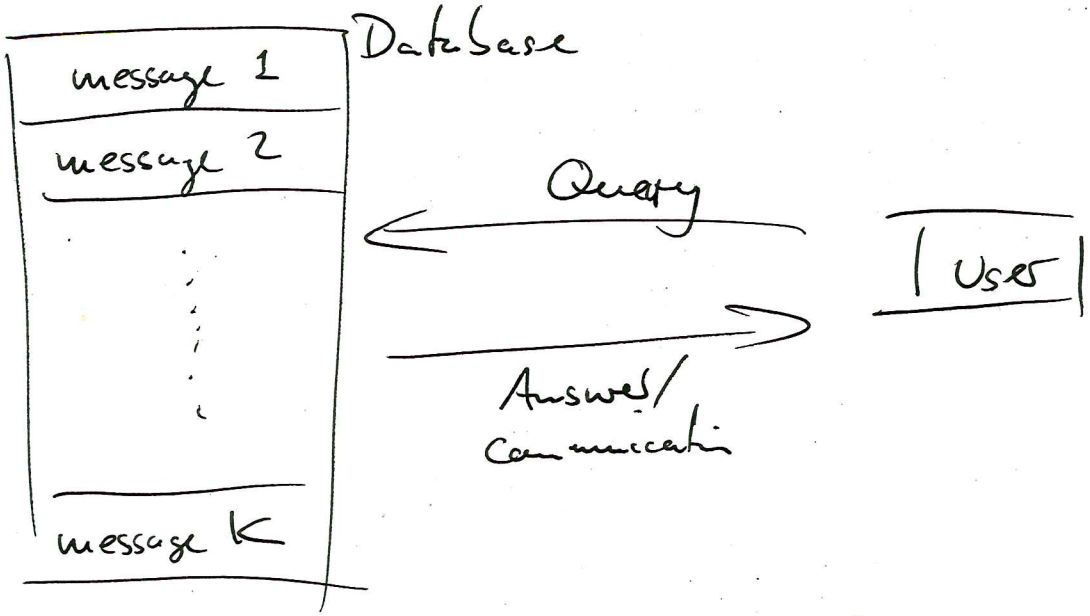
PRIVATE INFORMATION RETRIEVAL



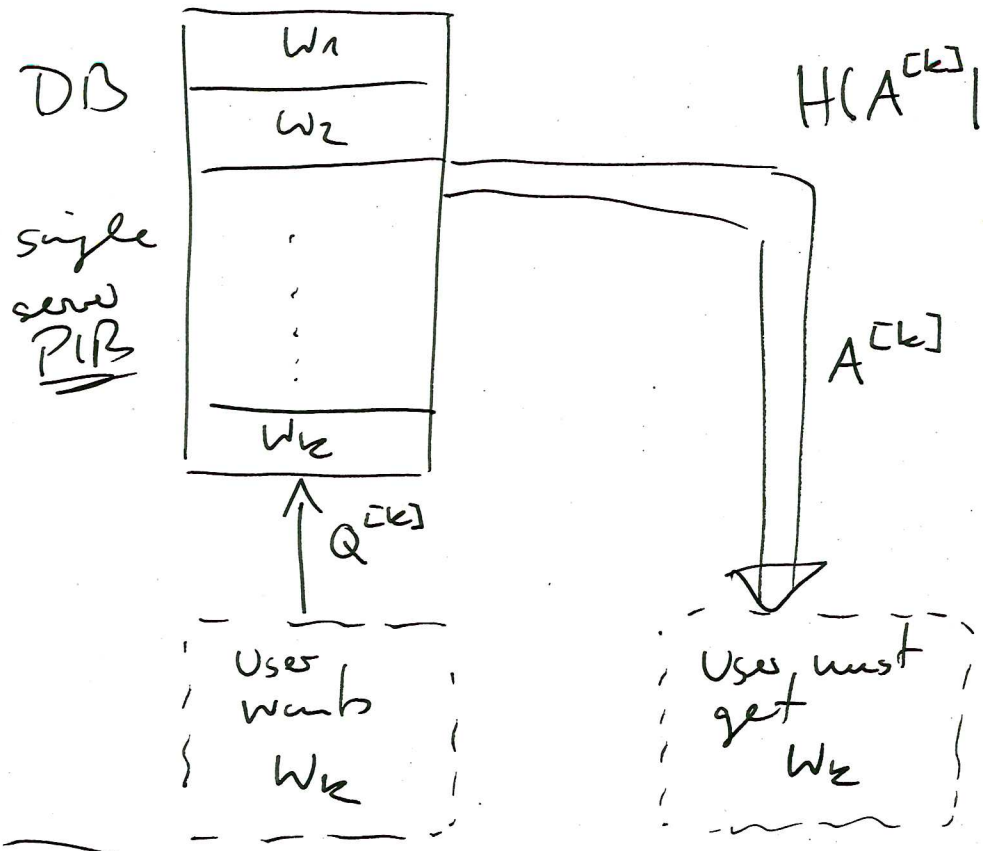
- user wishes to retrieve a single message
- user sends a Query to the DB
- Privacy : Query should not reveal to the DB which message the user is interested in!
- Correctness : User must be able to recover the desired message from the communication received from the DB.

Single Server/DB PIR

- K messages $W_1, W_2, \ldots, W_K$
- Assume messages are independent
- $H(W_1, W_2, \ldots, W_K) = H(W_1) + H(W_2) + \ldots + H(W_K)$
- $H(W_K) = L$ i.e. each message is of size L bits

→ User is interested in a message W_Θ $\Theta \in \{1, 2, \ldots, K\}$

→ User generates a query $Q^{[k]}$ when $\Theta = k$



$$H(A^{[k]} \mid Q^{[k]}, W_1, \dots, W_K) = 0$$

Privacy constraint: $(Q^{[1]}, A^{[1]}, W_1, \dots, W_K) \sim (Q^{[2]}, A^{[2]}, W_1 \dots, W_K)$

$\dots \sim (Q^{[K]}, A^{[K]}, W_1 \dots, W_K)$

or $I(\Theta; Q^{[Q]}, A^{[Q]}, W_1, \dots, W_K) = 0$

Correctness constraint: $\dfrac{H(W_K \mid A^{[k]}, Q^{[k]})}{L} = \underbrace{\dfrac{o(L)}{L}}_{\to 0 \text{ as } L \to \infty}$ (Fano's inequality)

(W_k must be "decodable" from $A^{[k]}$ and $Q^{[k]}$)

$$\text{Communication Cost of PIR} = \frac{\text{Total \# of bits downloaded}}{\text{Total \# of bits requested}} = \frac{H(A^{[L]})}{L} = \frac{D}{L}$$

$$\text{Communication Cost} = \frac{D}{L} \quad \begin{array}{l} \rightsquigarrow \text{\# of downloaded bits} \\ \rightsquigarrow \text{\# size of a message} \end{array}$$

Capacity of PIR $C^* = \frac{L}{D}$

Single Server PIR

$$C^* = \frac{1}{K}$$

Achievability:

- Users request all K messages

→ Server sends KL bits

$$\Rightarrow D = KL$$

$$\rightarrow \quad C^* \geq \frac{L}{KL} = \frac{1}{K}$$

Converse:

Q: Can we do better than the above trivial scheme for single server PIR?

A: No, we will now show that

$$C^* \leq \frac{1}{K}$$

- We have to show that $C^* \leq \frac{1}{K}$ is satisfied for single server PIR

Lemma

$$I(W_{[2:K]}; Q^{[1]}, A^{[1]} \mid W_1) \leq D - L + O(L) \cdot L$$

Proof:

$$I(W_{[2:K]}; Q^{[1]}, A^{[1]} \mid W_1)$$

$$= I(W_{[2:K]}; Q^{[1]}, A^{[1]}, W_1) \quad \text{(since } W_1 \perp W_{[2:K]} \text{ independent)}$$

$$= I(W_{[2:K]}; Q^{[1]}, A^{[1]}) + I(W_{[2:K]}; W_1 \mid Q^{[1]} A^{[1]})$$

$$\leq I(W_{[2:K]}; Q^{[1]}, A^{[1]}) + H(W_1 \mid Q^{[1]}, A^{[1]})$$

$$\leq I(W_{[2:K]}; Q^{[1]}, A^{[1]} + O(L) \cdot L \quad \text{(correctness / Fano)}$$

$$= \underbrace{I(W_{[2:K]}; Q^{[1]})}_{=0 \text{ (since queries are independent of message content)}} + I(W_{[2:K]}; A^{[1]} \mid Q^{[1]}] + O(L) \cdot L$$

$$\leq H(A^{[1]} \mid Q^{[1]}) - H(A^{[1]} \mid Q^{[1]}, W_{[2:K]}) + O(L) \cdot L$$

$$\leq H(A^{[1]}) - H(W_1, A^{[1]} \mid Q^{[1]}, W_{[2:K]}) + \underbrace{H(W_1 \mid A^{[1]}, Q^{[1]}, W_{[2:K]})}_{\leq O(L) \cdot L} + O(L) \cdot L$$

$$\leq D - H(W_1, A^{[1]} \mid Q^{[1]}, W_{[2:K]}) + L \cdot O(L)$$

$$= D - \underbrace{H(W_1 \mid Q^{[1]}, W_{[2:K]})}_{= H(W_1) = L} - \underbrace{H(A^{[1]}, W_1 W_{[2:K]})}_{\geq 0} + L \cdot O(L)$$

$$\Rightarrow I(W_{[2:K]}; Q^{[1]}, A^{[1]} | W_1) \leq D - L + L\cdot o(L).$$

- We will now lower bound the above term (recursive bounding)

$$I(W_{[2:K]}; Q^{[1]}, A^{[1]} | W_1)$$

$$= I(W_{[2:K]}; Q^{[2]}, A^{[2]} | W_1) \quad \text{(privacy constraint)}$$

$$= I(W_2; Q^{[2]}, A^{[2]} | W_1) + I(W_{[3:K]}; Q^{[2]}, A^{[2]} | W_1 W_2)$$

$$= \underbrace{H(W_2|W_1)}_{=L} - \underbrace{H(W_2 | Q^{[2]}, A^{[2]}, W_1)}_{\leq L\cdot o(L)} + I(W_{[3:K]}; Q^{[2]} A^{[2]} | W_1$$

$$\geq L - L\cdot o(L) + I(W_{[3:K]}; Q^{[2]}, A^{[2]} | W_{[1:2]})$$

$$\geq (L - L\cdot o(L)) + (L - L\cdot o(L)) + I(W_{[4:K]}; Q^{[3]} A^{[3]} | W_{[1:3]}$$

$$\vdots$$

applying the same bit recursively (i.e. using privacy constraint)

$$\geq (K-1)\cdot L - L\cdot o(L)$$

$$\Rightarrow \text{we have}$$

$$(K-1)\cdot L - L\cdot o(L) \leq I(W_{[2:K]}; Q^{[1]} A^{[1]} | W_1)$$

$$\leq D - L + L\cdot o(L)$$

$$\Rightarrow D \geq K \cdot L \quad - L \cdot o(L)$$

$$\frac{D}{L} \geq K - o(L) \Rightarrow \lim_{L \to \infty} \frac{D}{L} \geq K$$

$$\Rightarrow \lim_{L \to \infty} \frac{D}{L} \geq K$$

$$\Rightarrow \boxed{C^* = \lim_{L \to \infty} \frac{L}{D} \leq \frac{1}{K}}$$

$$\Rightarrow \boxed{C^* = \frac{1}{K}} \quad \text{for single server PIR}$$