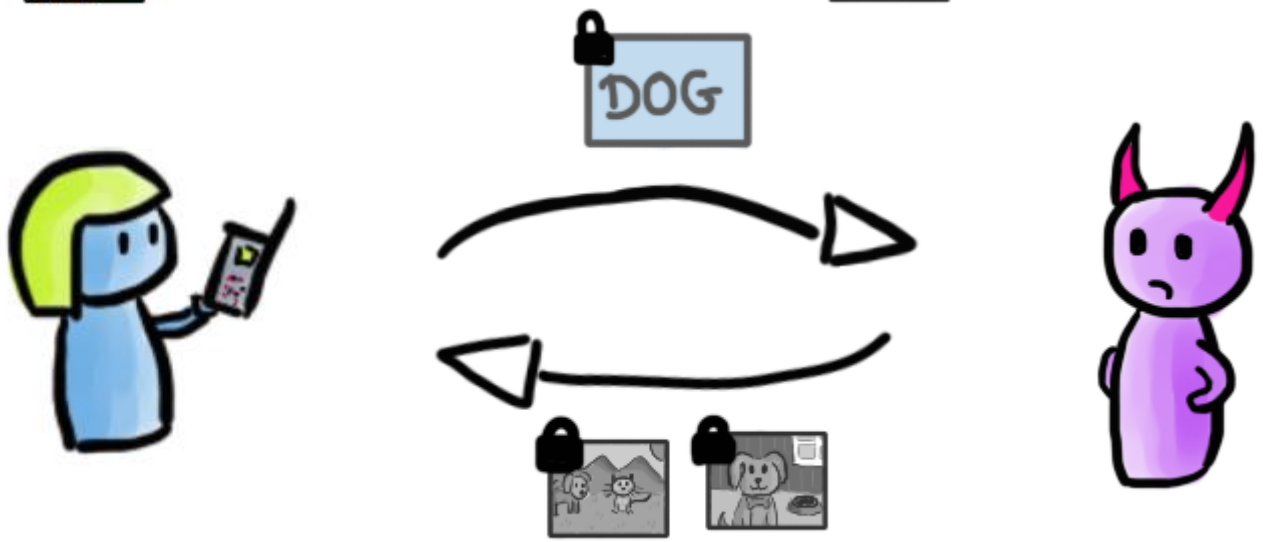
Overview: searchable encryption

Searchable encryption allows a user to send her encrypted database to a server while still being able to perform secure searches over it.



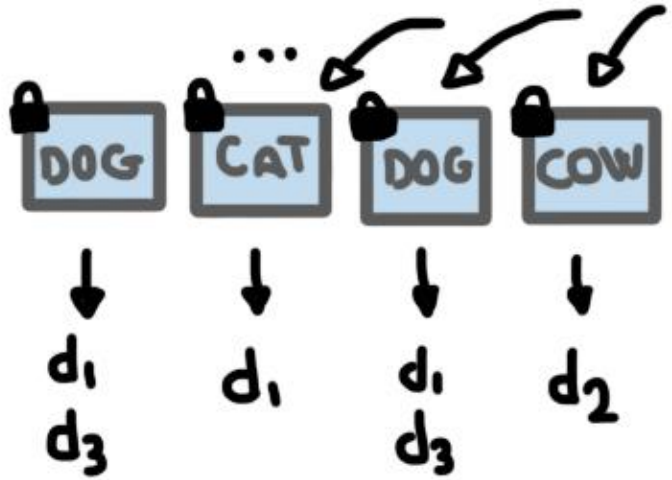
Example:





Service Provider = (Passive) Adversary

Leakage

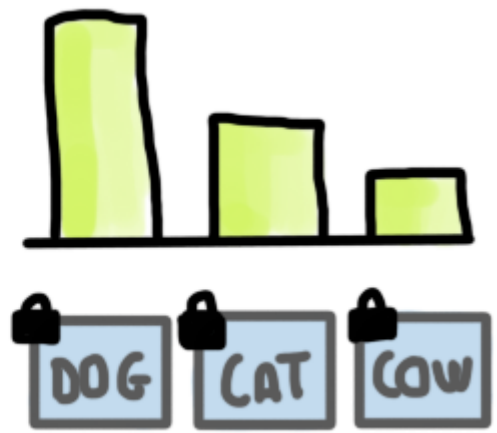


Efficient searchable encryption schemes leak the **search pattern** (which can be used to compute the frequency of each query token) and the **access pattern** (which reveals how many documents match the query).

Search pattern

↓

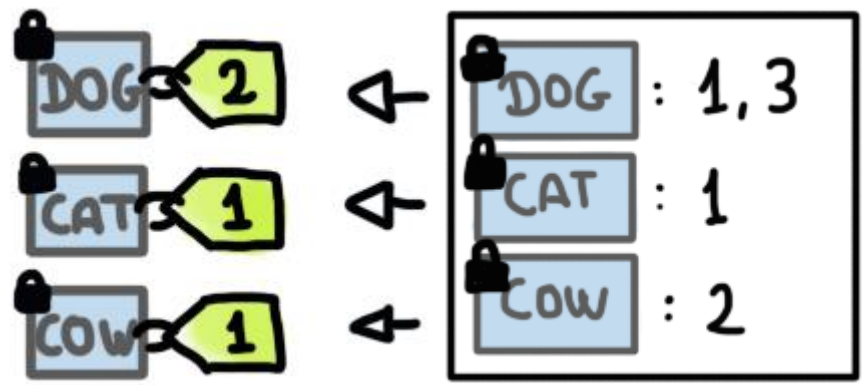
Query frequency



Access pattern

↓

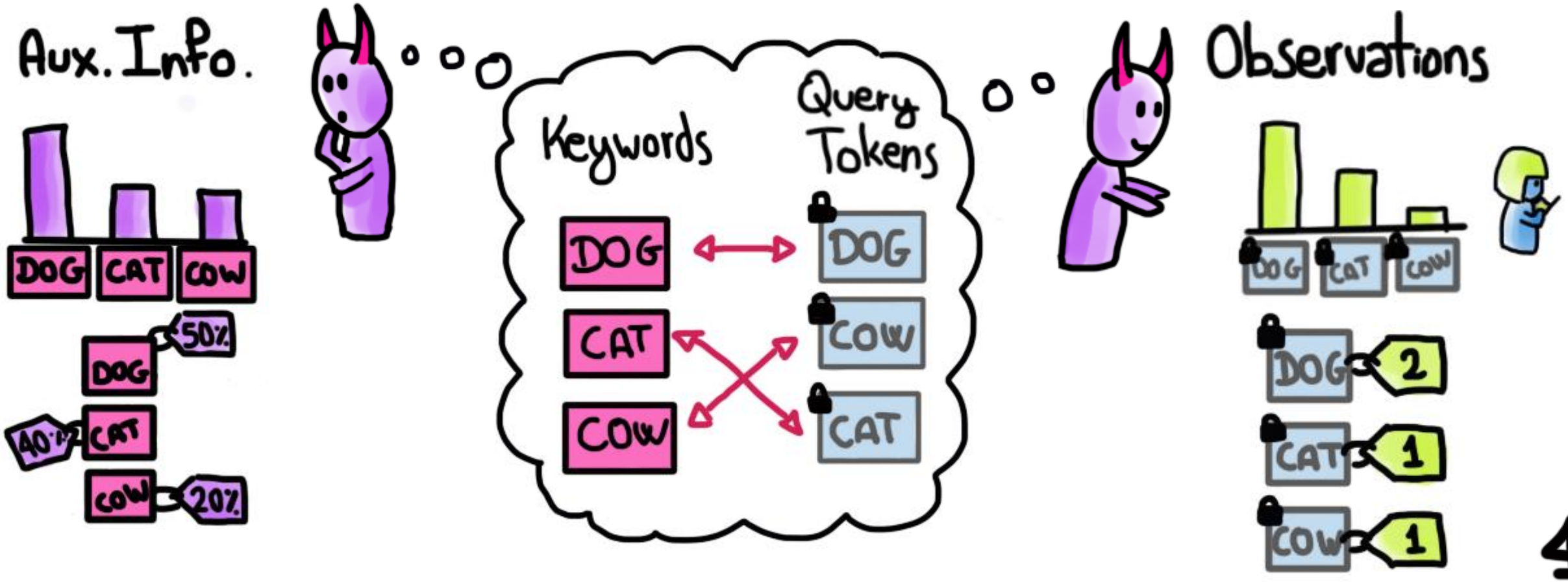
Keyword volume



New attack: SAP

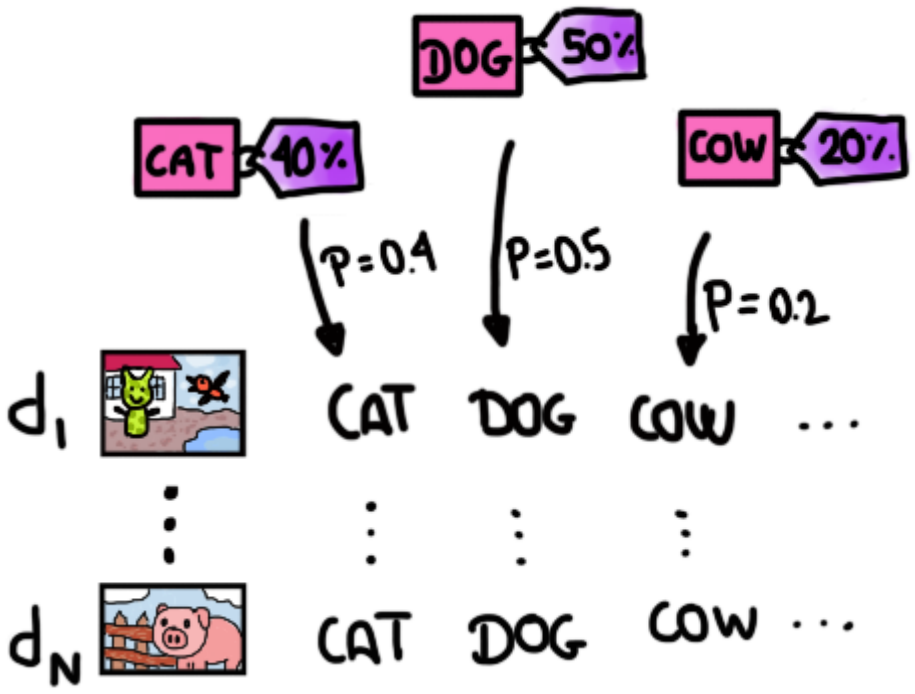
Search & Access Pattern-based attack

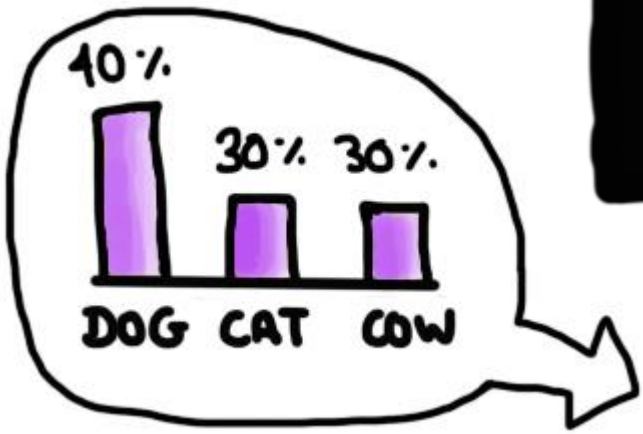
We propose a new attack (SAP) that uses both search and access pattern leakage, as well as auxiliary information (which is not necessarily ground-truth information).



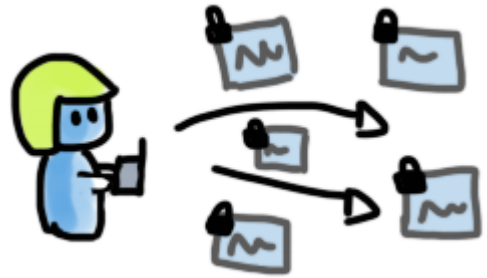
Modelling Observations

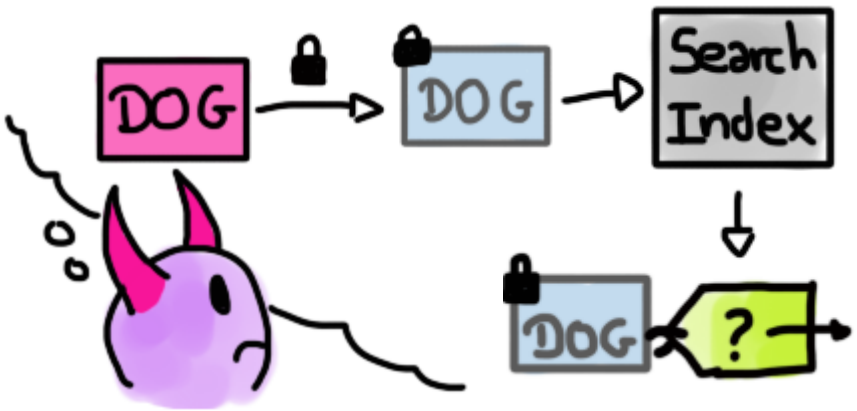
To derive the attack, we build a mathematical model of the response volume and frequency of each query, based on the auxiliary information.





η queries



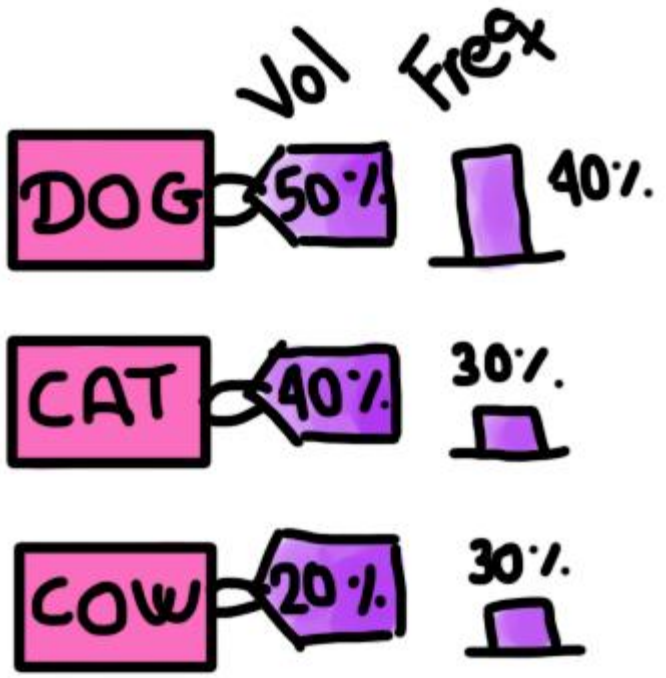


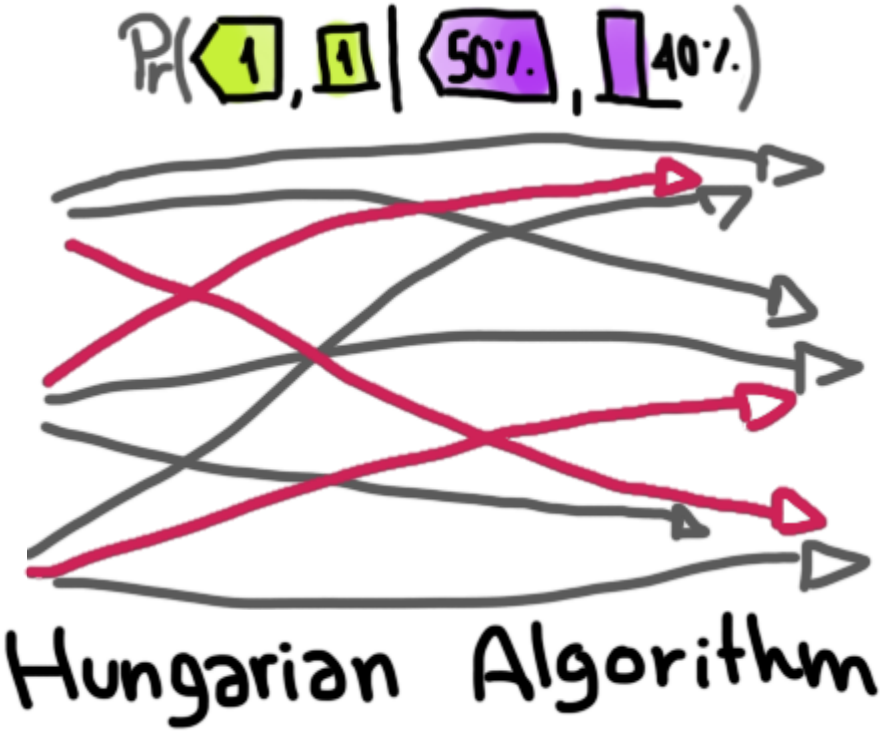
$$(\#\text{DOG}, \#\text{CAT}, \#\text{COW}) \sim \text{Multi}(\eta, (40\%, 30\%, 30\%))$$

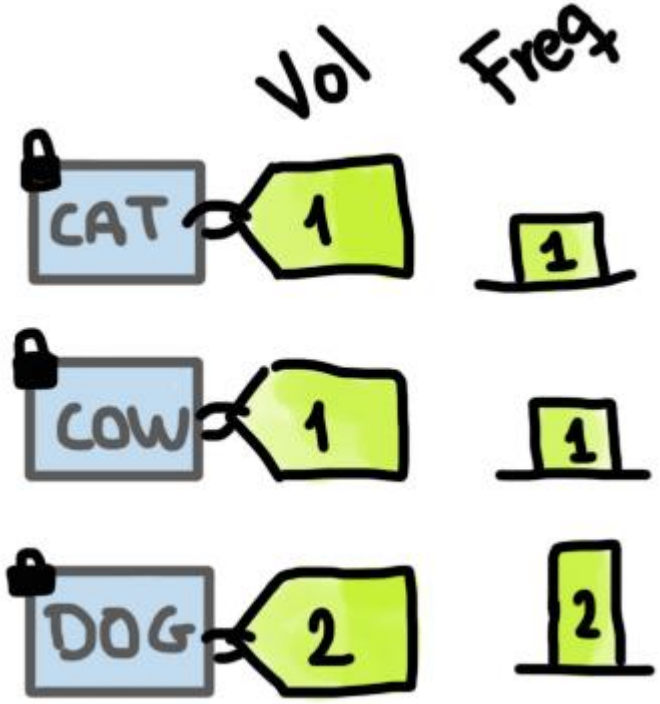
SAP uses ML (Maximum Likelihood)

$$\hat{P} = \underset{P \in \mathcal{P}}{\operatorname{argmax}} \Pr(Obs \mid Aux, P)$$

The attack finds the **maximum likelihood** matching of keywords to query tokens given the previous mathematical model. It uses the Hungarian Algorithm to find the optimal matching.



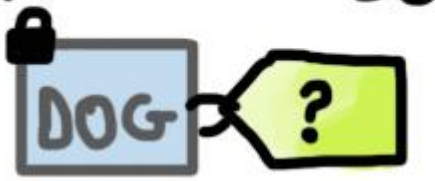




SAP vs. Defenses

SAP is easy to adapt against different volume-hiding defenses (padding) by just taking the defense into account in the mathematical model.

#docs with "DOG"



No defense:

$? \sim Bino(N, 50\%)$

CLRZ

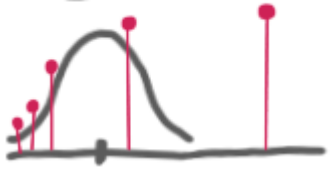
$? \sim Bino(N, 50\% \cdot TPR + (1 - 50\%) \cdot FPR)$

PPYY

$? \sim Bino(N, 50\%) + Lap(b) + c$

SEAL

$? \sim Pad_x(Bino(N, 50\%))$

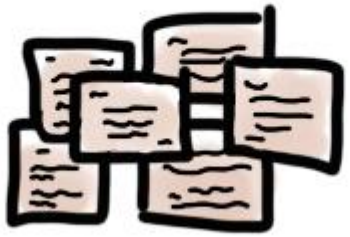


CLRZ: Chen et al. “Differentially private access patterns for searchable symmetric encryption”. INFOCOM’18
PPYY: Patel et al. “Mitigating leakage in secure cloud-hosted data structures”. CCS’19
SEAL: Demertzis et al. “SEAL: Attack mitigation for encrypted databases via adjustable leakage”. USENIX’20

Evaluation:

We evaluate SAP using real datasets (Enron and Lucene) and use query frequencies grabbed from Google Trends. We give the adversary imperfect auxiliary information to run the attack.

Real datasets



50% → Aux Info.

50% → Actual dataset

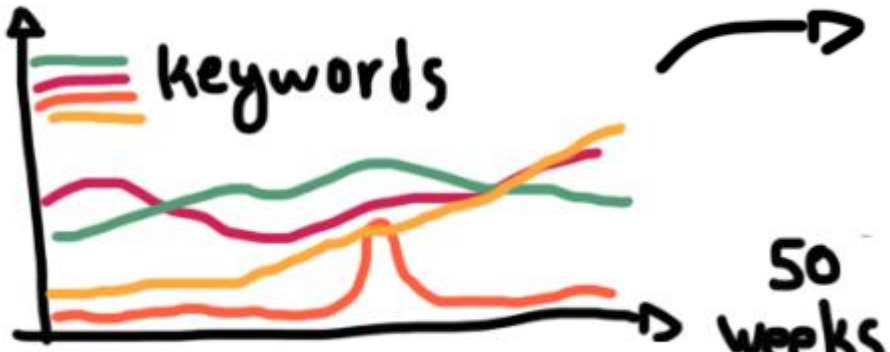
Enron: 30k emails

Lucene: 66k emails

3000 most popular keywords → Δ Enron

3000 most popular keywords → Δ Lucene

From Google Trends



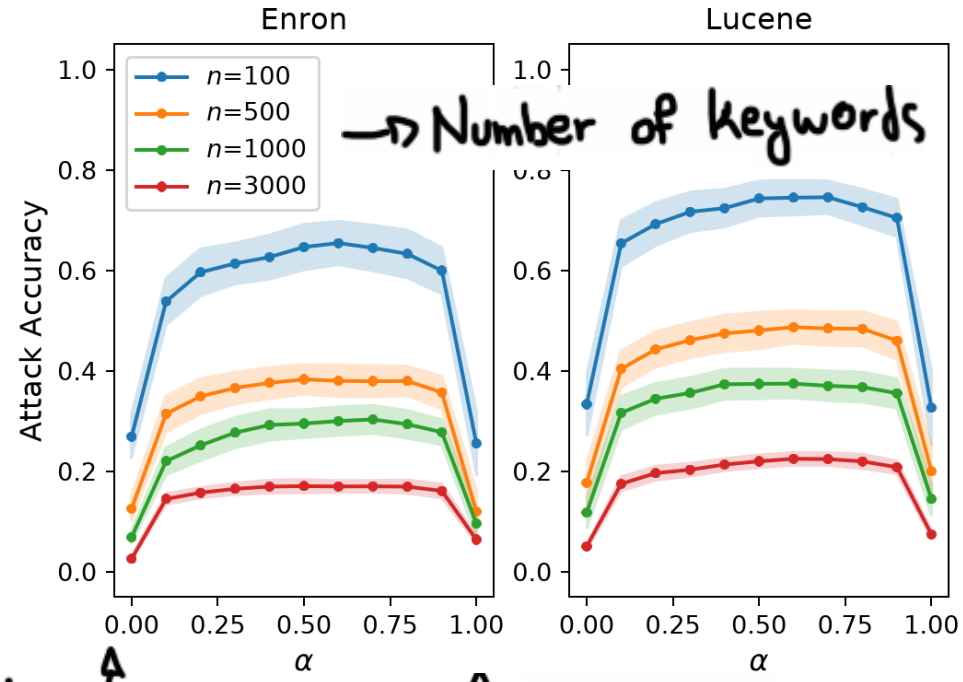
Aux info has 5-week offset

Results I

5 queries / week
50 weeks

By combining both volume and frequency information, SAP achieves high query recovery (left).
SAP outperforms current state-of-the-art attacks (right).

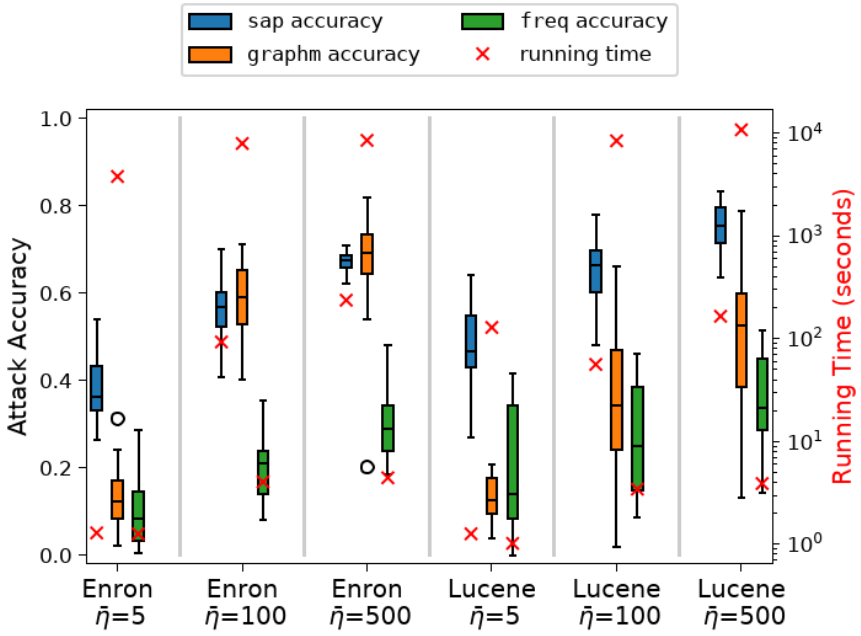
Importance of search+access pattern



Only volume (from access pattern)

Only query frequency (from search pattern)

Attack comparison

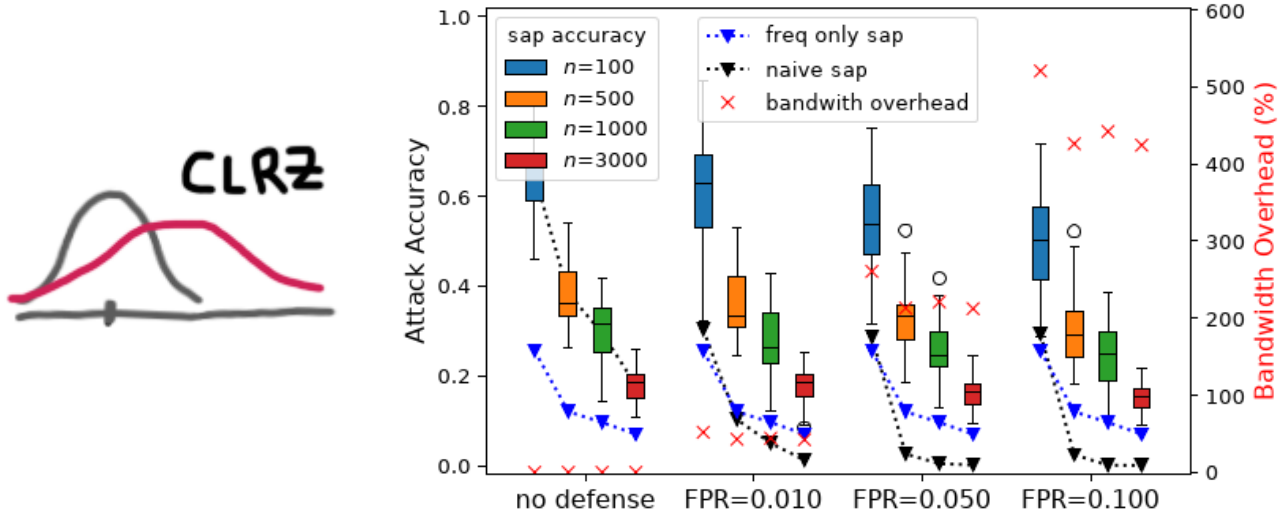


freq: Liu et al. “Search pattern leakage in searchable encryption: Attacks and new construction”. Information Sciences, 2014.
graphm: Pouliot and Wright. “The shadow nemesis: Inference attacks on efficiently deployable, efficiently searchable encryption”. CCS’16

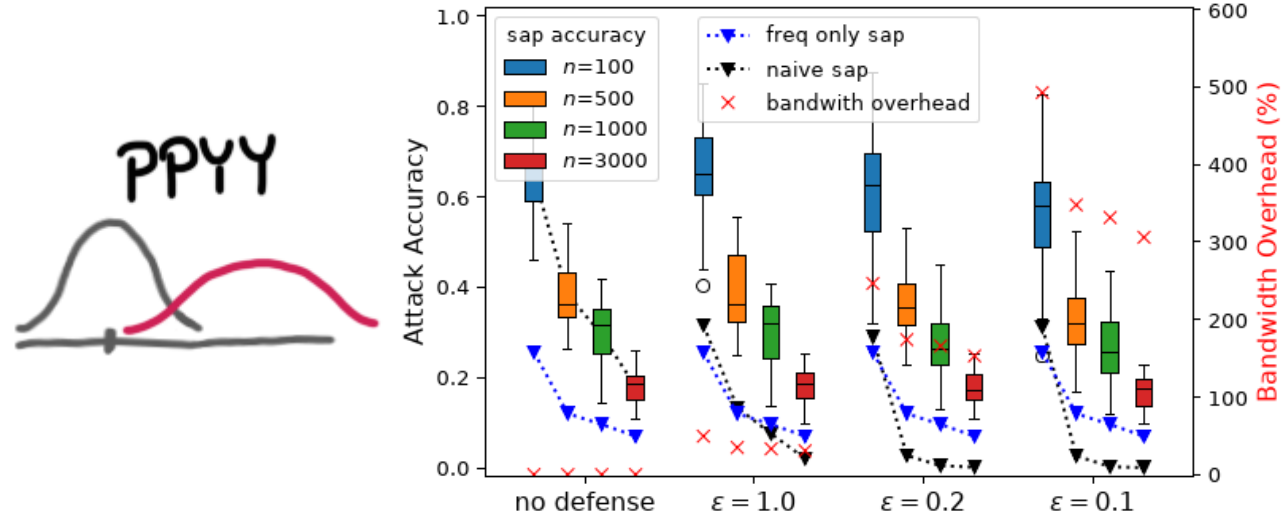
Results II

By adapting SAP against previous defenses, we are able to practically bypass two of them (CLRZ and PPYY) and we still achieve non-trivial recovery rates for the third one (SEAL).

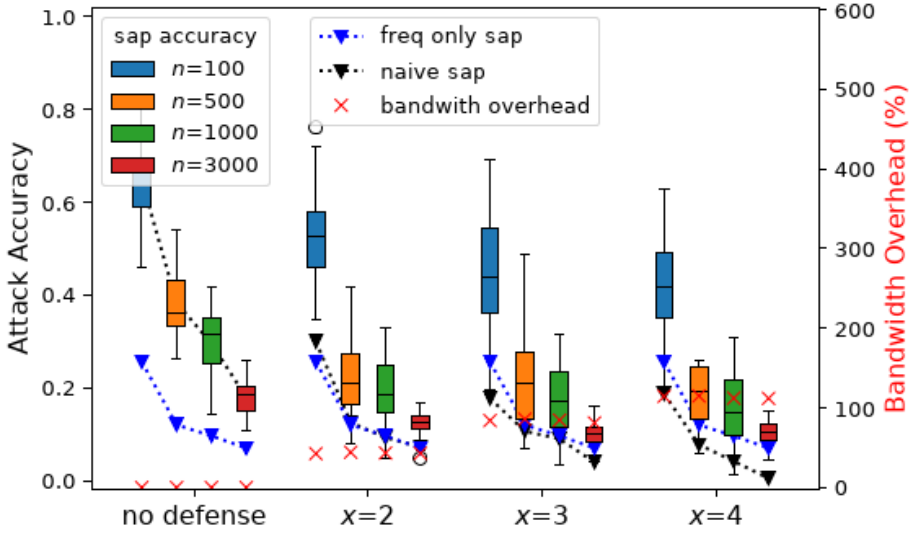
CLRZ



PPYY



SEAL



Conclusions

- SAP is efficient and strong
- Frequency + Volume leakage is dangerous
- Padding strategies are not very effective
- Hiding search pattern and/or frequency is important

Our results show the importance of hiding search patterns and/or frequency leakage. Recent works that are moving in this direction seem promising.

There is hope!

PANCAKE
Grubbs et al.
USENIX'20

OSSE
Shang et al.
NDSS'21

SWiSSSE
Gui et al.

Thank you!

simon.oya@uwaterloo.ca