Obfuscated Access and Search Patterns in Searchable Encryption

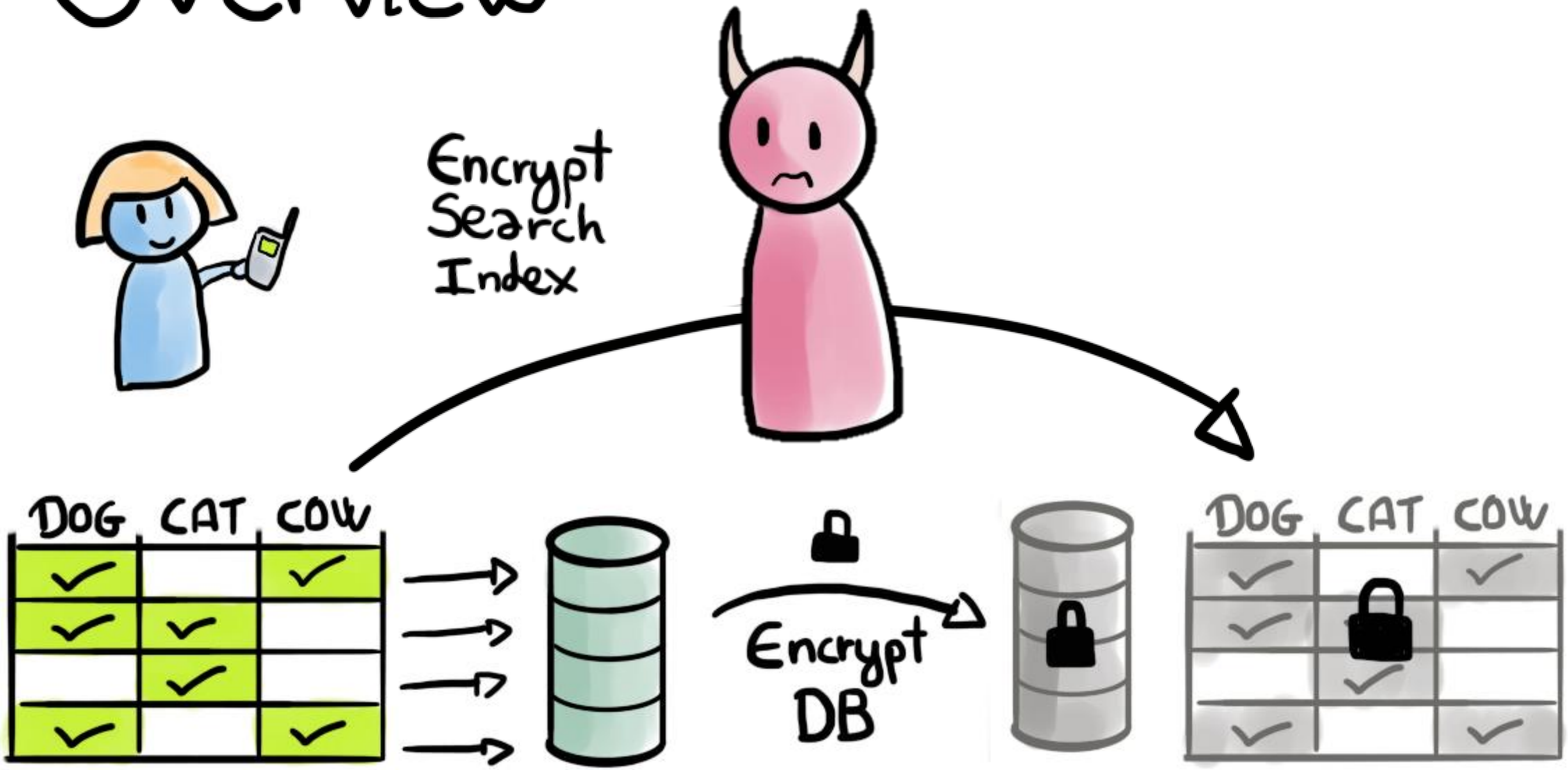
Zhiwei Shang*, Simon Oya*, Andreas Peter*, Florian Kerschbaum*

University of Waterloo

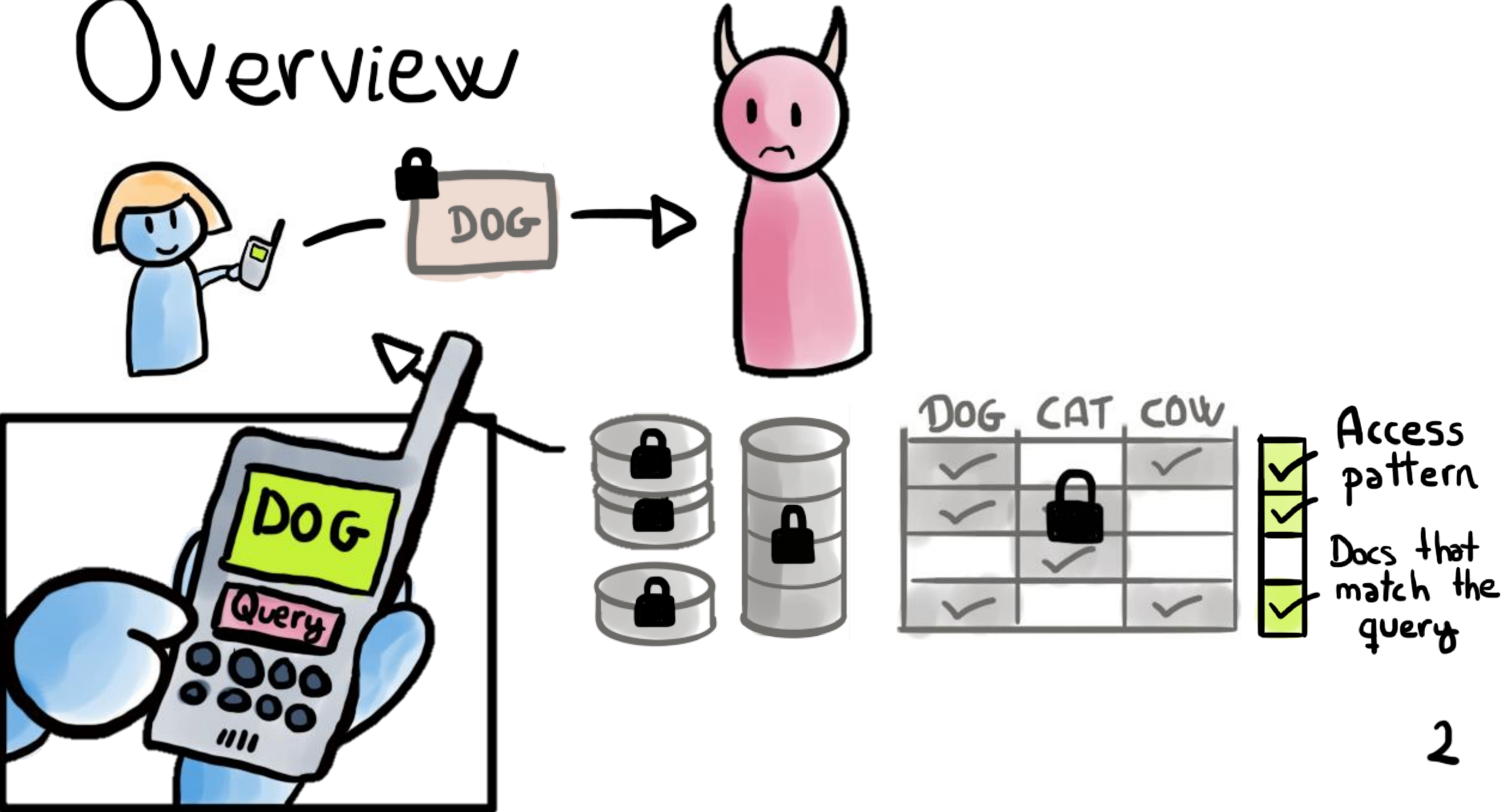
University of Twente

NDSS'21

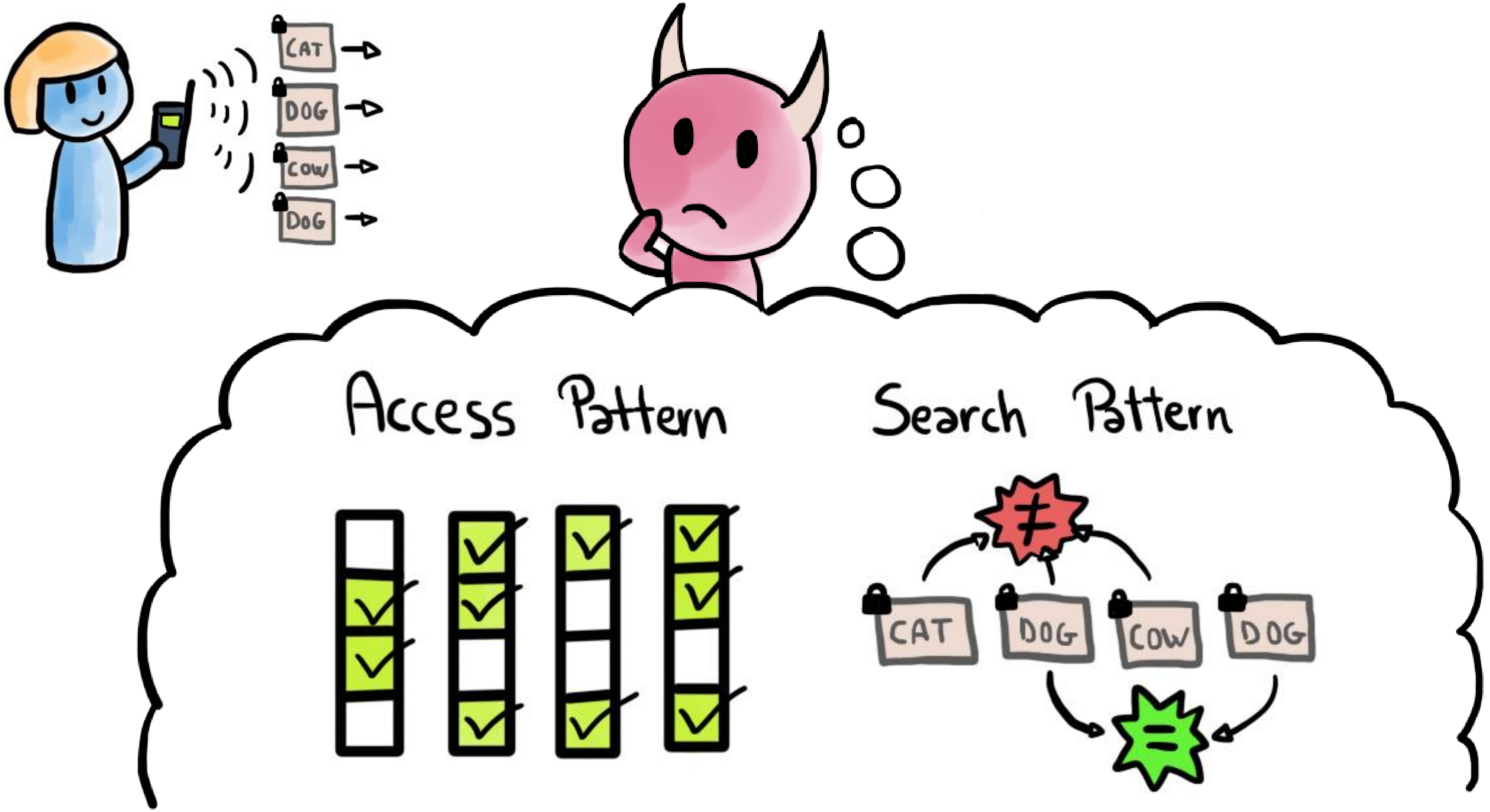
Overview



Overview

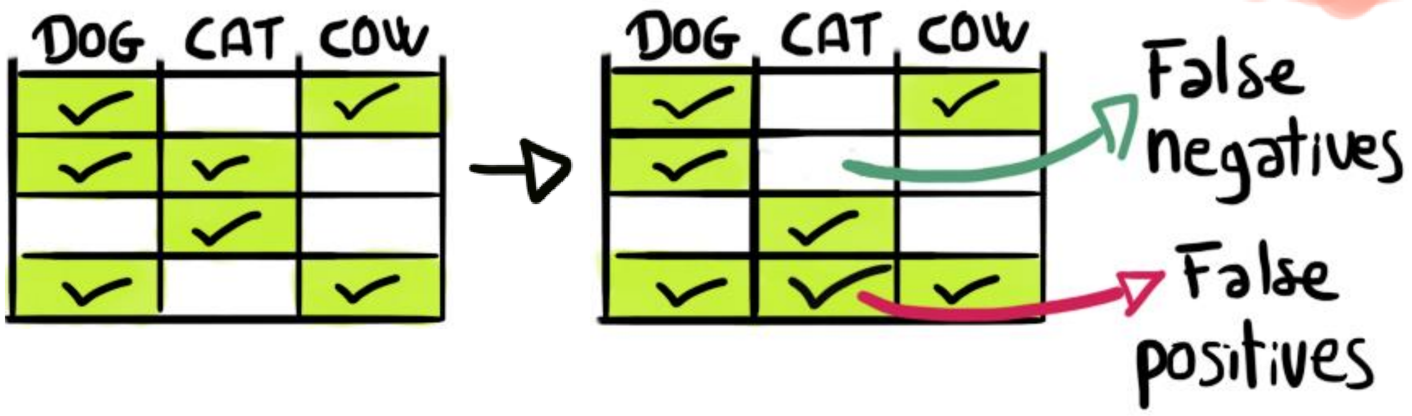

DOG
DOG
Query
DOG
CAT
COW
Access pattern
Docs that match the query

Overview


CAT
DOG
COW
DOG
Access Pattern
Search Pattern
≠
CAT
DOG
COW
DOG
=

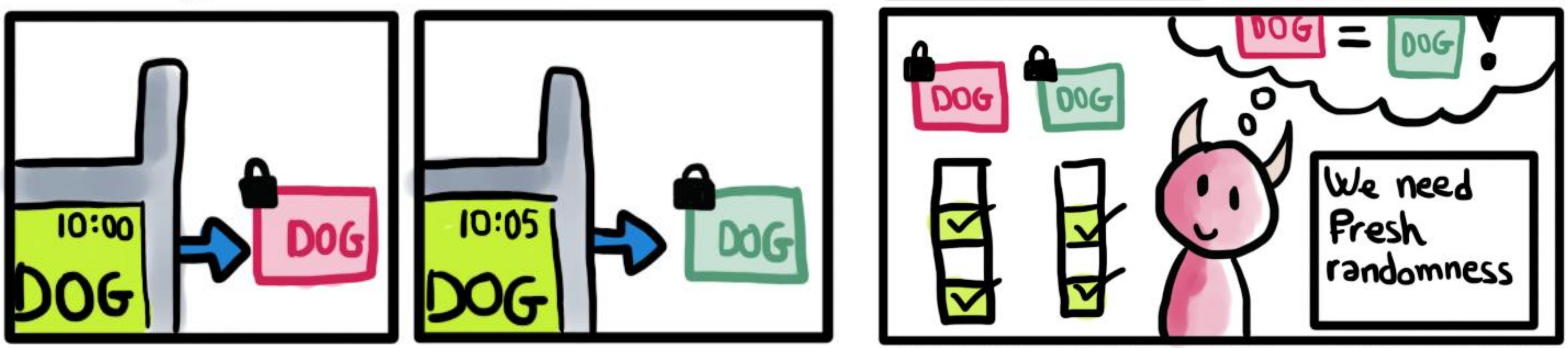
Hiding Access Pattern

CLRZ

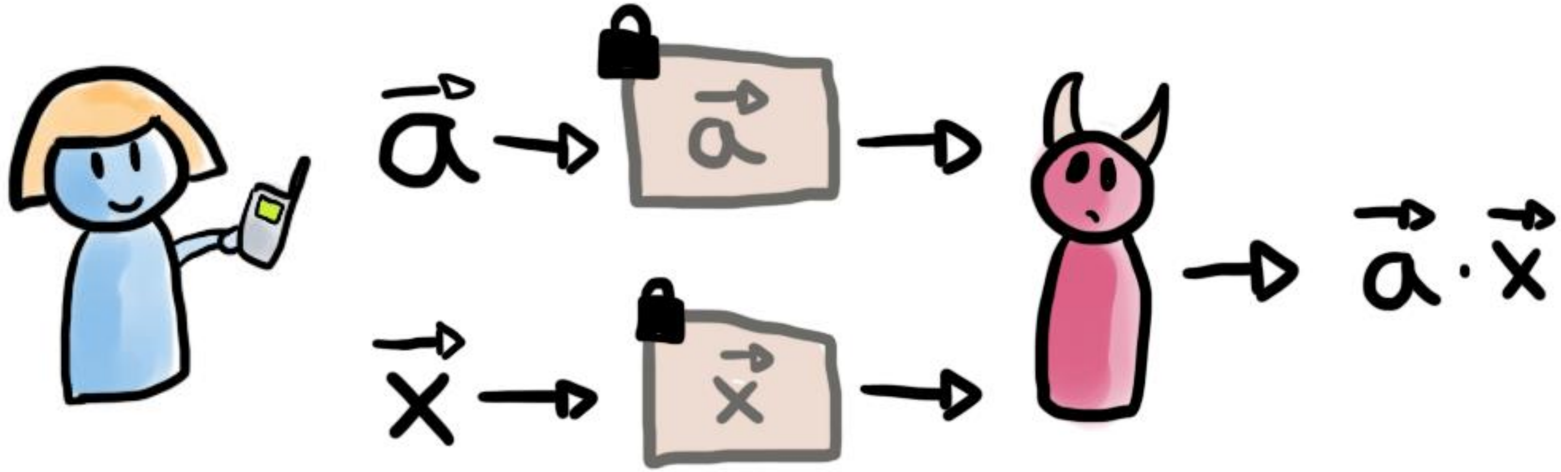


G. Chen, T.-H. Lai, M. K. Reiter, and Y. Zhang, "Differentially private access patterns for searchable symmetric encryption," in *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 2018, pp. 810–818.

Hiding Search Pattern?



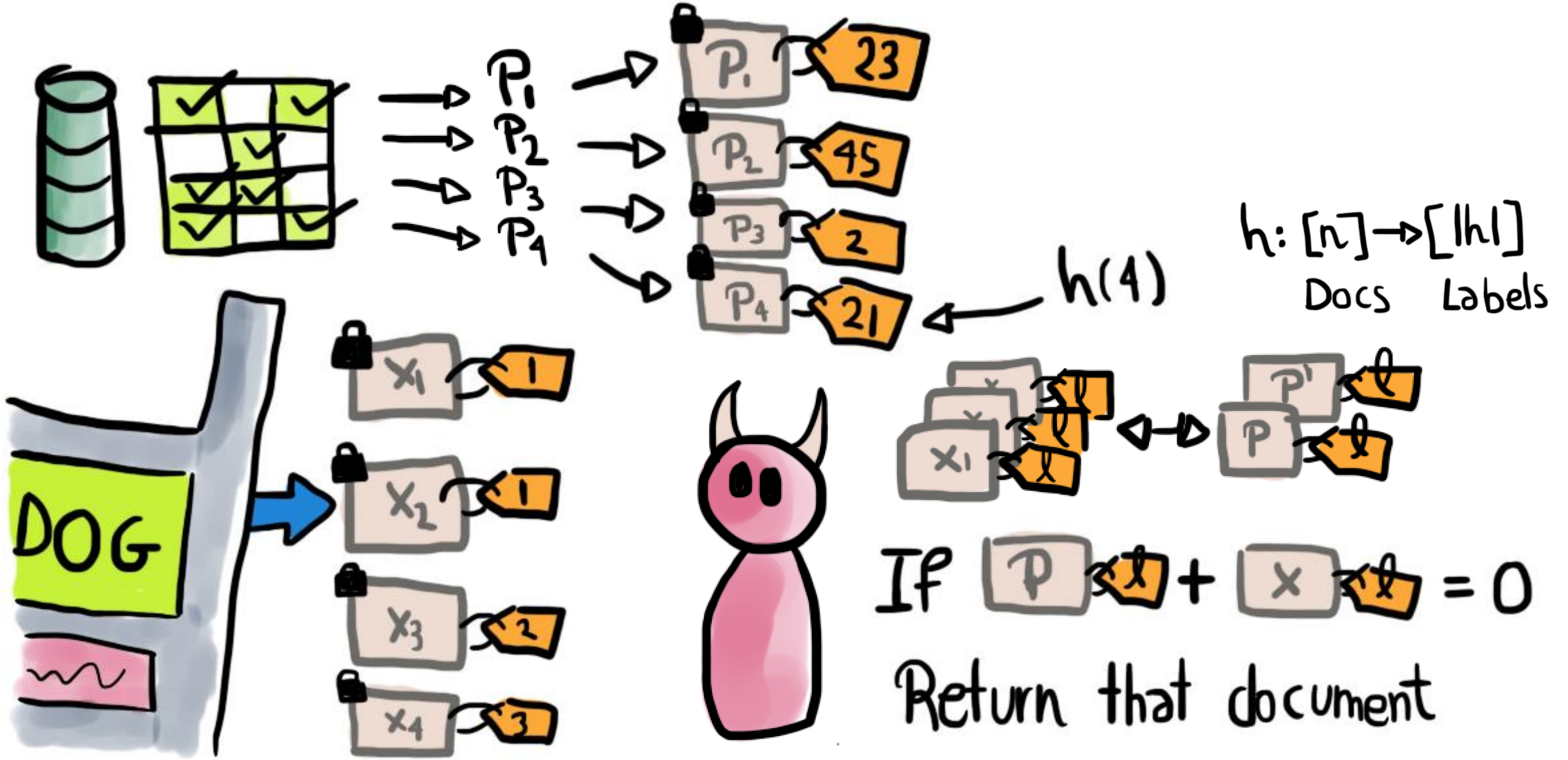
IPPE: Inner Product Predicate Encryption



$$P(x) = (x - r_1)(x - r_2)\cdots(x - r_d) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d \cdot x^d = \vec{a} \cdot \vec{x}$$

$$\vec{x} = (x^0, x^1, x^2, \ldots)$$

OSSE: Obfuscated SSE



Polynomial Generation

$r_1 = (DOG \| \ell \| 5)$ ← There are 5 $(DOG\|\ell\|\cdots)$ already

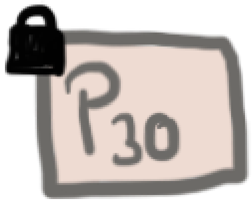
$r_2 = (COW \| \ell \| 0)$

$r_3 = (RAT \| \ell \| 1)$

$r_4 = (AAA \| \ell \| 0)$

$r_5 = (AAA \| \ell \| 0)$

$r_6 = (30 \| 0 \| -1)$



$D_{30} = \{DOG, COW, RAT\}$

$\ell = h(30)$

$S_{max} = \frac{\text{Max keywords}}{\text{per document}} = 5$

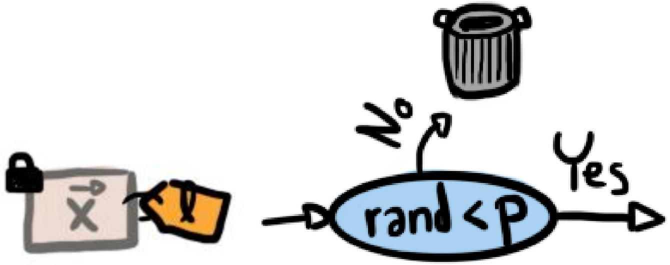
Token Generation

- **Find (database) with "DOG":**

For $\ell = 1 \rightarrow |h|$:

For $c = 0 \rightarrow c_{max}$:

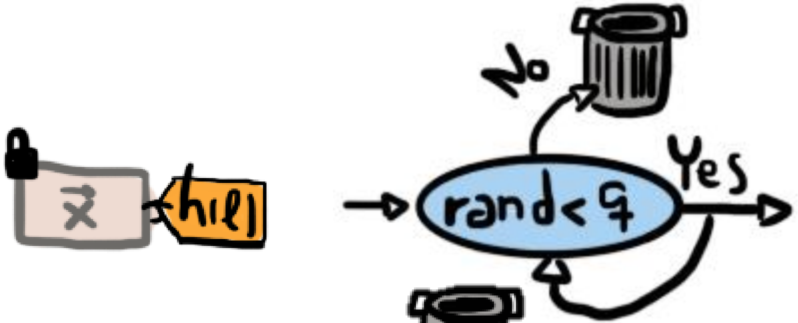
$x = (DOG \| \ell \| c) \rightarrow$

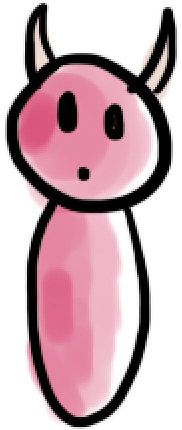


- **False positives:**

For $id = 1 \rightarrow n$:

$x = (id \| 0 \| -1) \rightarrow$

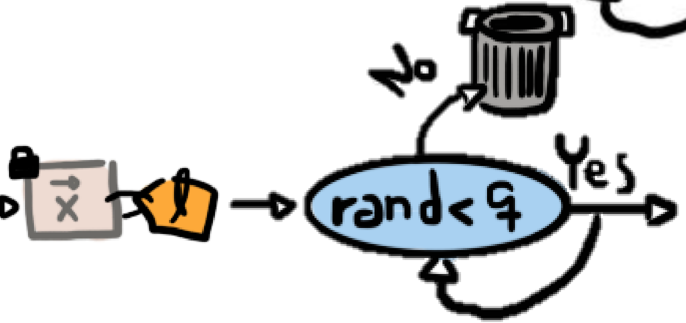




- **Non-matches:**

For $\ell = 1 \rightarrow |h|$:

$x = (AAA \| -1 \| 0) \rightarrow$



$r_1 = (DOG \| \ell \| 6)$

$r_2 = (COW \| \ell \| 0)$

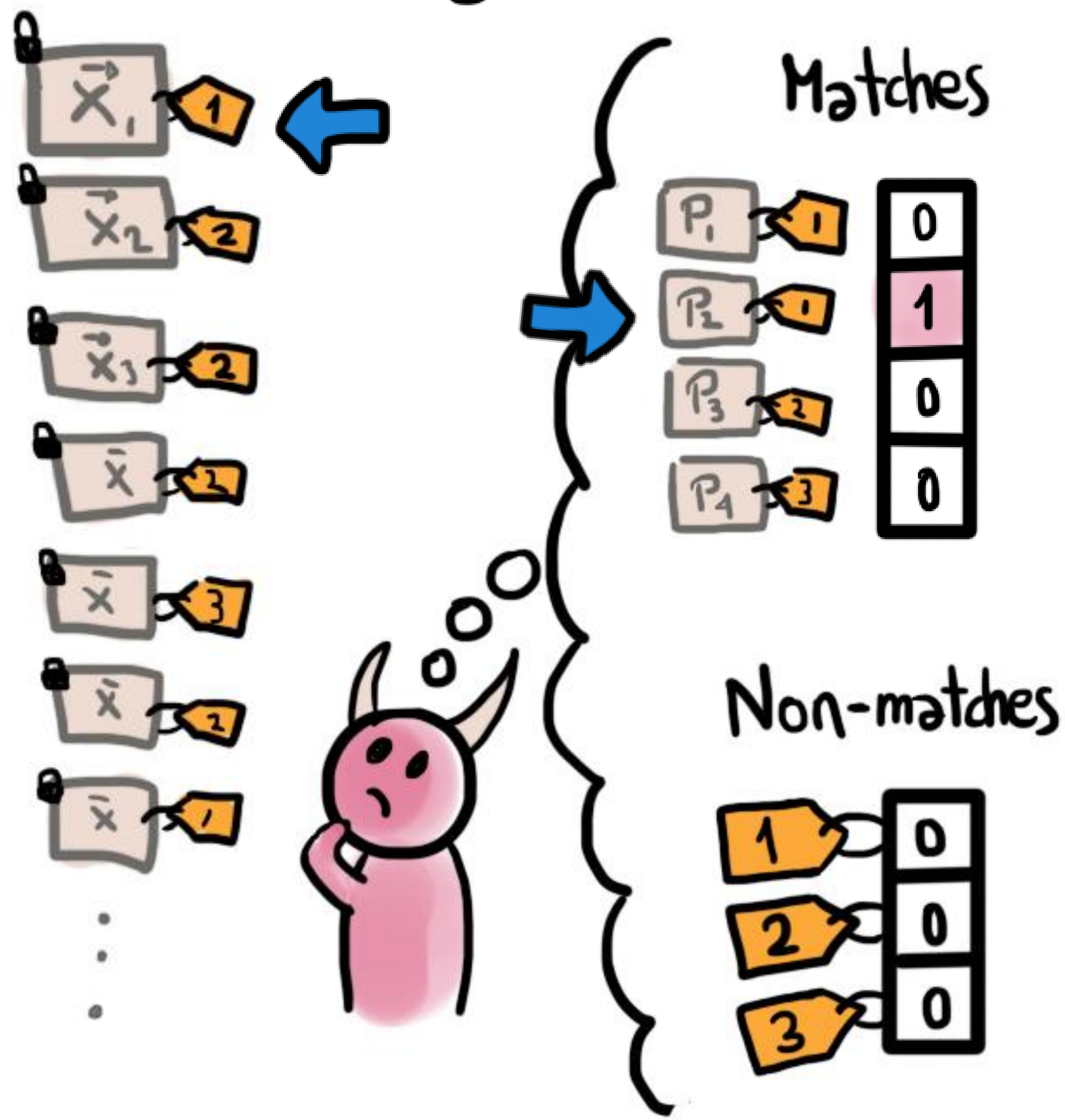
$r_3 = (RAT \| \ell \| 1)$

$r_4 = (AAA \| \ell \| 0)$

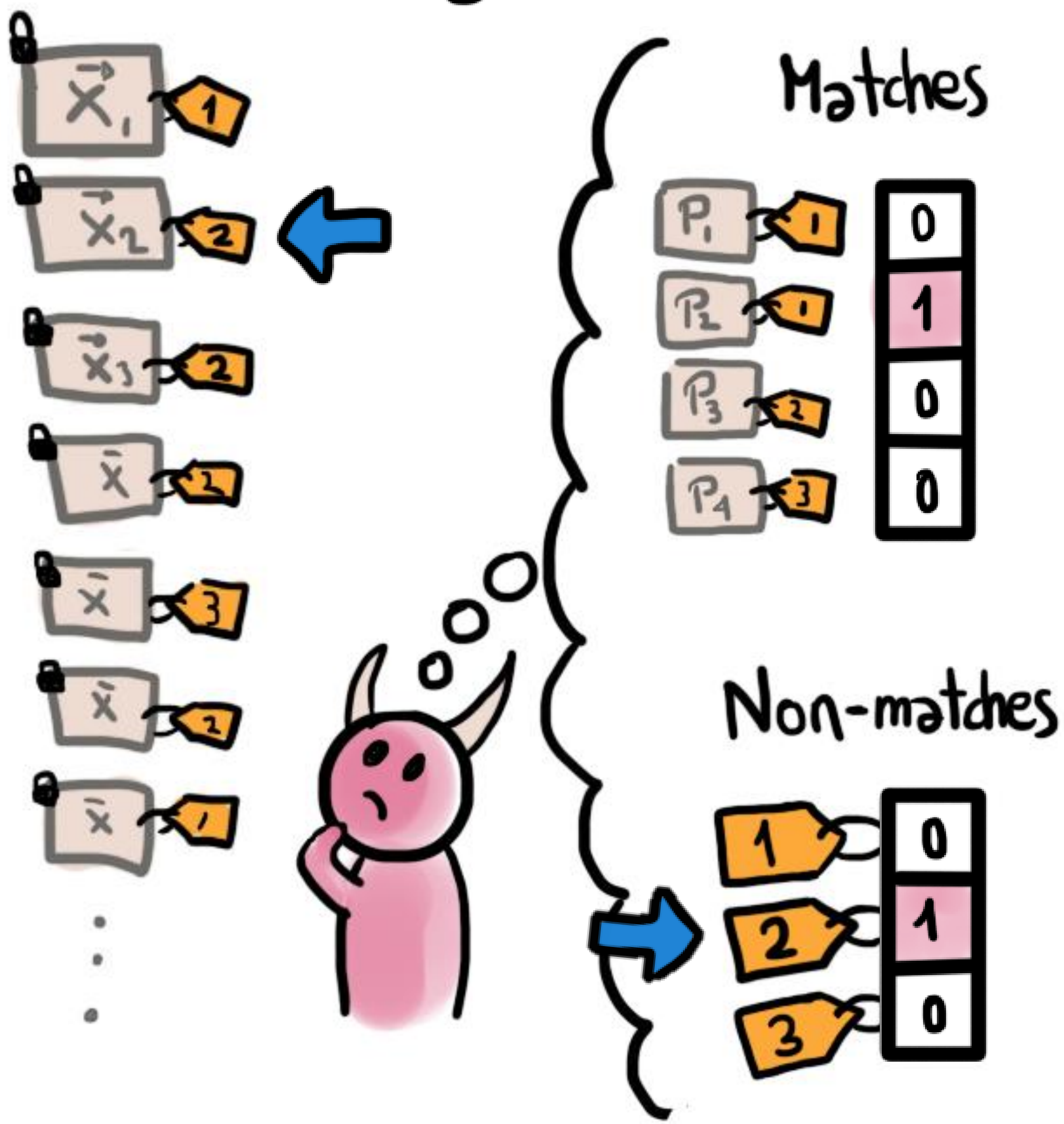
$r_5 = (AAA \| \ell \| 0)$

$r_6 = (30 \| 0 \| -1)$

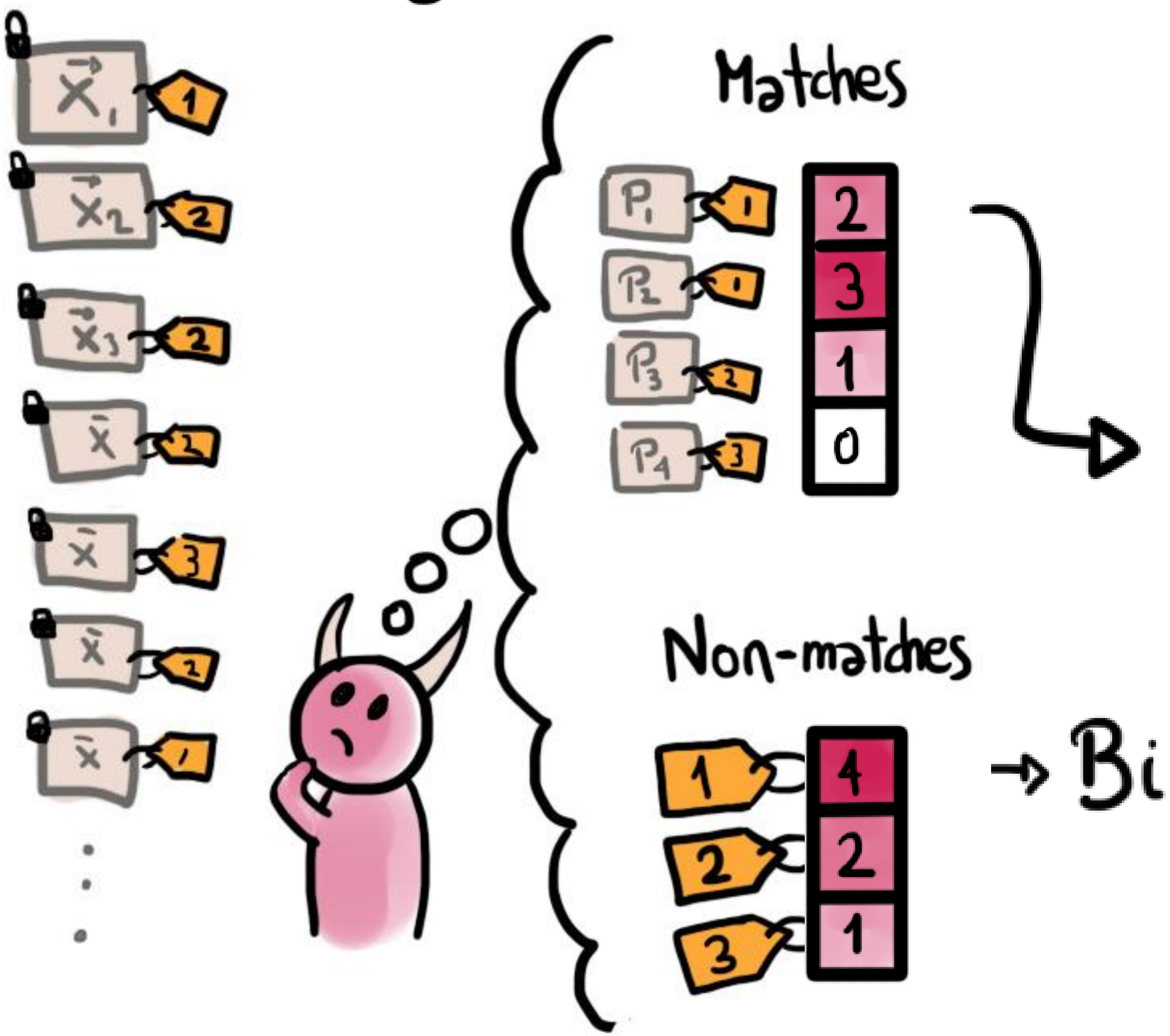
Adversary's View



Adversary's View



Adversary's View



$$\begin{cases} Ber(p) + Geo(1-q) & DOG \in D_1 \\ Geo(1-q) & DOG \notin D \end{cases}$$

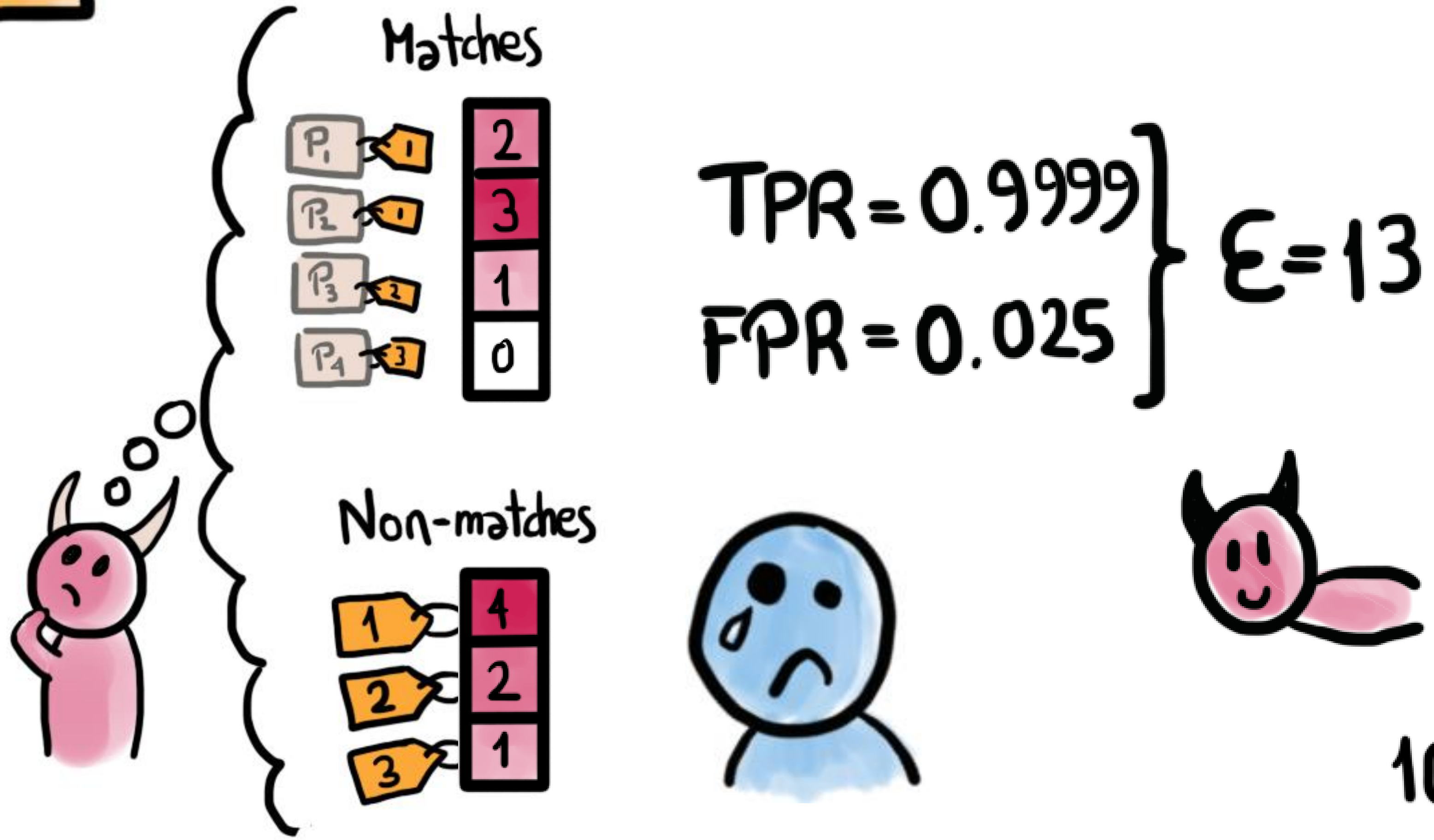
$$\rightarrow Bi(g_1, p) + Geo(1-q)$$

Security

We prove it holds by IPPE security

Differential Privacy

$$\varepsilon = \ln\left(\frac{TPR}{FPR} \cdot \frac{1-FPR}{1-TPR}\right)$$



Complexity Analysis

- Communication overhead (Zipf)

  $COMM = O(\log n_{keywords})$

  1 round

- Computational complexity

  $COMP < n \cdot (C_{max} + 1)$

- Client Storage: 

TWORAM (ORAM)

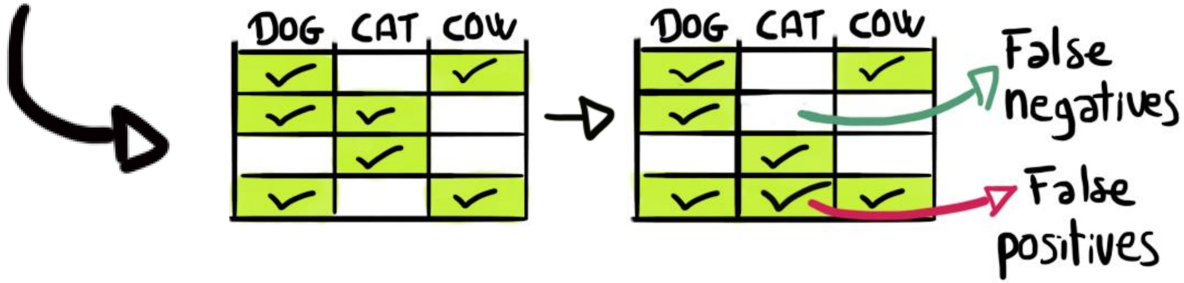
$O(\log n \cdot \log\log n)$

4 rounds at least

$O(\log^2 n)$ storage 

Evaluation:

- CLRZ vs. OSSE



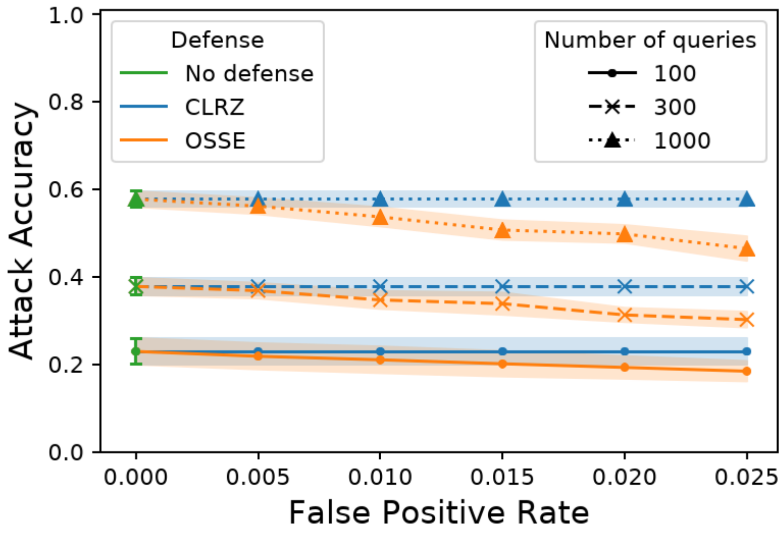
- Four different query recovery attacks
- Enron dataset
- We adapt the attacks against the defenses

Results

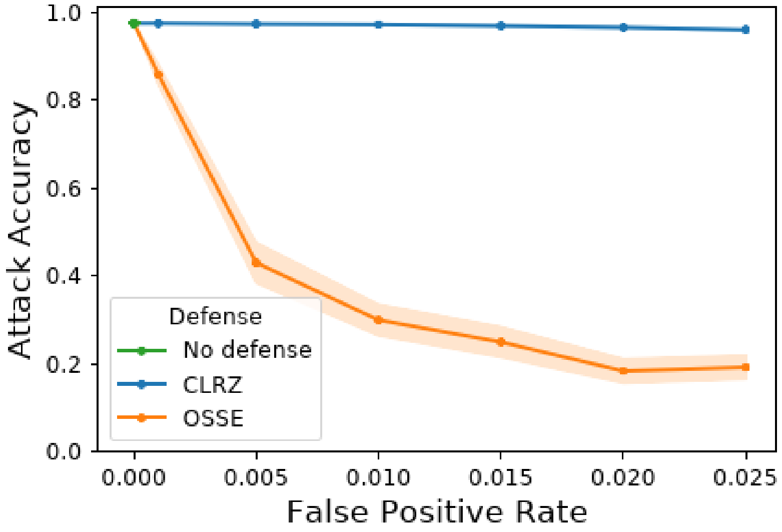
CLRZ

OSSE

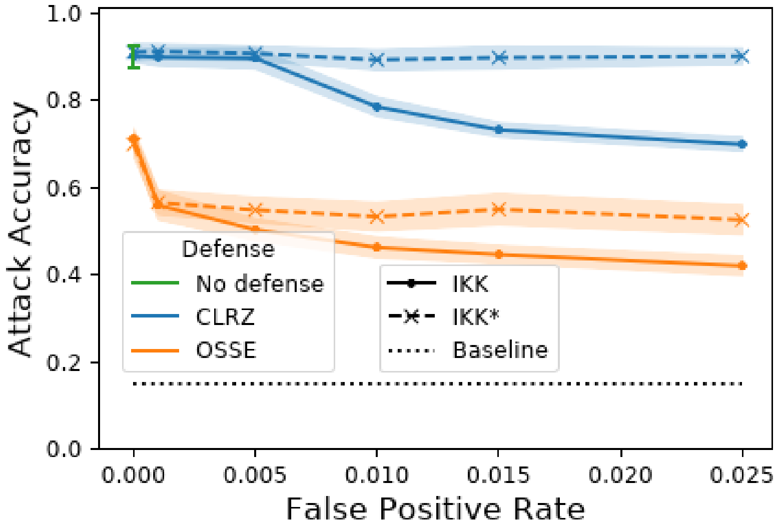
Liu et al.



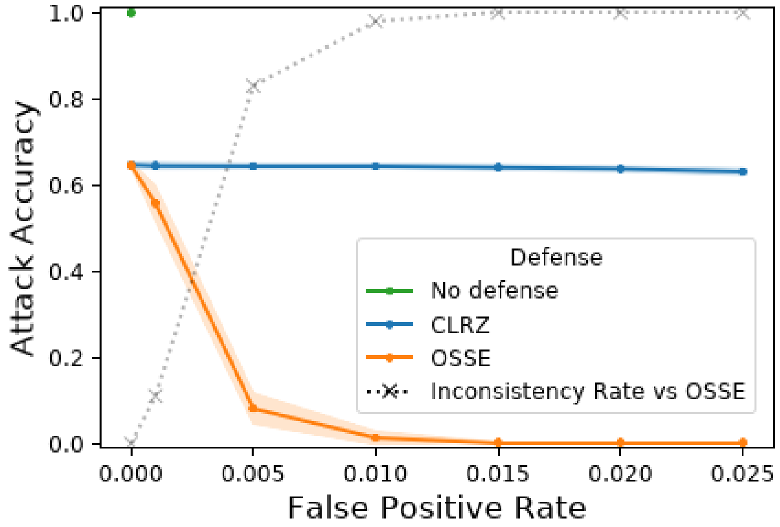
Pouliot & Wright



Islam et al.



Cash et al.



Conclusions

- Hiding search pattern is challenging but very effective against attacks!

- OSSE: SSE using IPPE

High computation

# cores	BuildIndex (min)	Trapdoor (s)	Search (min)
4	272.5	580.7	1099.1
8	136.3	290.5	549.6
16	68.2	145.3	274.8
32	34.1	72.8	137.4
64	17.1	36.4	68.7
128	8.5	18.2	34.4
160	6.9	14.7	27.5

TABLE V: Running Times

CLRZ=200 ms

Conclusions

- Hiding search pattern is challenging but very effective against attacks!

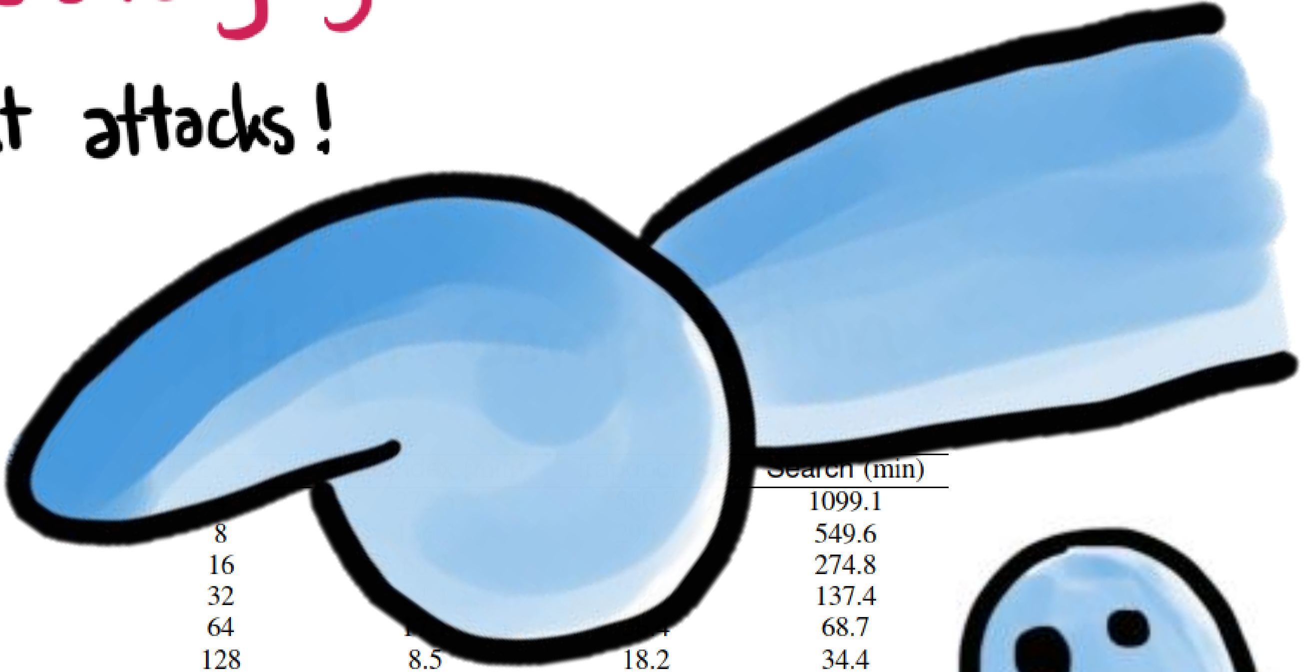
- OSSE: SSE using IPPE

1 comm round!

No client storage!

Better asymp. Comm than ORAM

Hides search pattern!



			Search (min)
			1099.1
8			549.6
16			274.8
32			137.4
64			68.7
128	8.5	18.2	34.4
160	6.9	14.7	27.5

TABLE V: Running Times

CLRZ=200 ms

Thanks!

simon.oya@uwaterloo.ca