



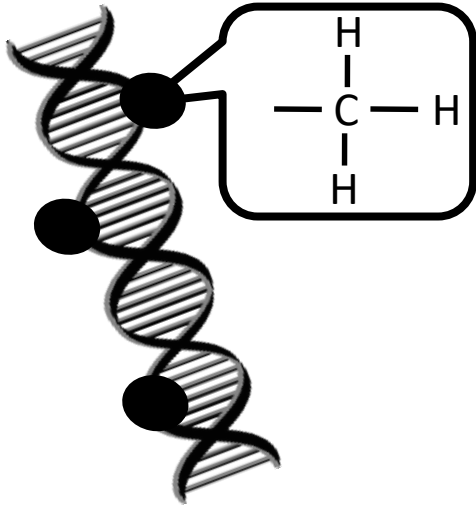
CISPA

HELMHOLTZ CENTER FOR
INFORMATION SECURITY

MBeacon: Privacy-Preserving Beacons for DNA Methylation Data

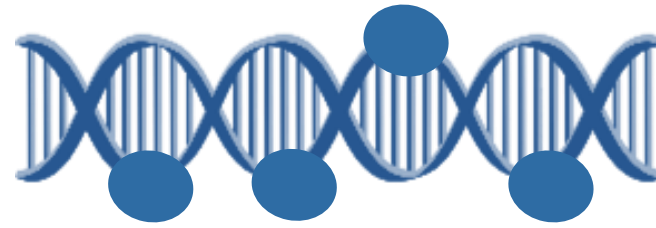
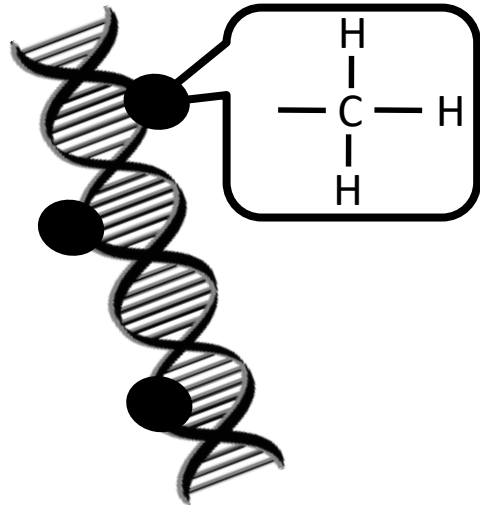
Inken Hagestedt, Yang Zhang, Mathias Humbert, Pascal Berrang, Haixu Tang, XiaoFeng Wang, Michael Backes

CISPA Helmholtz Center for Information Security, Swiss Data Science Center, ETH Zurich & EPFL, Indiana University Bloomington



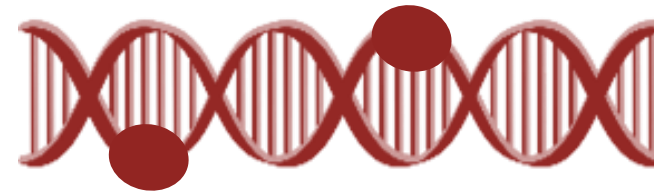
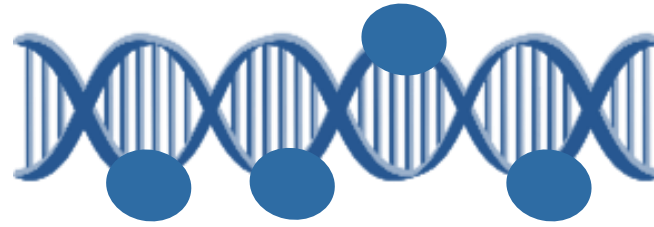
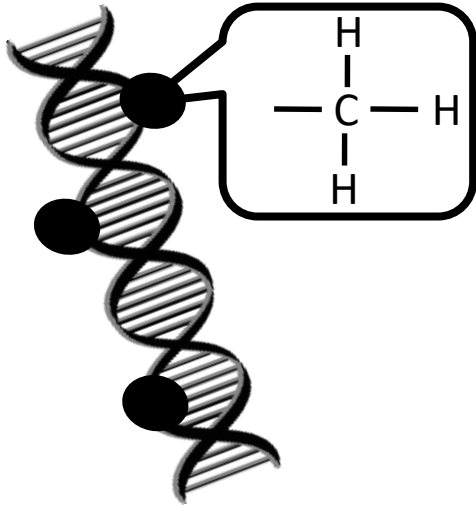
- many molecules influence cell life
- most important molecule: methyl group added to DNA
- methylation changes how DNA can be copied in the cell

Methylation Data



different tissue type

Methylation Data: Interesting for Biomedical Research

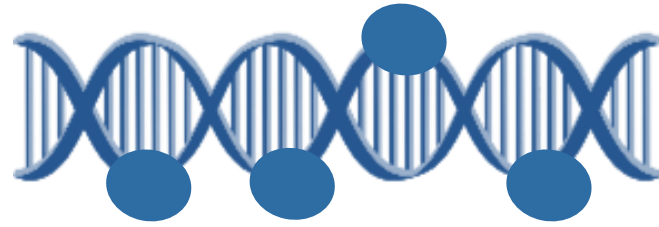
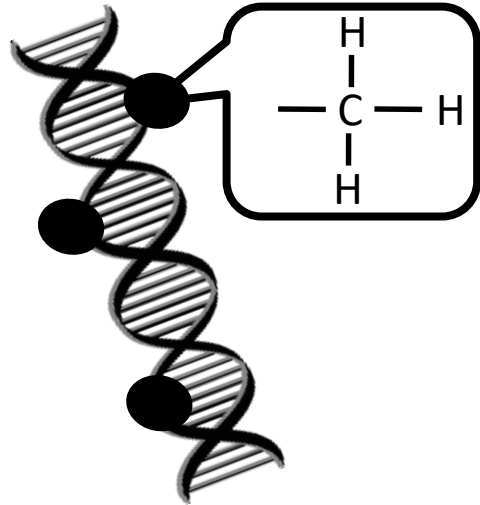


different tissue type

different environment

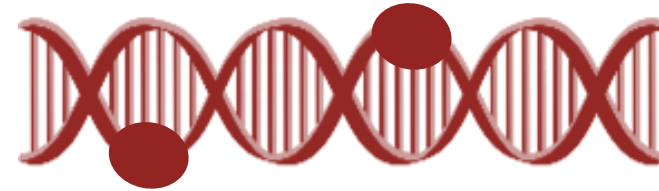
disease

Methylation Data: Interesting for Privacy Research



smokes?

stressed?



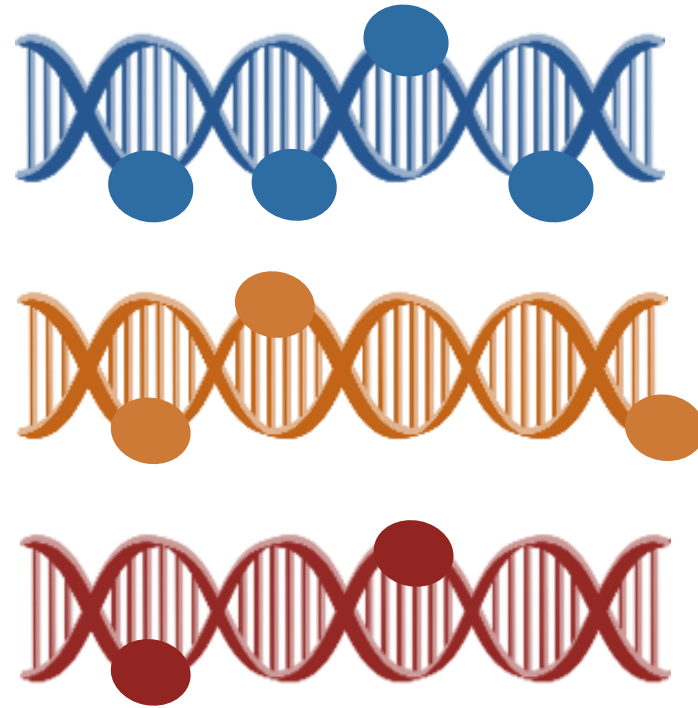
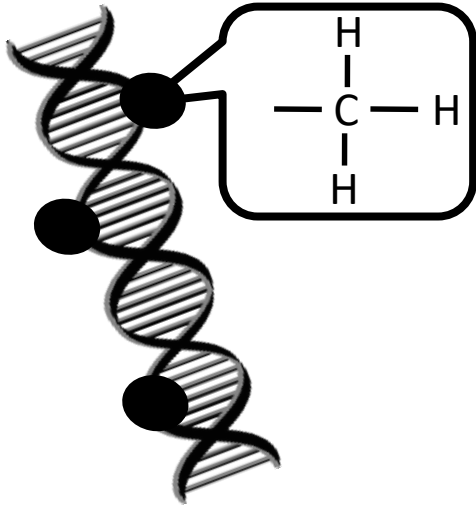
different environment

has cancer?

disease



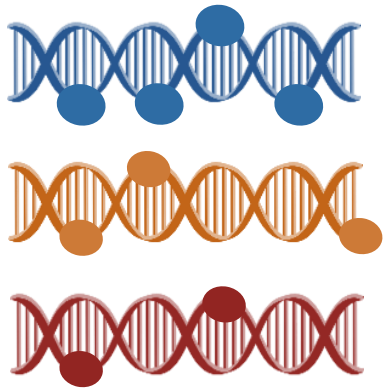
Methylation Data Format



[...0.0, 1.0, 0.6, 0.6, 0.3, 0.3 ...]

$\in R_{[0,1]}^n \quad n = 450000$

Motivation



\$\$

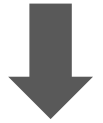
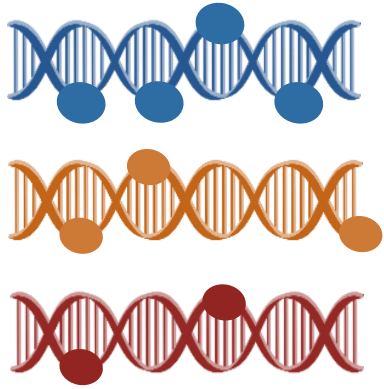
methylation sequencing:
costly process



Research
Institute

[...0.0, 1.0, 0.6, 0.6, 0.3, 0.3 ...]

Motivation: The Problem



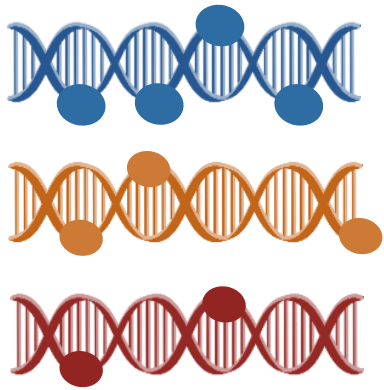
\$\$

methylation sequencing:
costly process

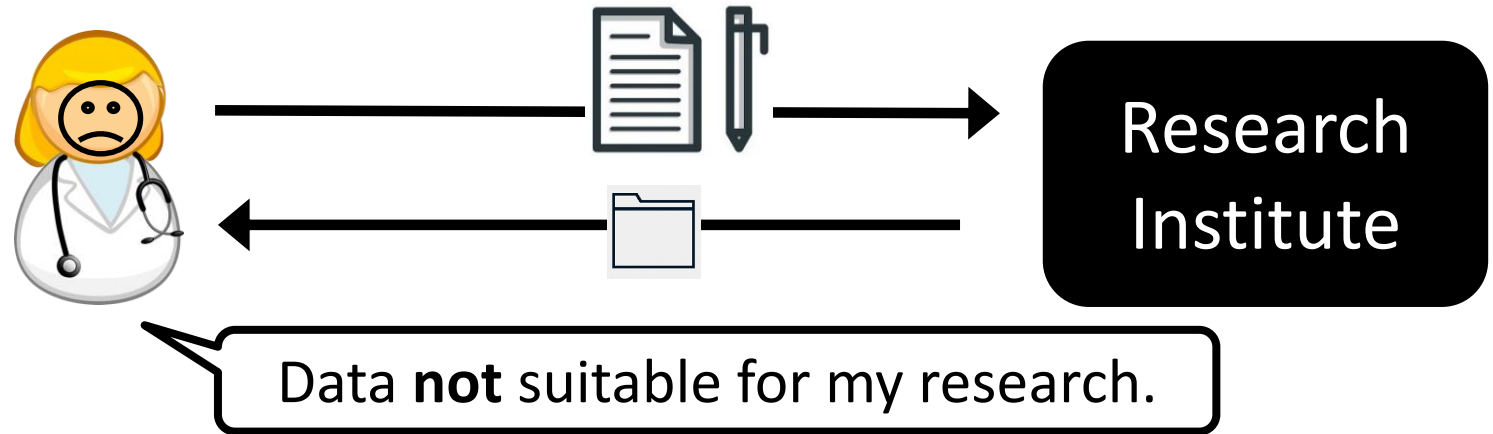


Research
Institute

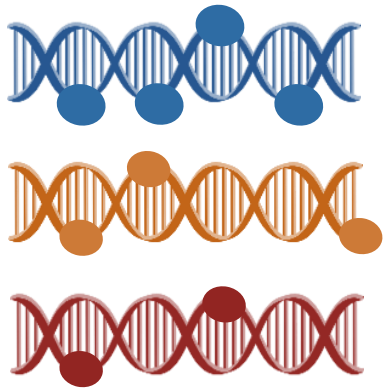
Motivation: The Problem



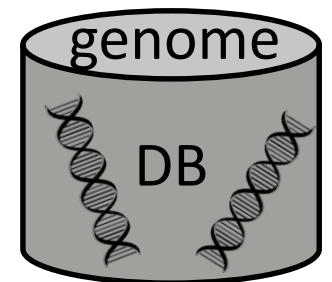
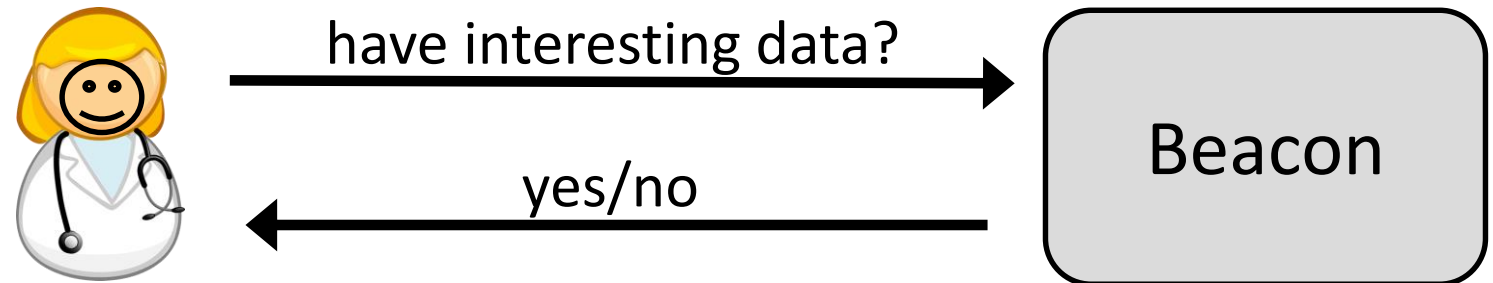
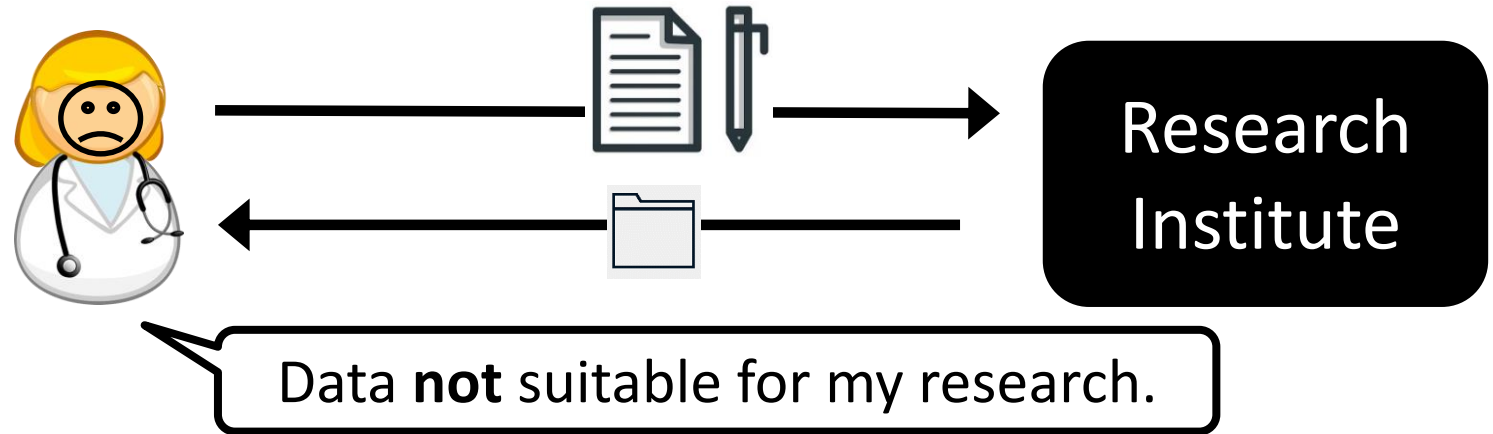
[...0.0, 1.0, 0.6, 0.6, 0.3, 0.3 ...]



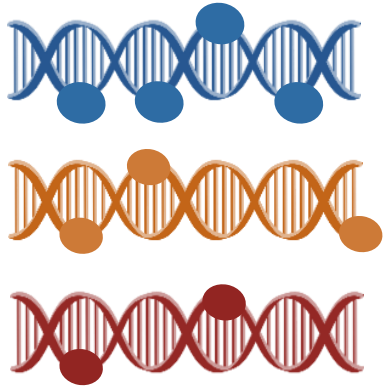
Motivation: First Step Towards Solution



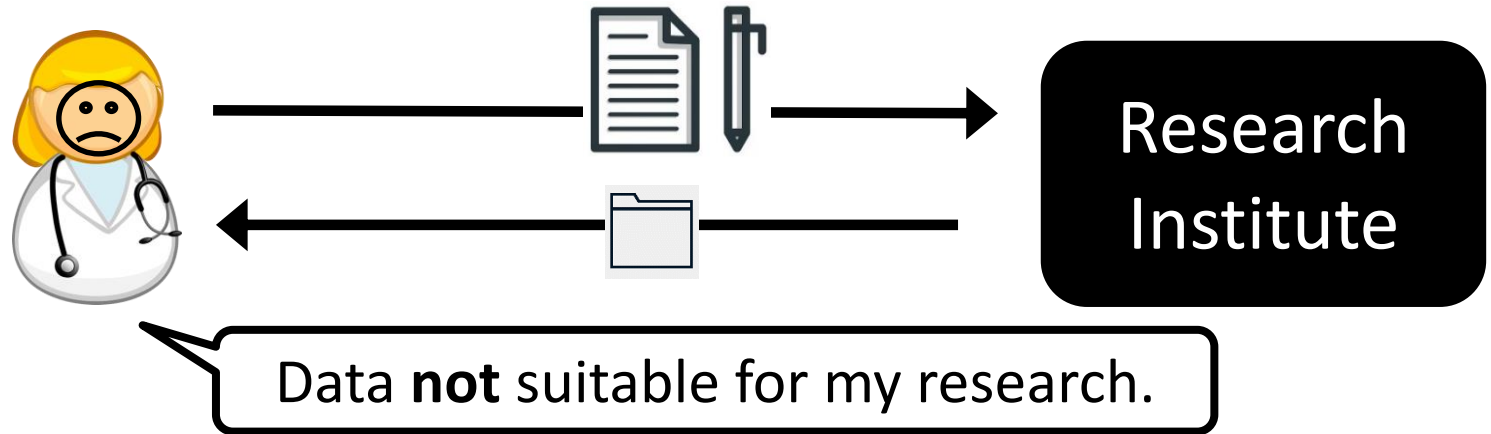
[...0.0, 1.0, 0.6, 0.6, 0.3, 0.3 ...]



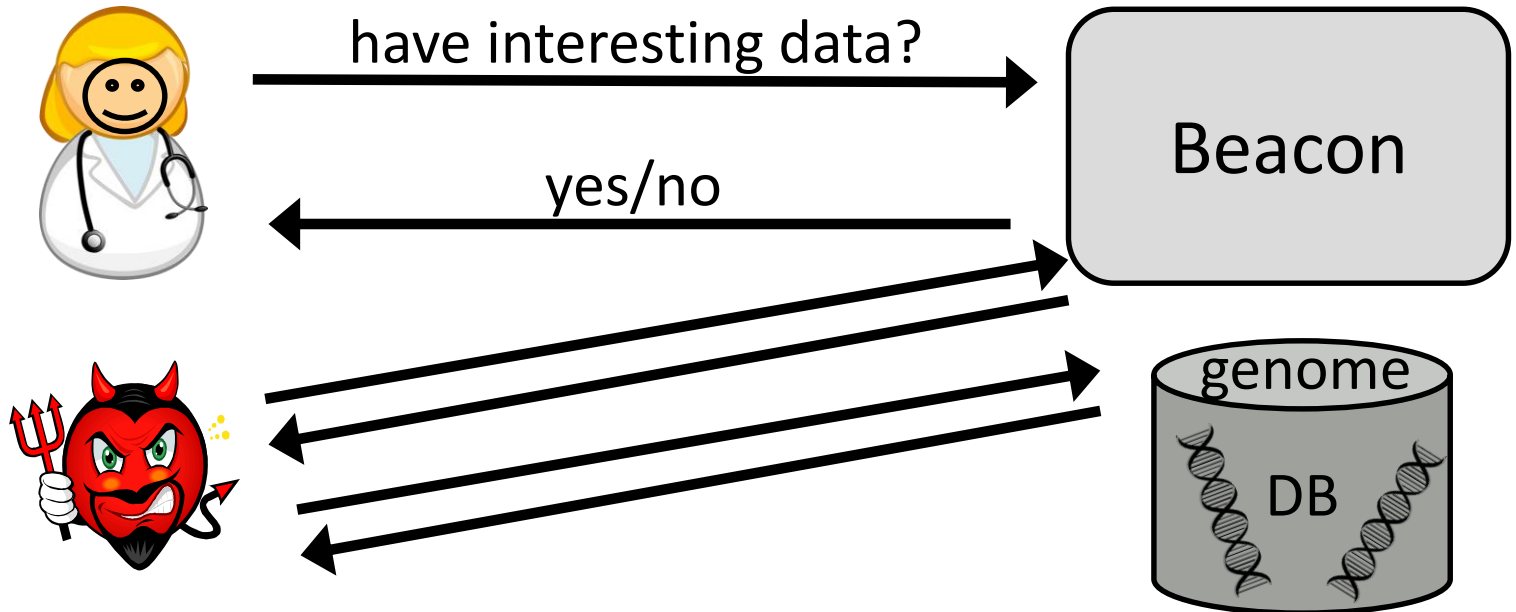
Motivation: First Step Towards Solution



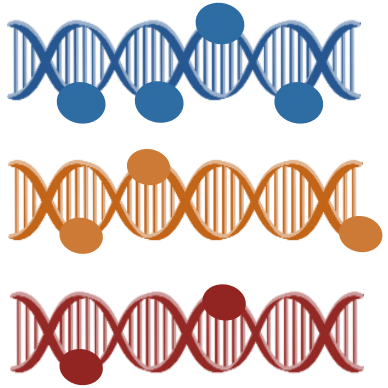
[...0.0, 1.0, 0.6, 0.6, 0.3, 0.3 ...]



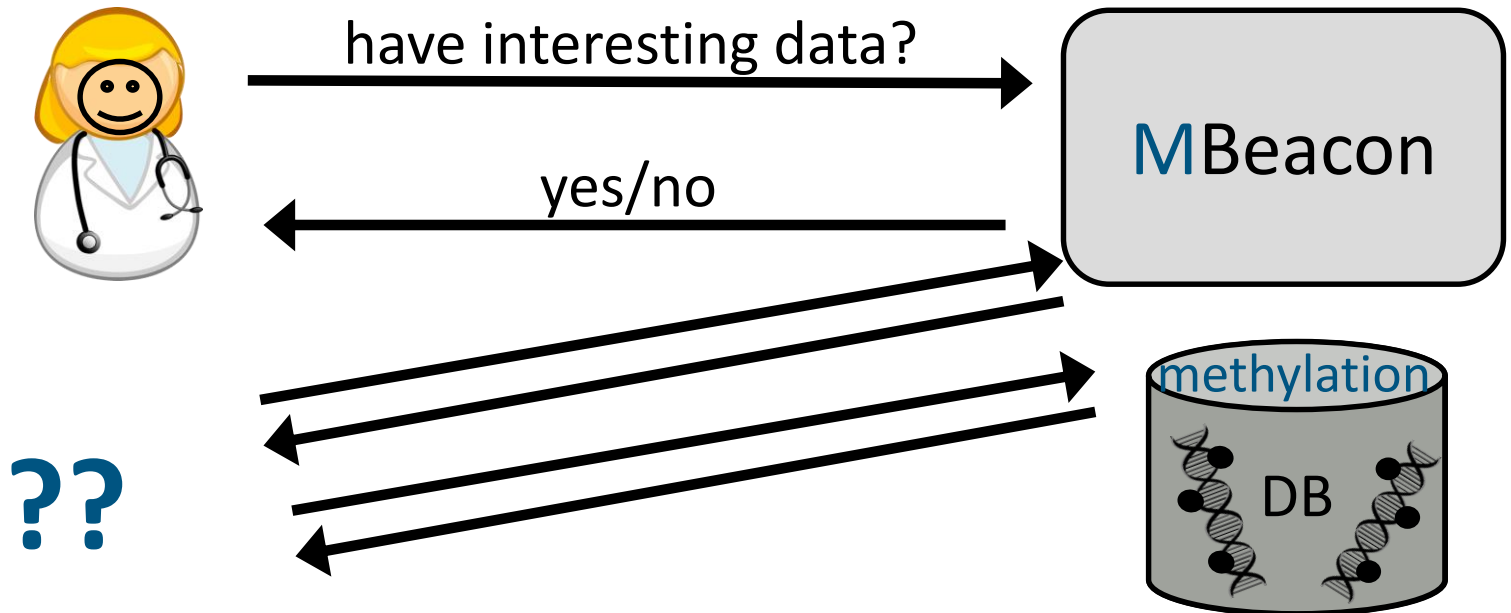
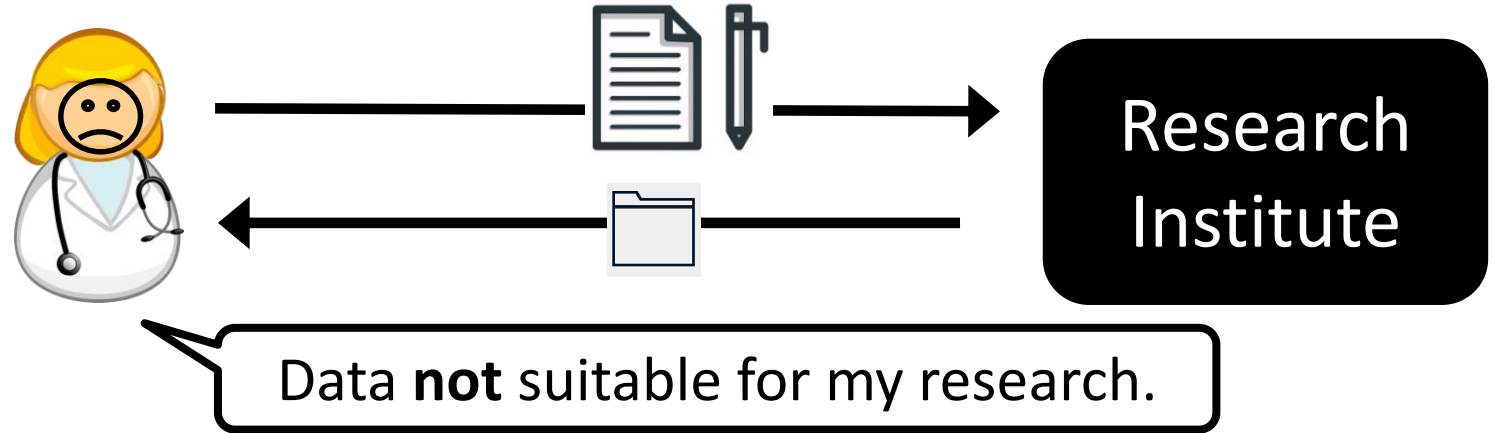
target has cancer



Motivation: Our Solution



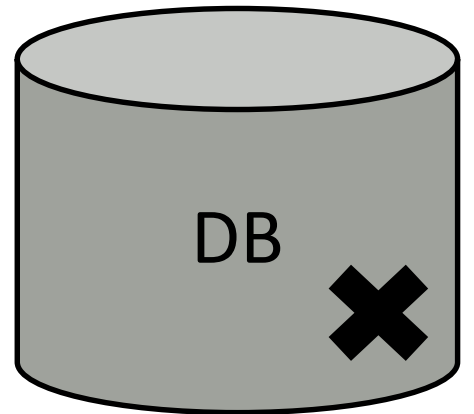
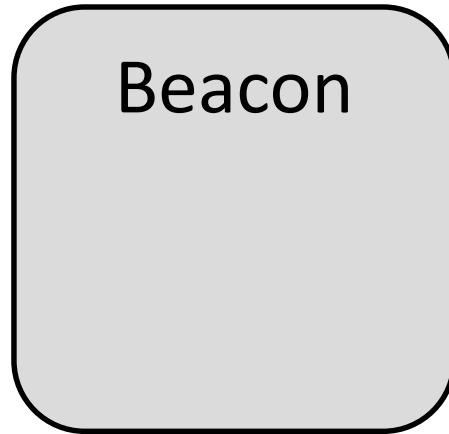
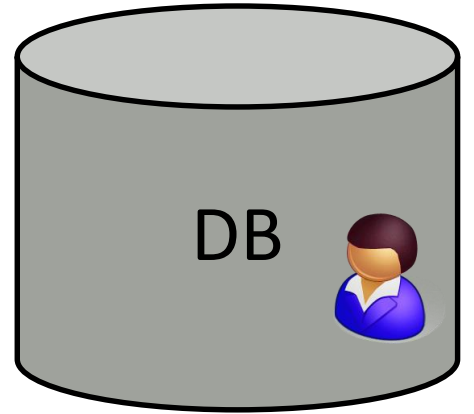
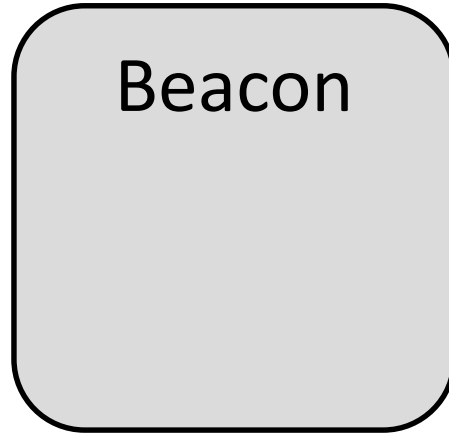
[...0.0, 1.0, 0.6, 0.6, 0.3, 0.3 ...]





Attack Overview

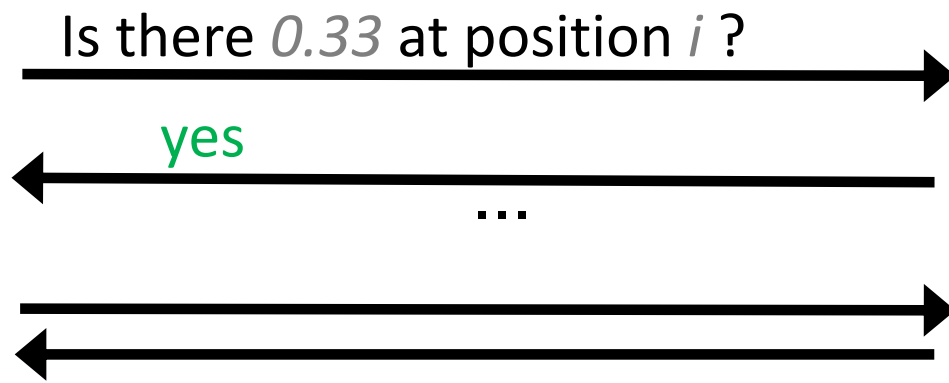


[...
0.33,
1.0,
0.723,
0.615,
0.001,
0.51,
...]



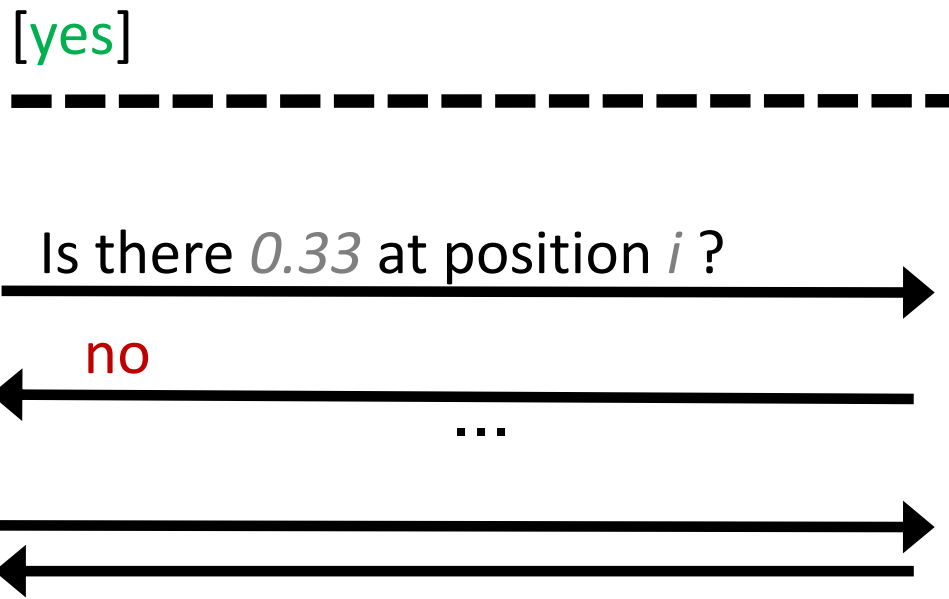
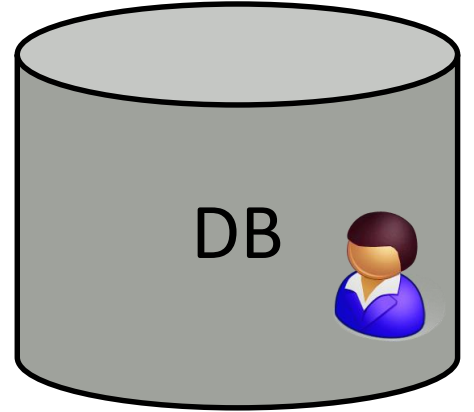
Attack Overview

- 
-
- 
-
- [...
-
- 0.33,
-
- 1.0,
-
- 0.723,
-
- 0.615,
-
- 0.001,
-
- 0.51,
-
- ...]



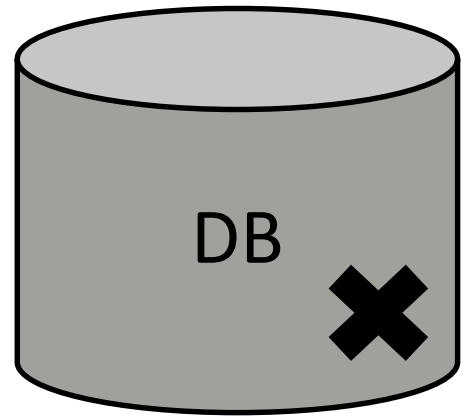
MBeacon

Is there
 $0.3 < x < 0.4$
at position *i* ?





MBeacon

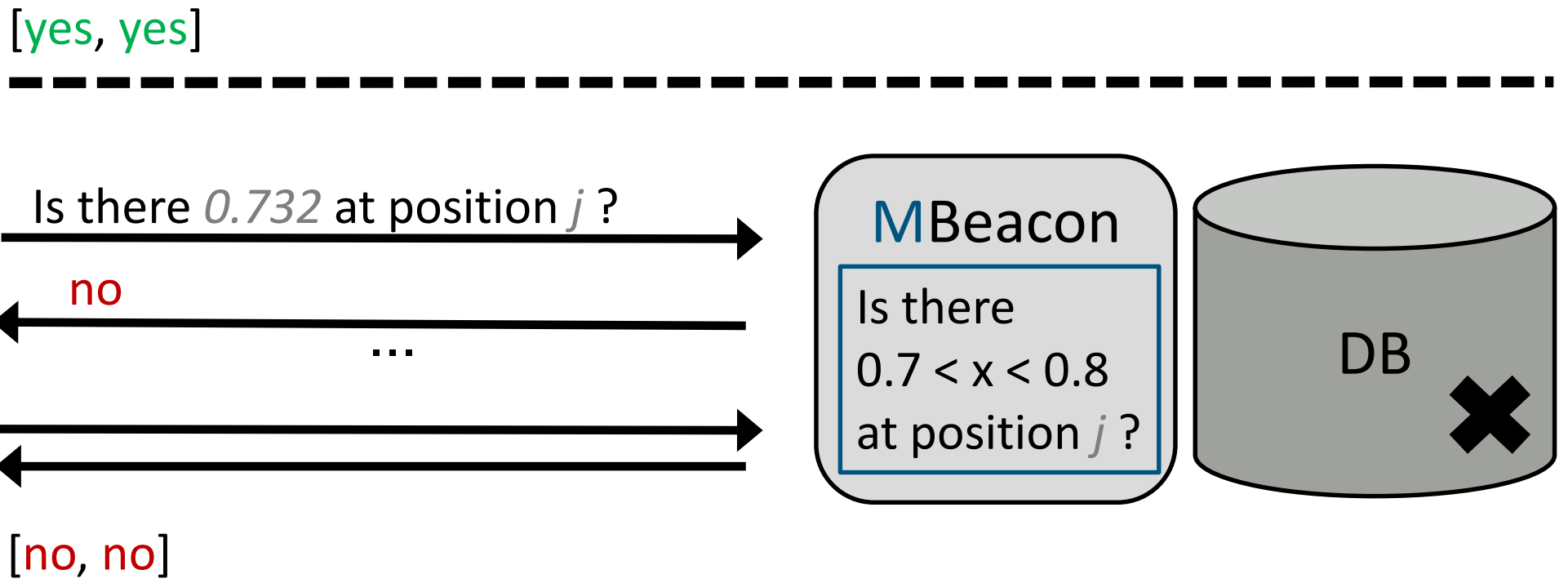
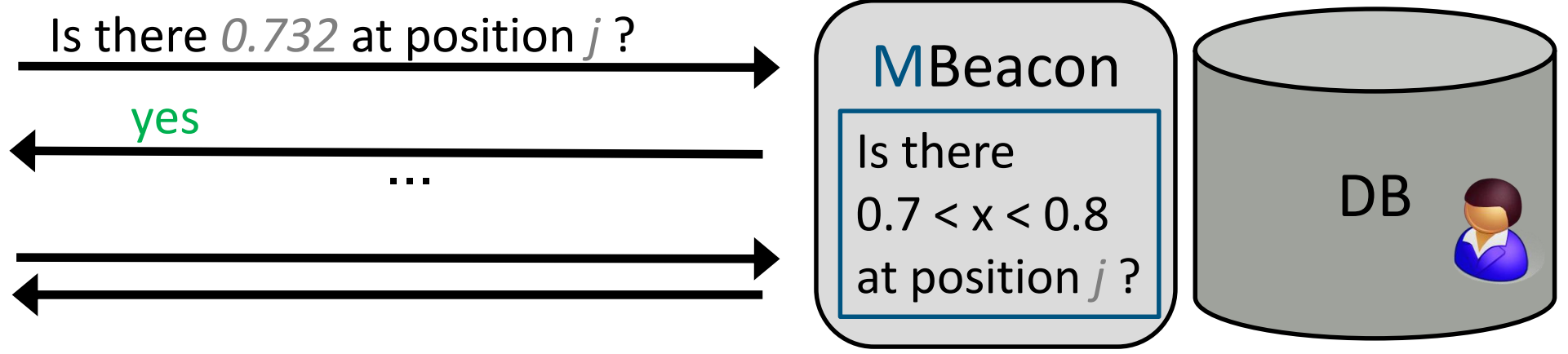
Is there
 $0.3 < x < 0.4$
at position *i* ?



[no]

Attack Overview

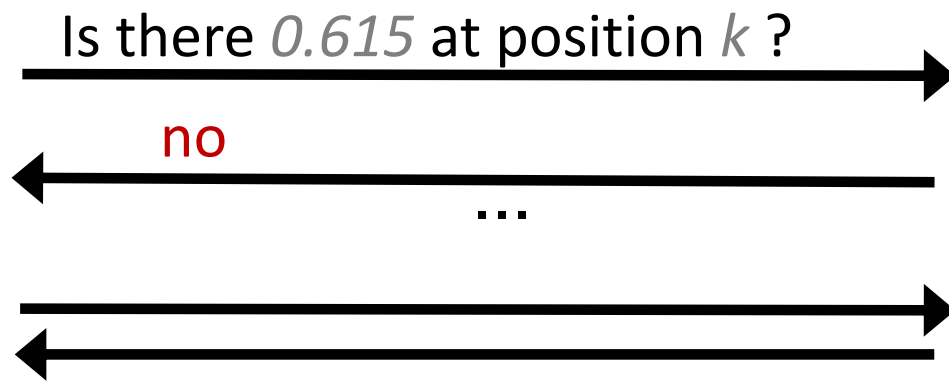
- 
-
- 
-
- [...
-
- 0.33,
-
- 1.0,
-
- 0.723,
-
- 0.615,
-
- 0.001,
-
- 0.51,
-
- ...]



[no, no]

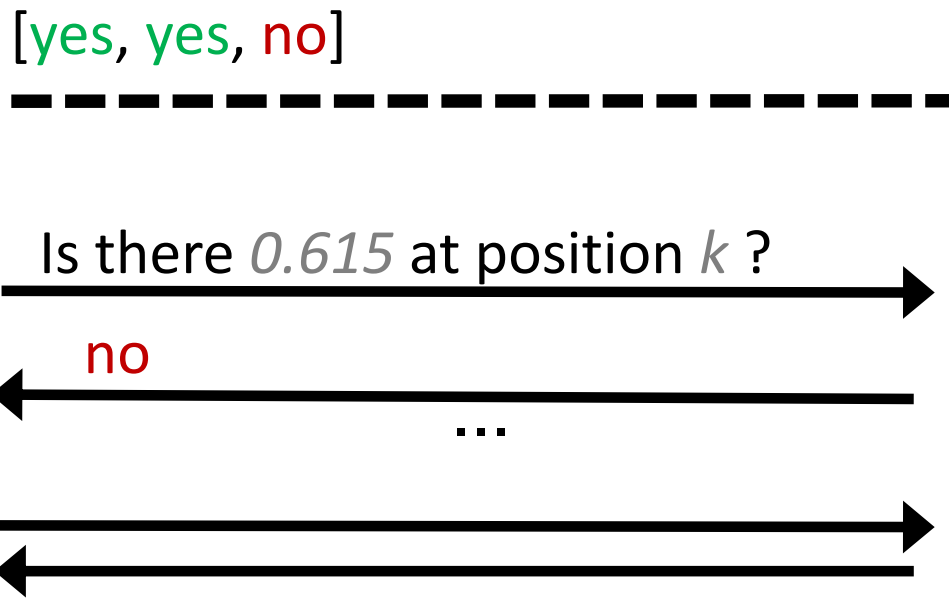
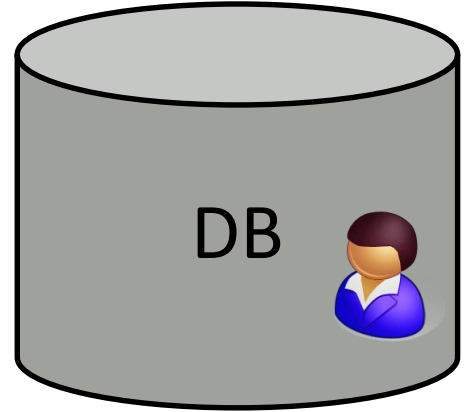
Attack Overview

- [...
0.33,
1.0,
0.723,
0.615,
0.001,
0.51,
...]



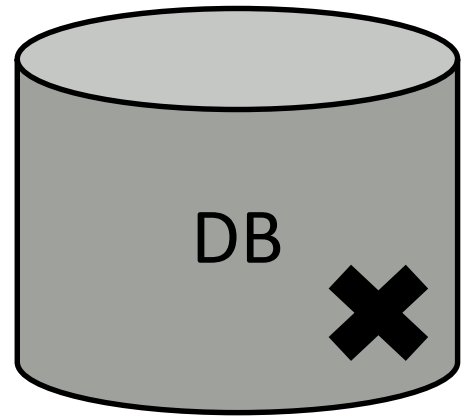
MBeacon

Is there
 $0.6 < x < 0.7$
at position *k*?



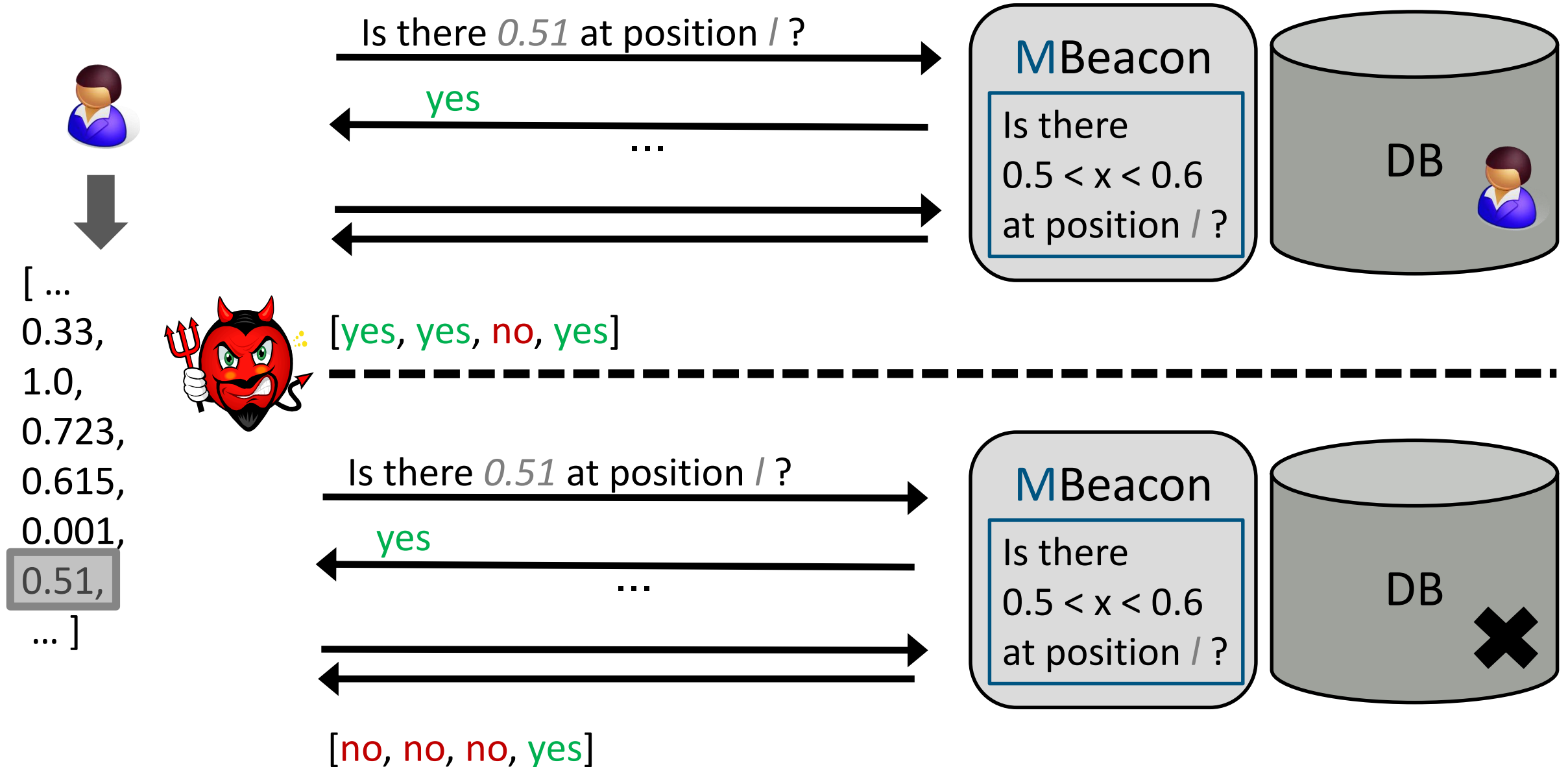
MBeacon

Is there
 $0.6 < x < 0.7$
at position *k*?



[no, no, no]

Attack Overview



Attack Overview



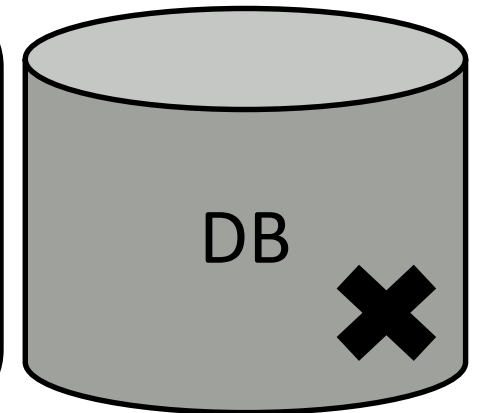
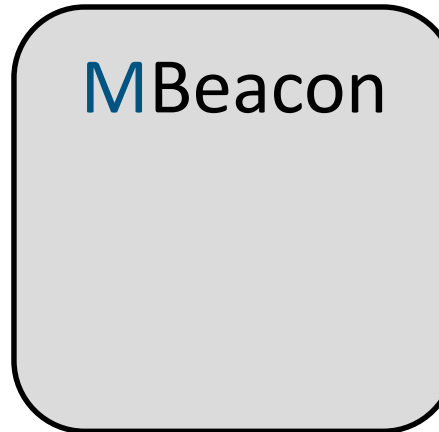
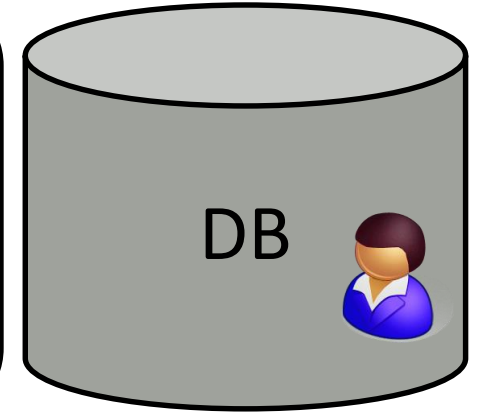
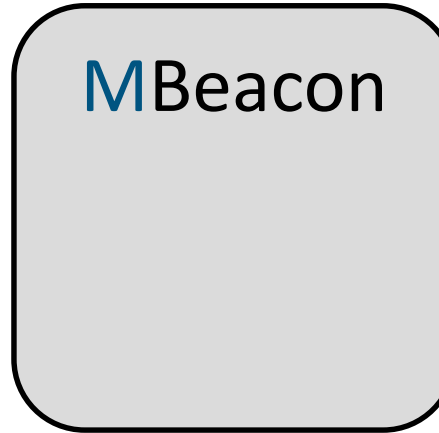
[...
0.33,
1.0,
0.723,
0.615,
0.001,
0.51,
...]



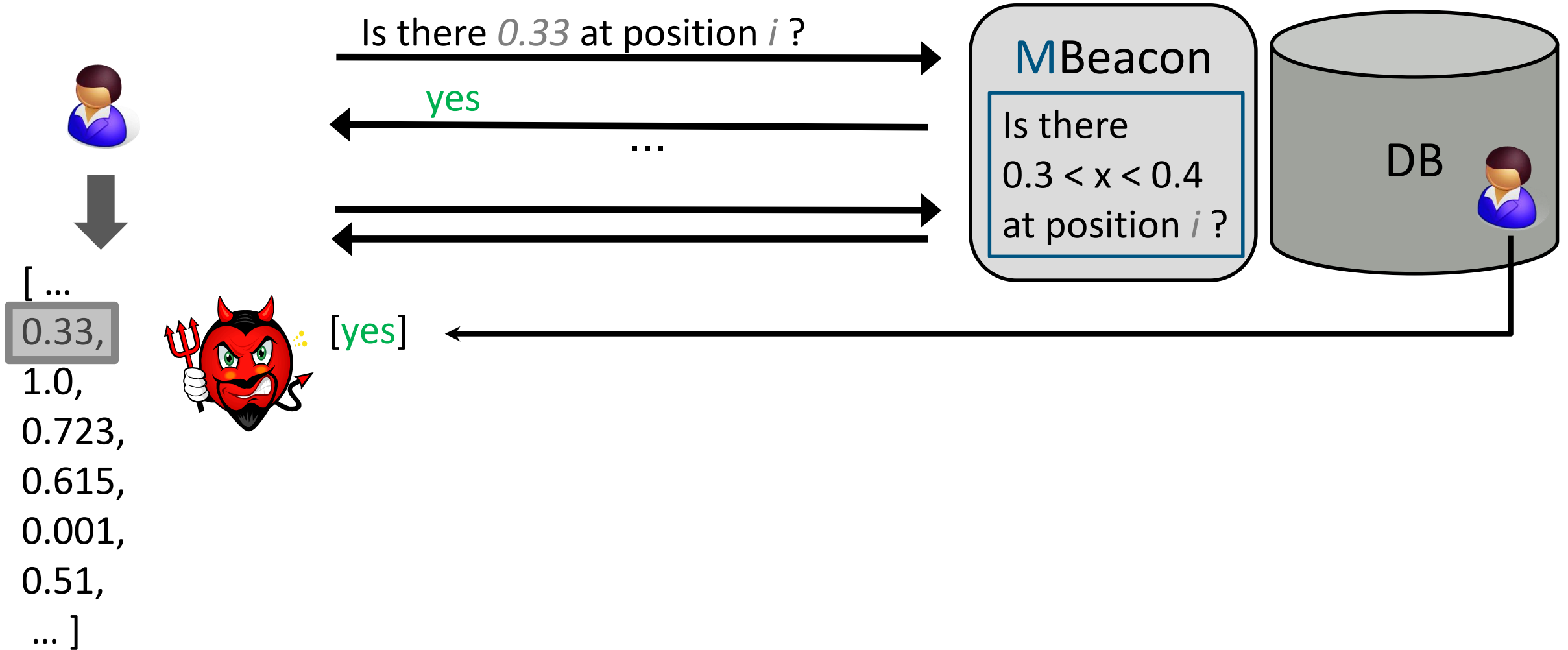
[yes, yes, no, yes, yes, no, yes, yes]



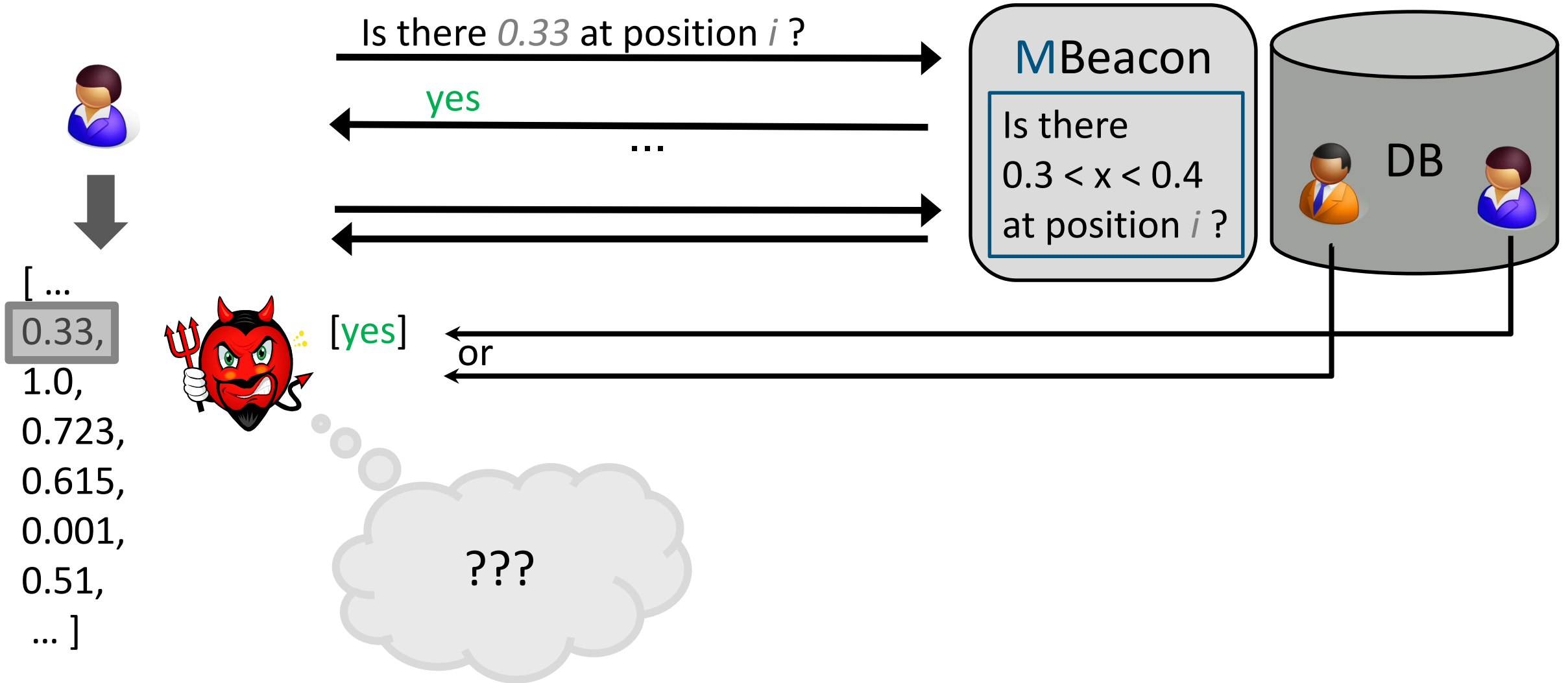
[no, no, no, yes, no, yes, no, no]



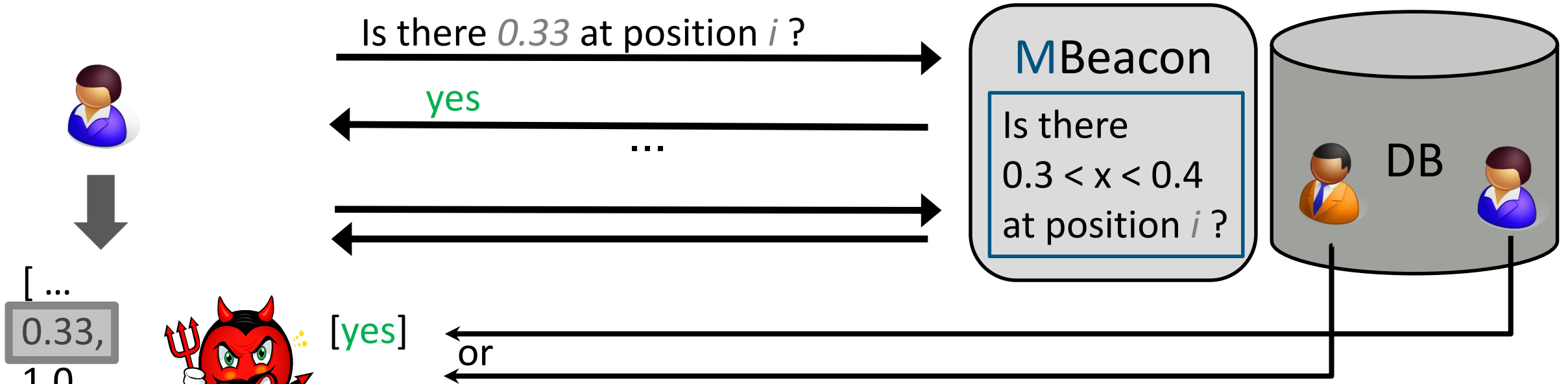
Attack: Estimation of **yes** Answers



Attack: Estimation of **yes** Answers



Attack: Estimation of **yes** Answers

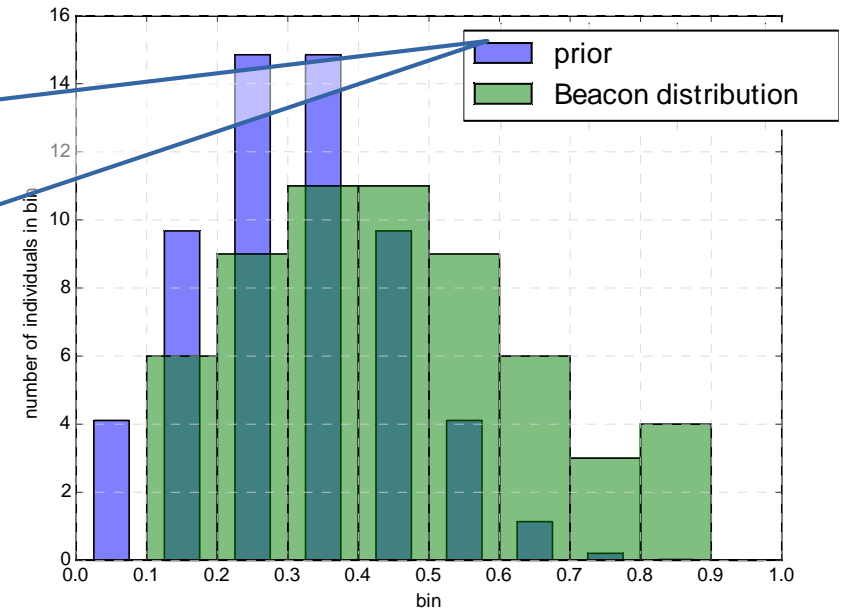


- [...
- 0.33,
- 1.0,
- 0.723,
- 0.615,
- 0.001,
- 0.51,
- ...]

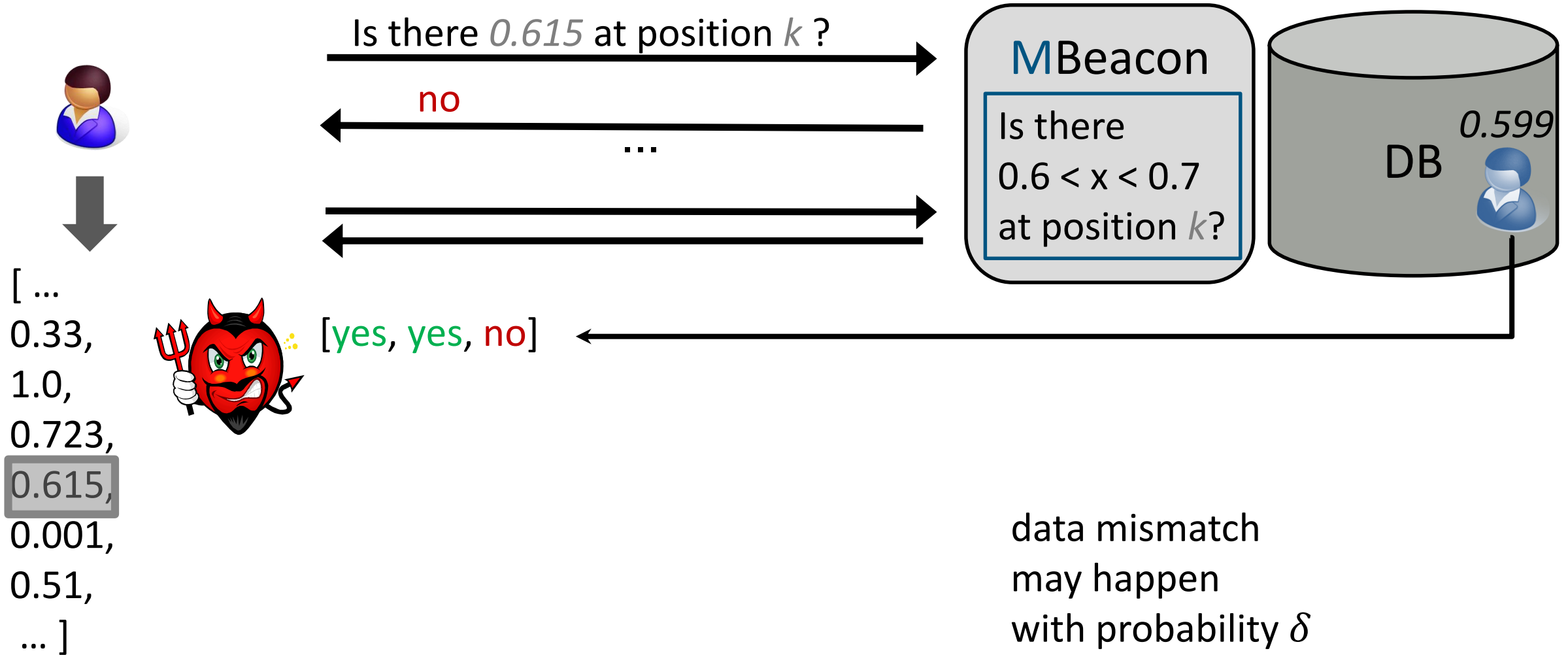


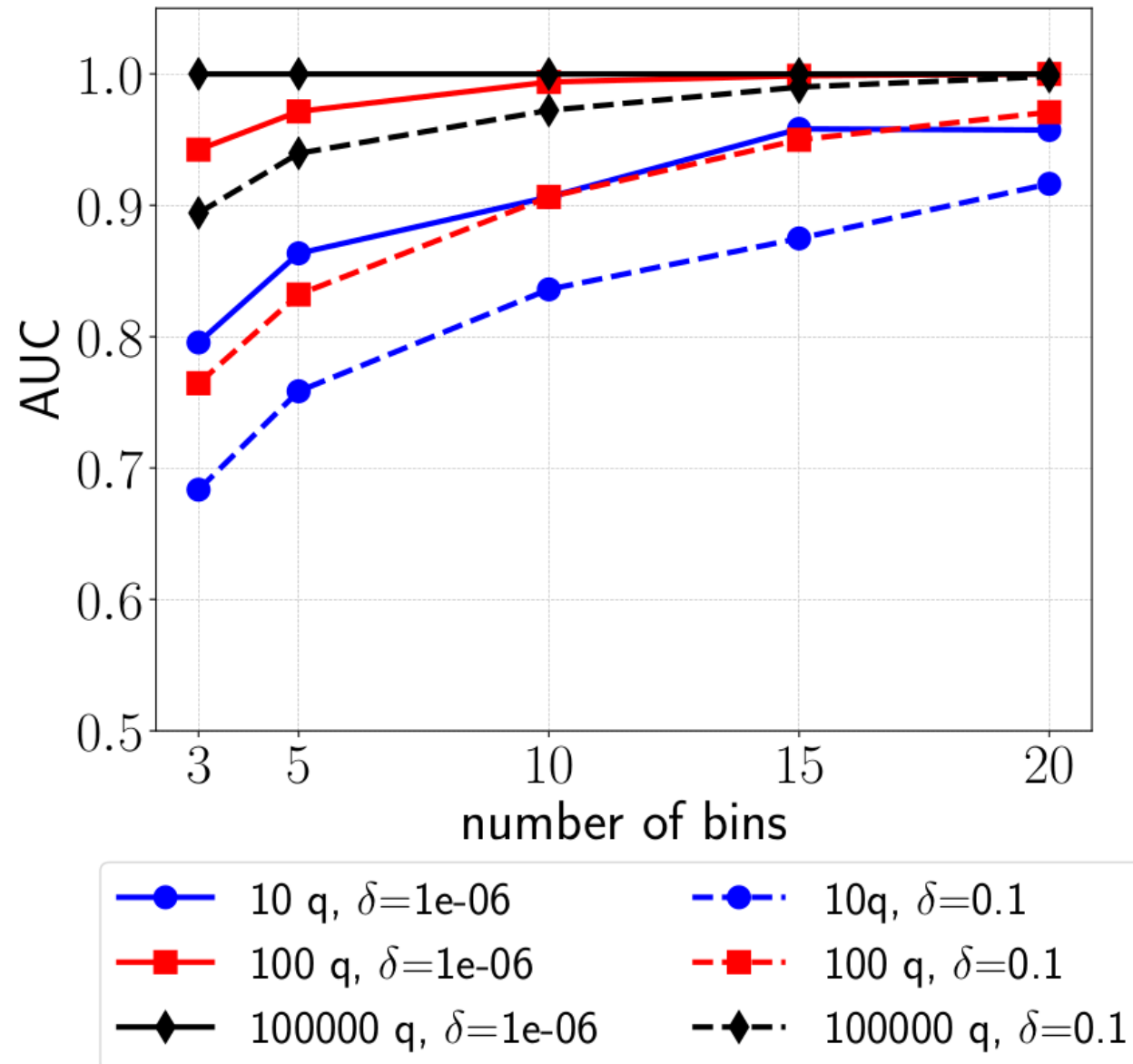
???

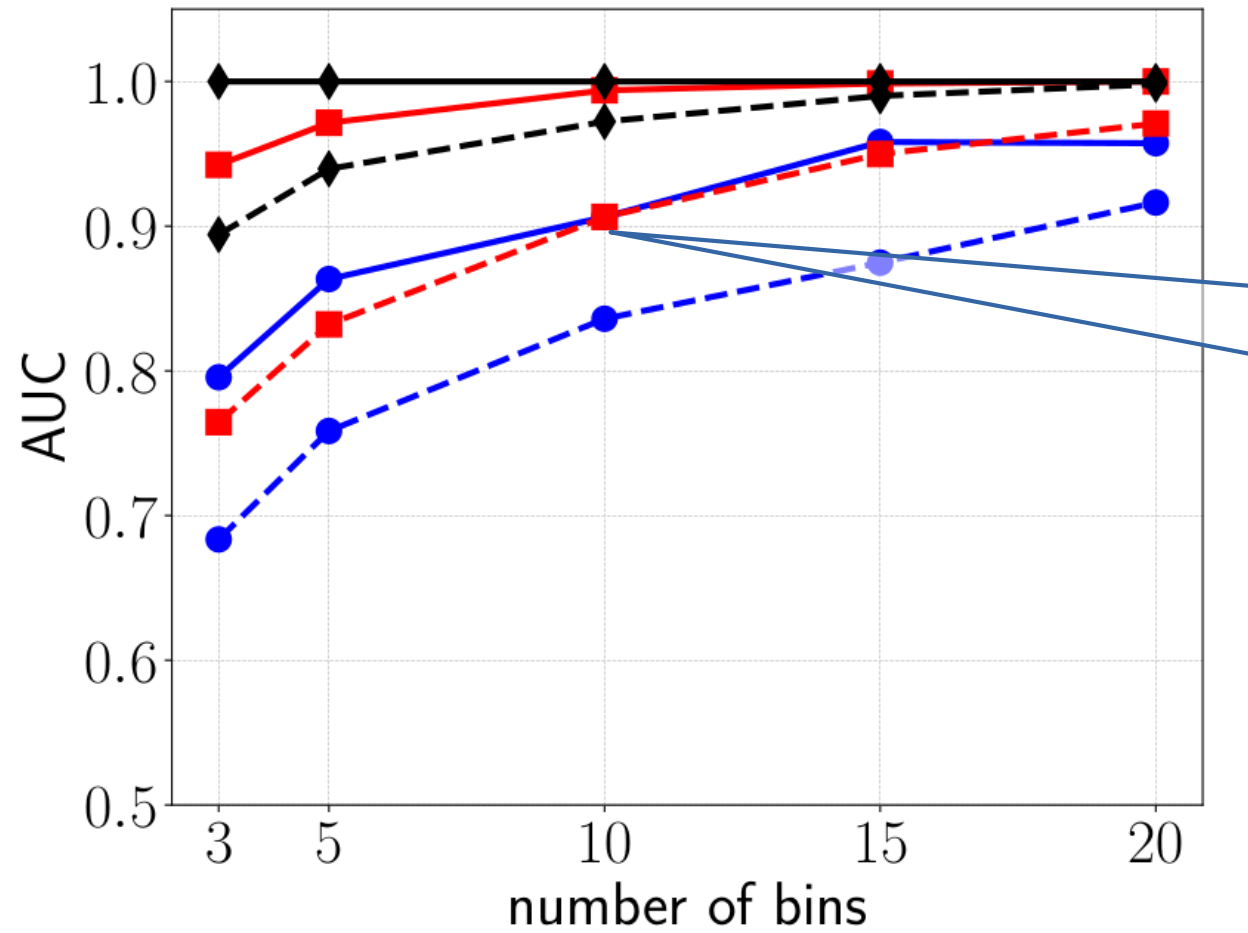
normal distribution with known mean and standard deviation



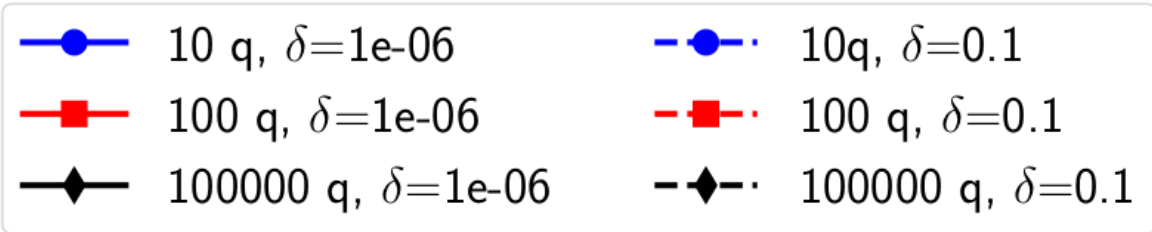
Attack: Estimation of no Answers

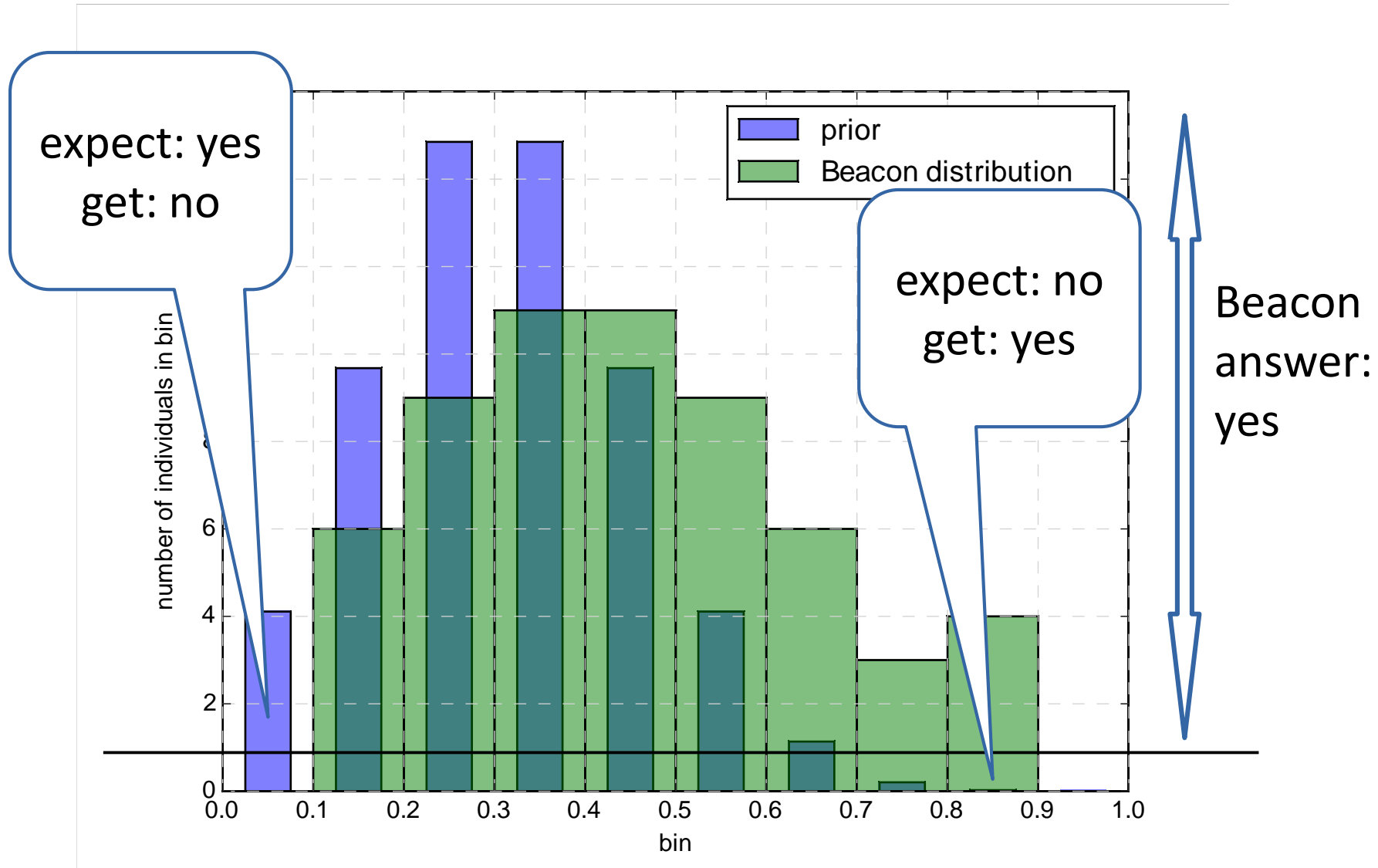




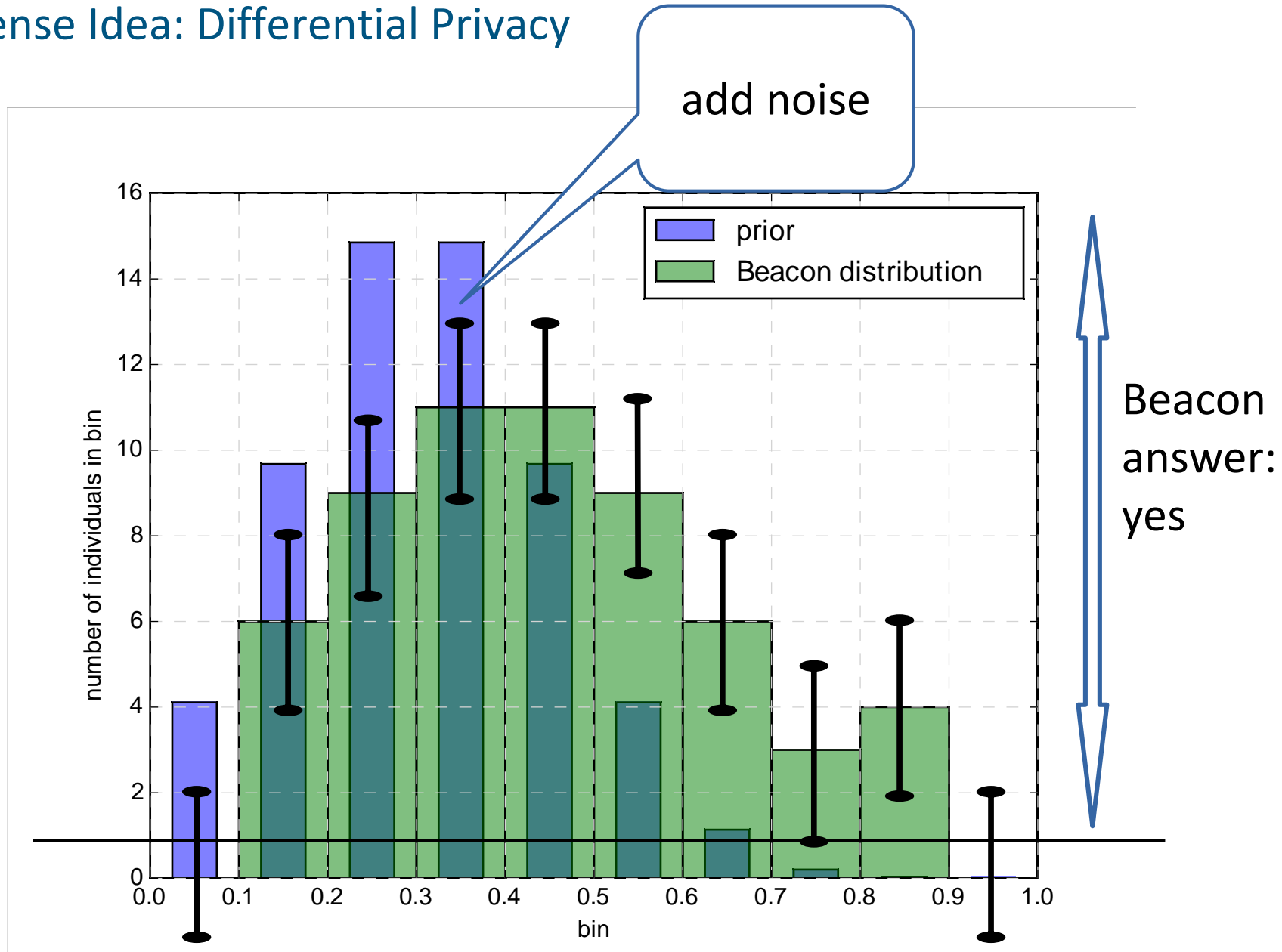


10 queries,
10 bins:
> 0.9 AUC

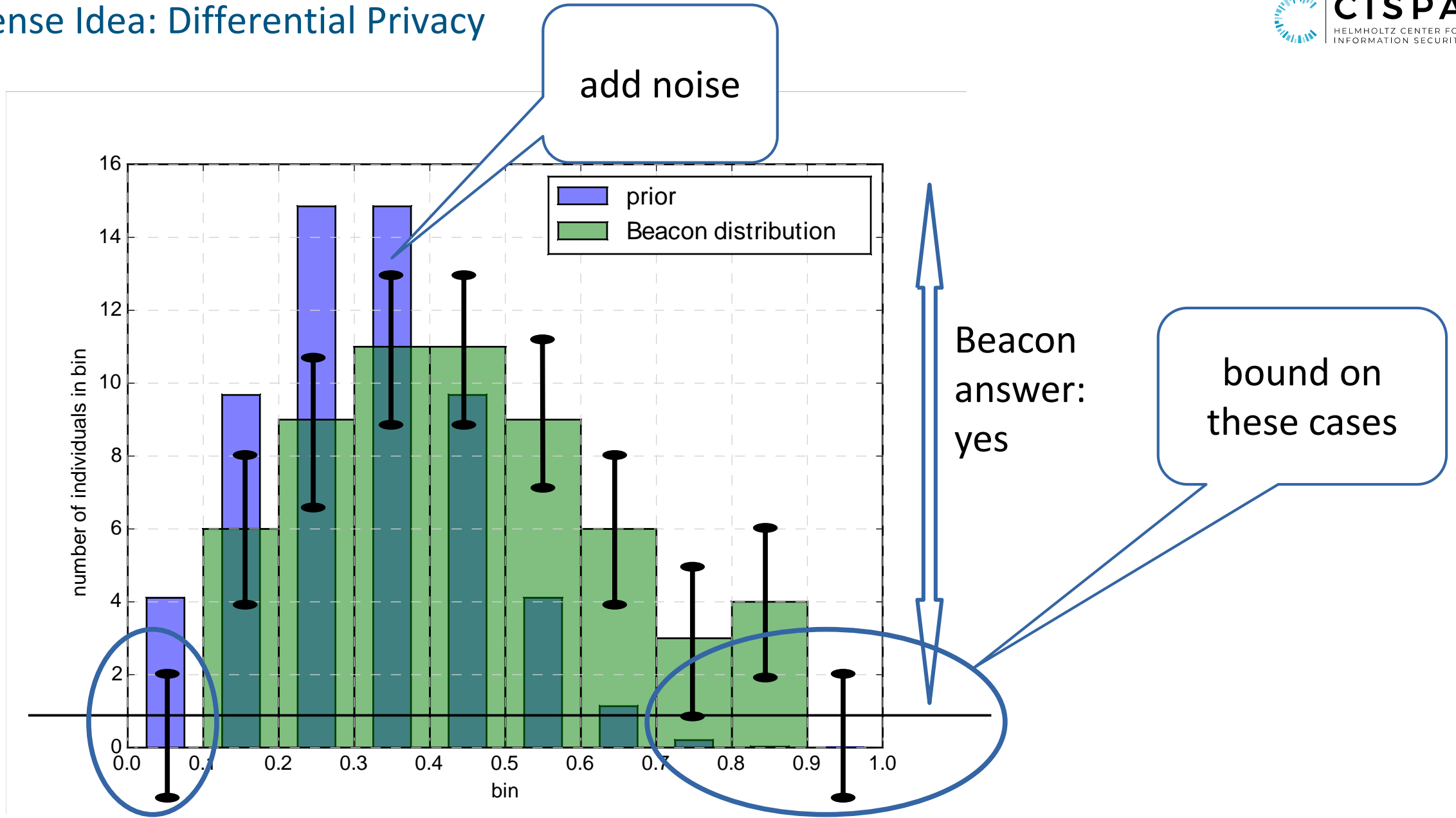




Defense Idea: Differential Privacy



Defense Idea: Differential Privacy

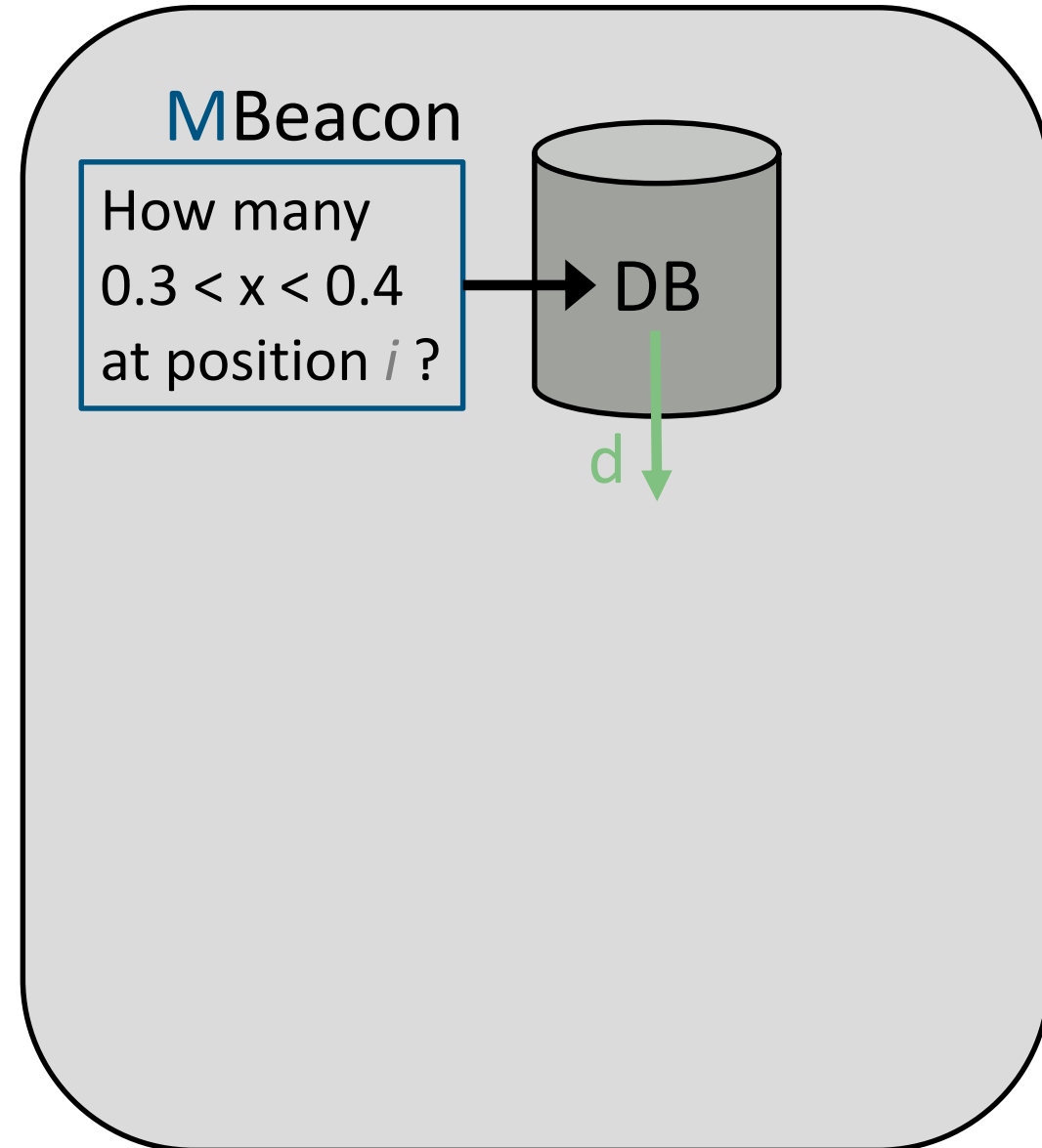


Our Defense in Detail: SVT²

Is there 0.33 at position i ?

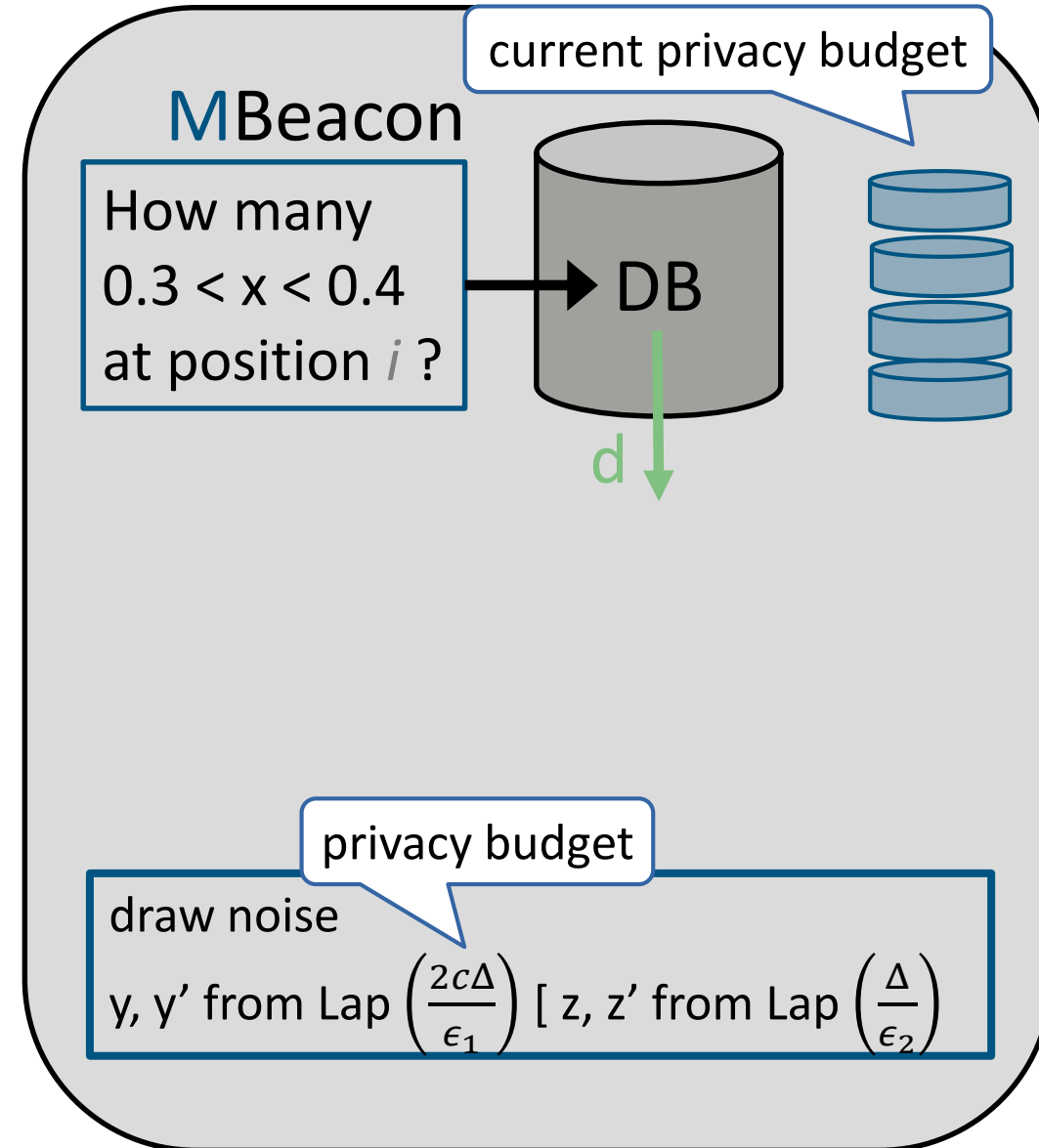
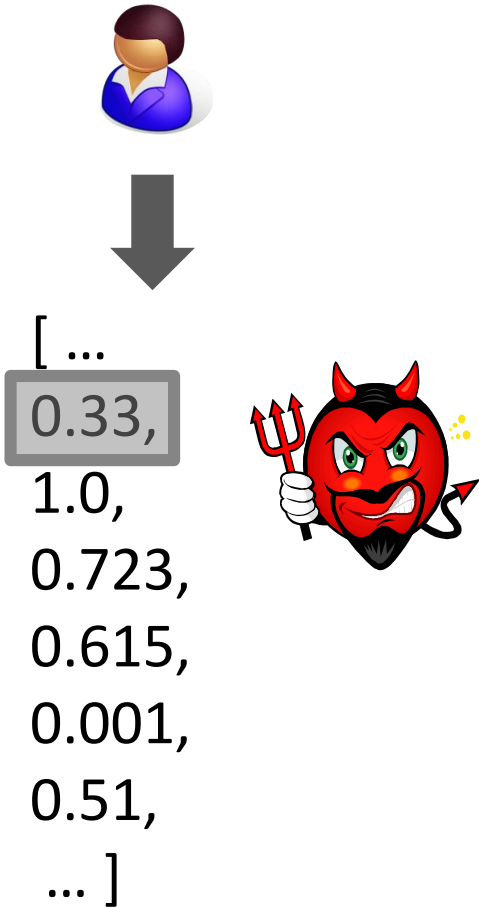


[...
0.33,
1.0,
0.723,
0.615,
0.001,
0.51,
...]



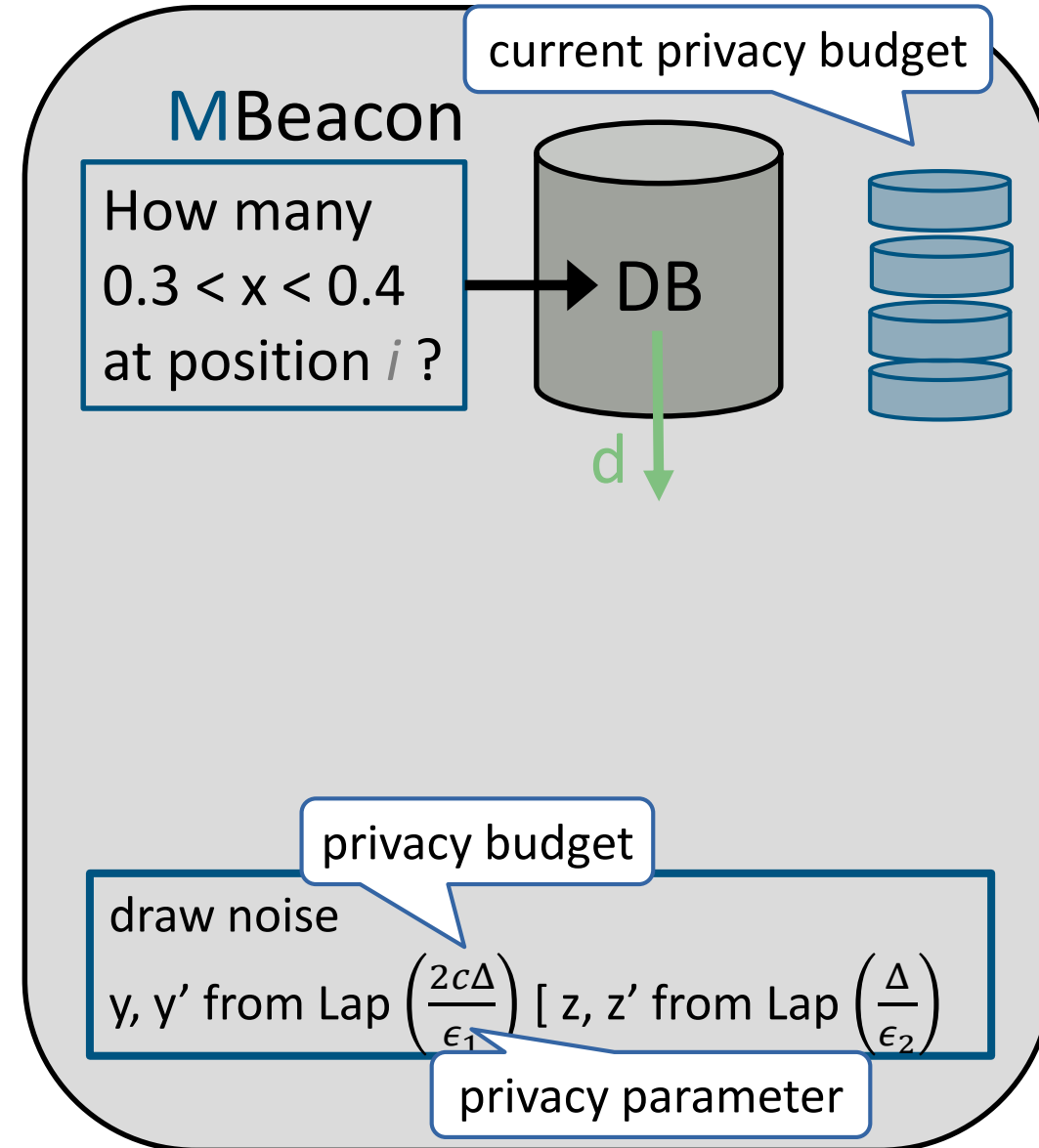
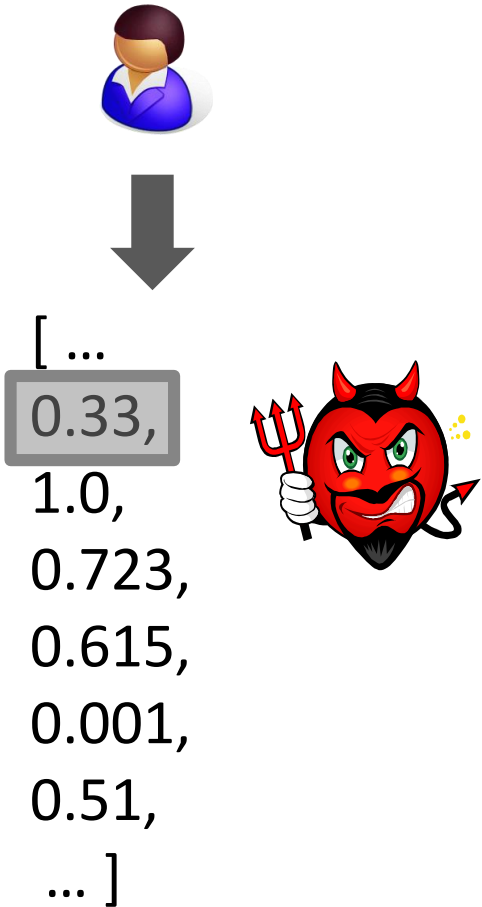
Our Defense in Detail: SVT²: Laplace Noise

Is there *0.33* at position *i*?



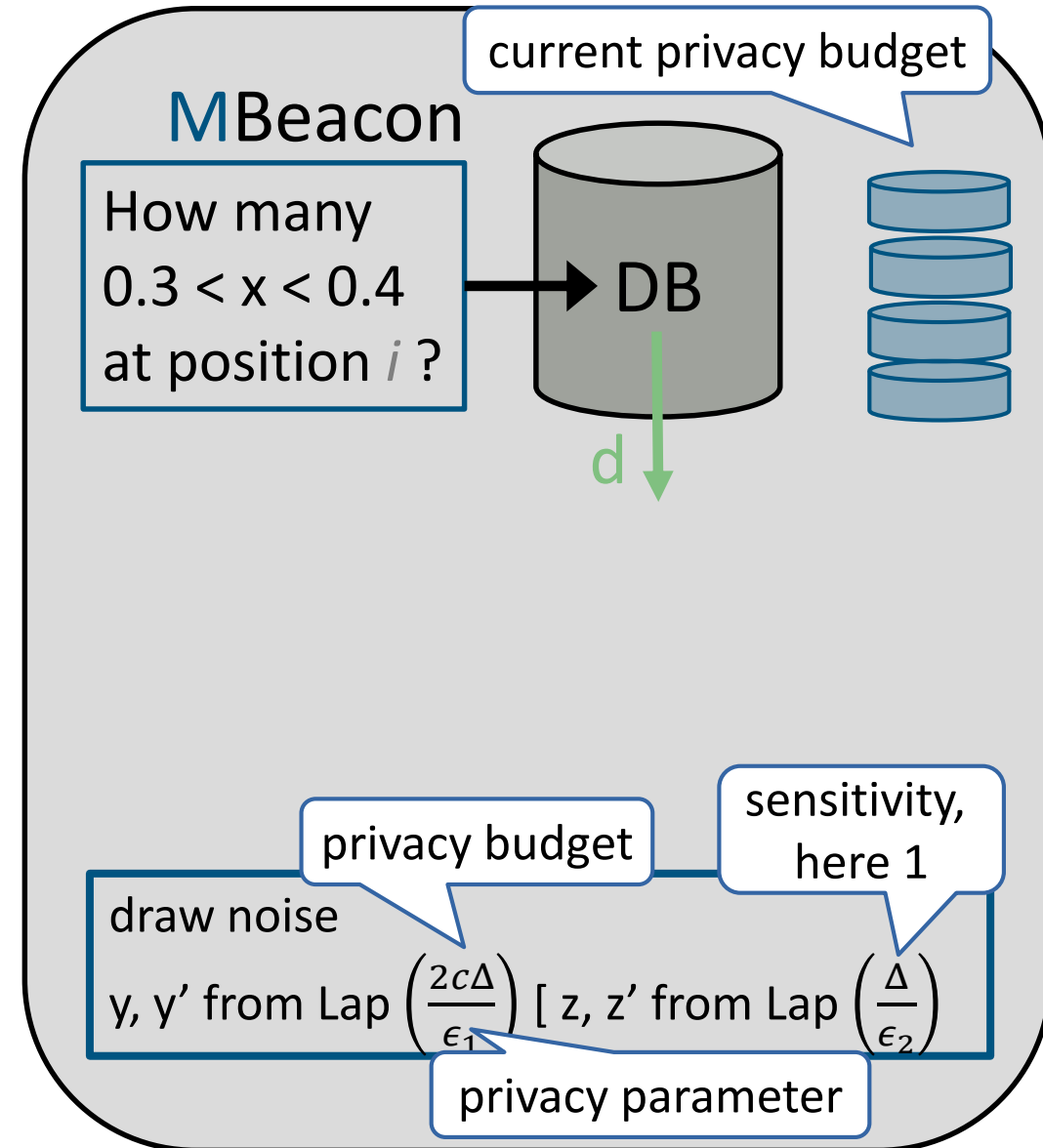
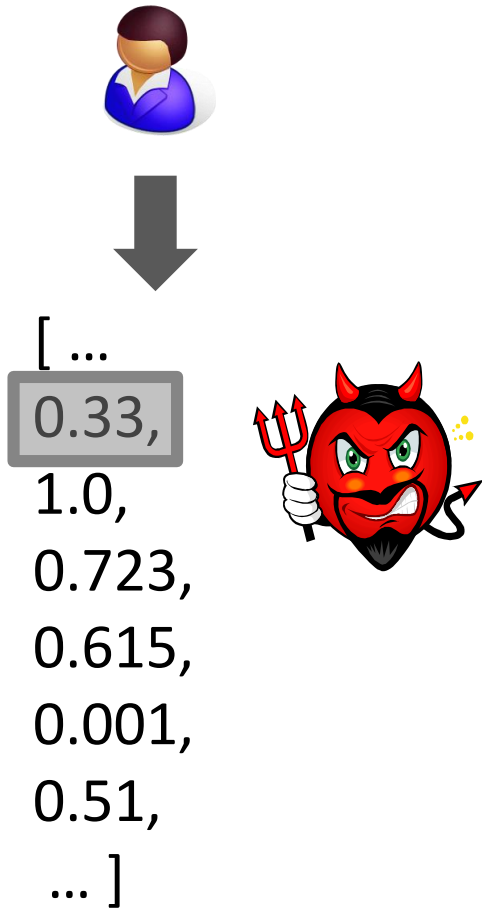
Our Defense in Detail: SVT²: Laplace Noise

Is there 0.33 at position i ?



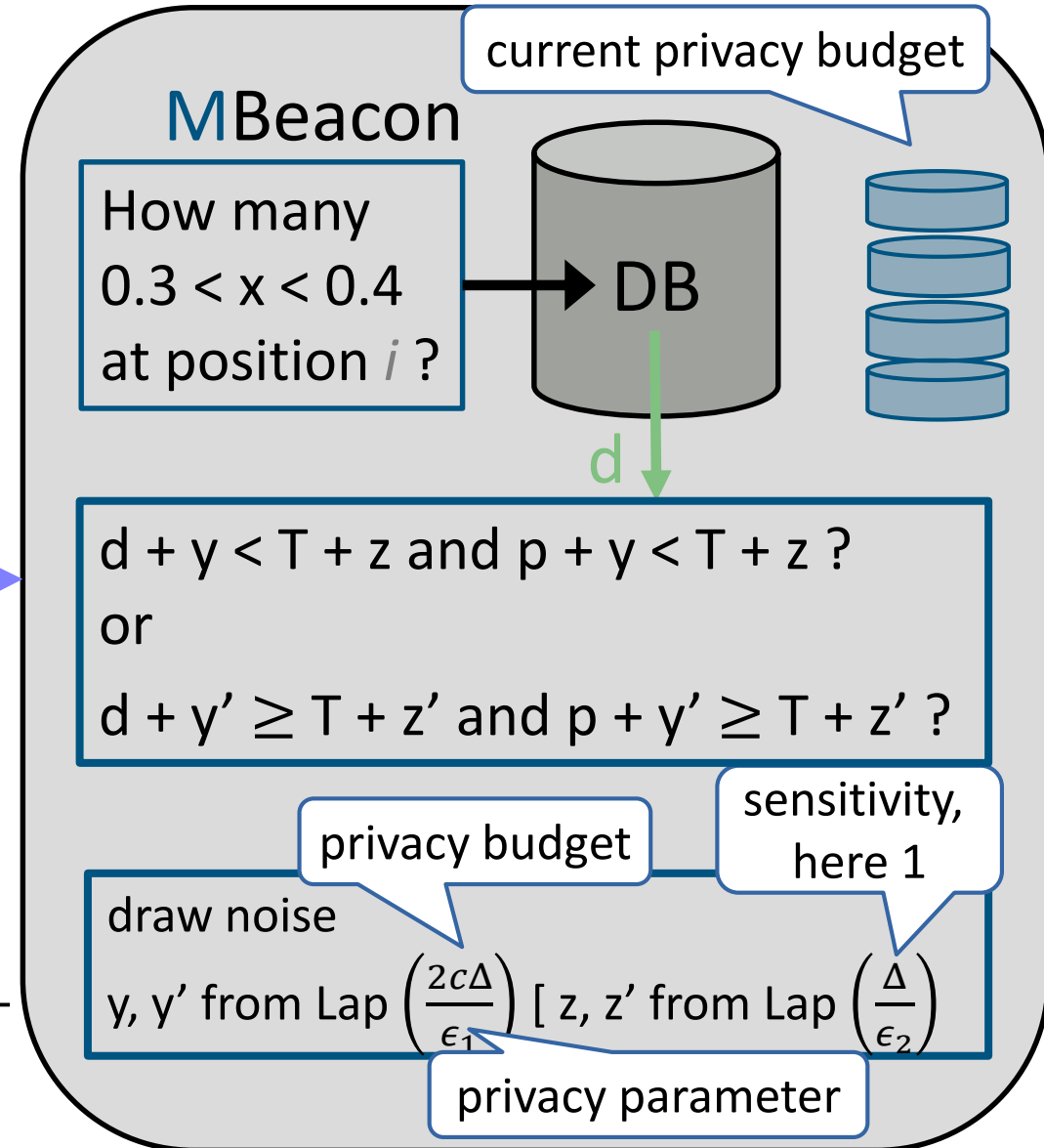
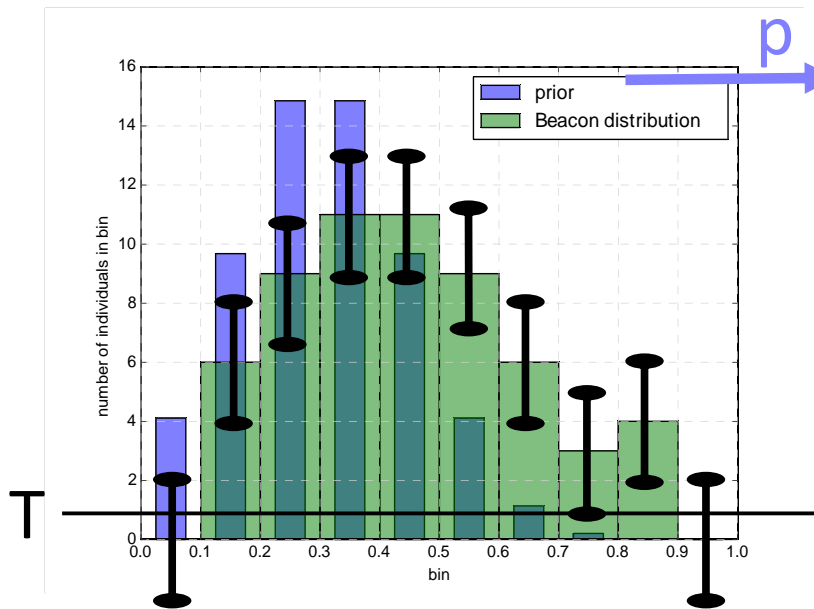
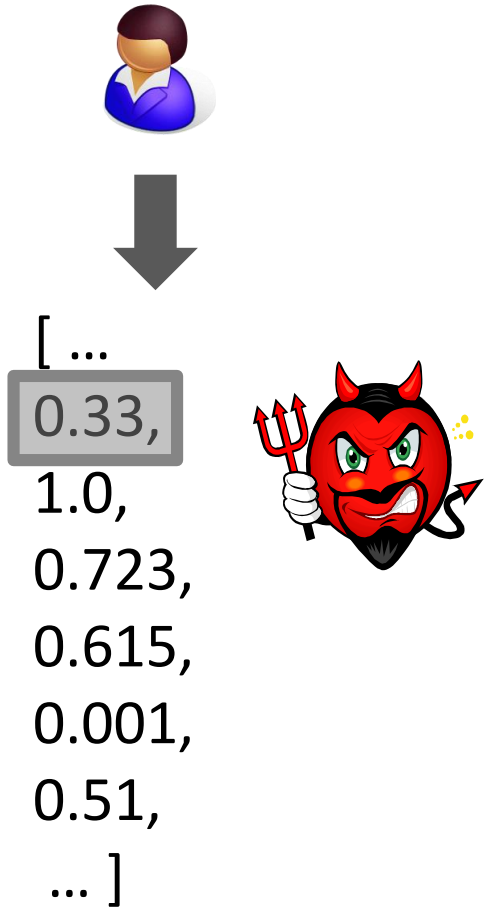
Our Defense in Detail: SVT²: Laplace Noise

Is there 0.33 at position i ?

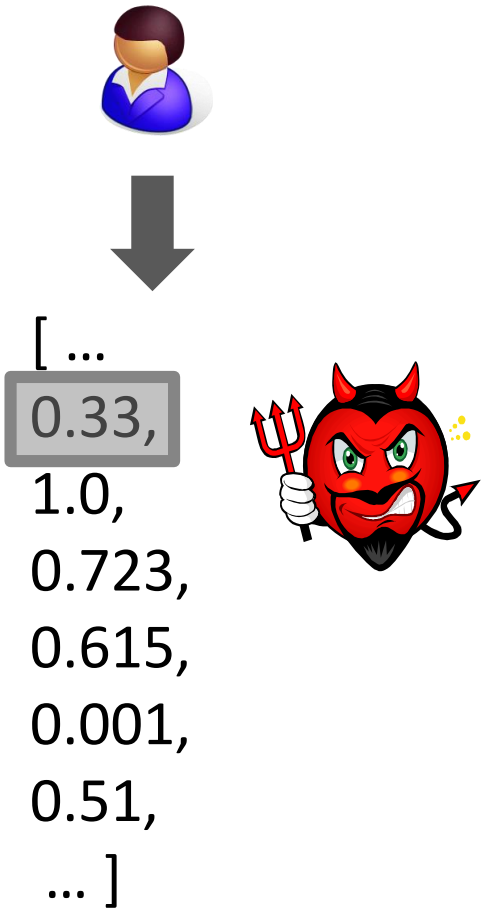


Our Defense in Detail: SVT²: Noisy Comparison

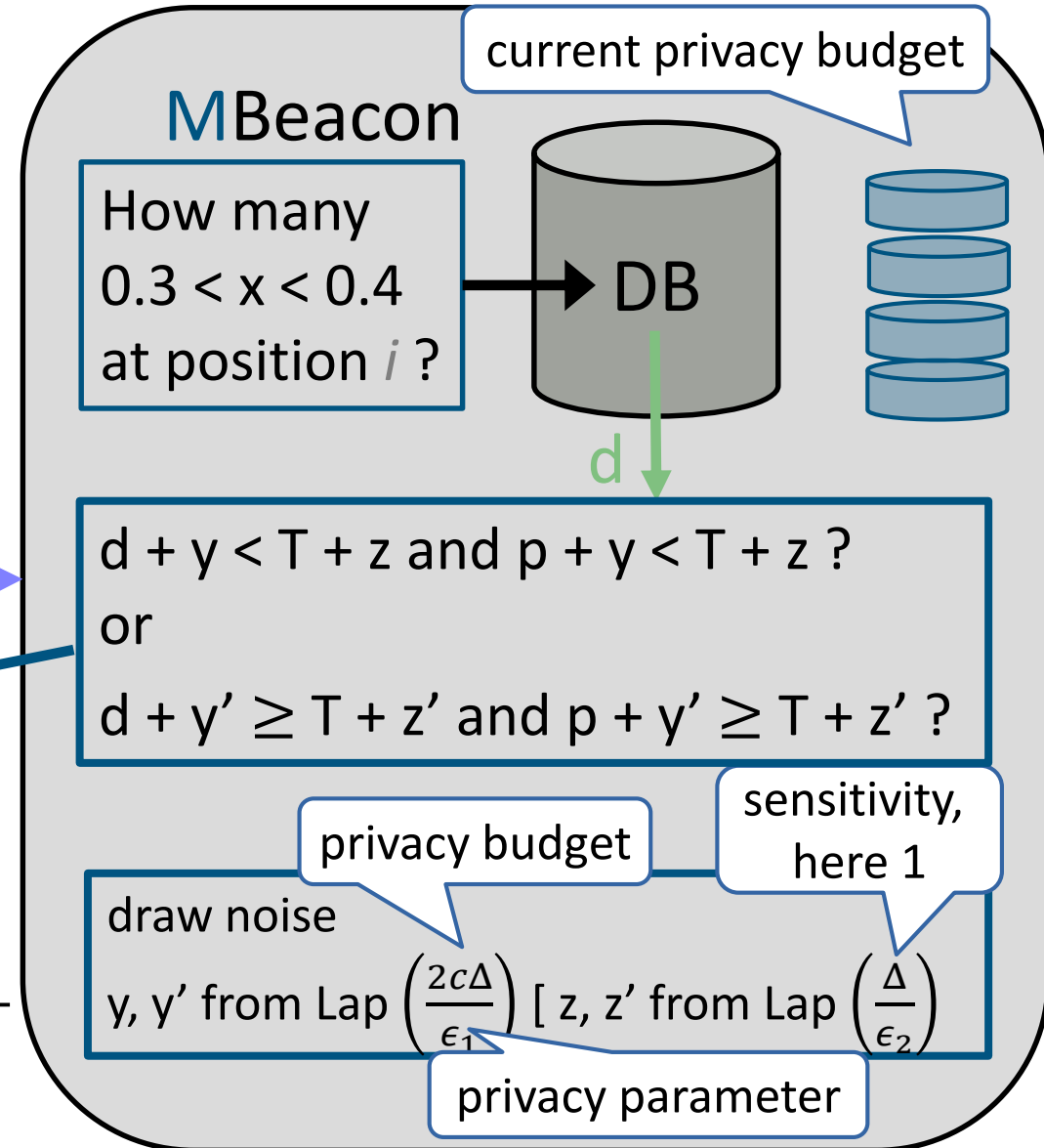
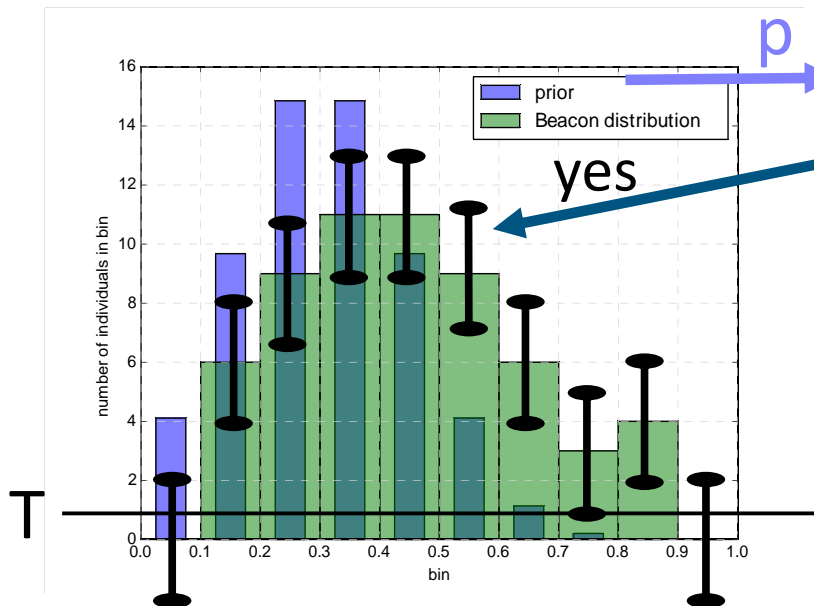
Is there 0.33 at position i ?



Our Defense in Detail: SVT²: Noisy Comparison

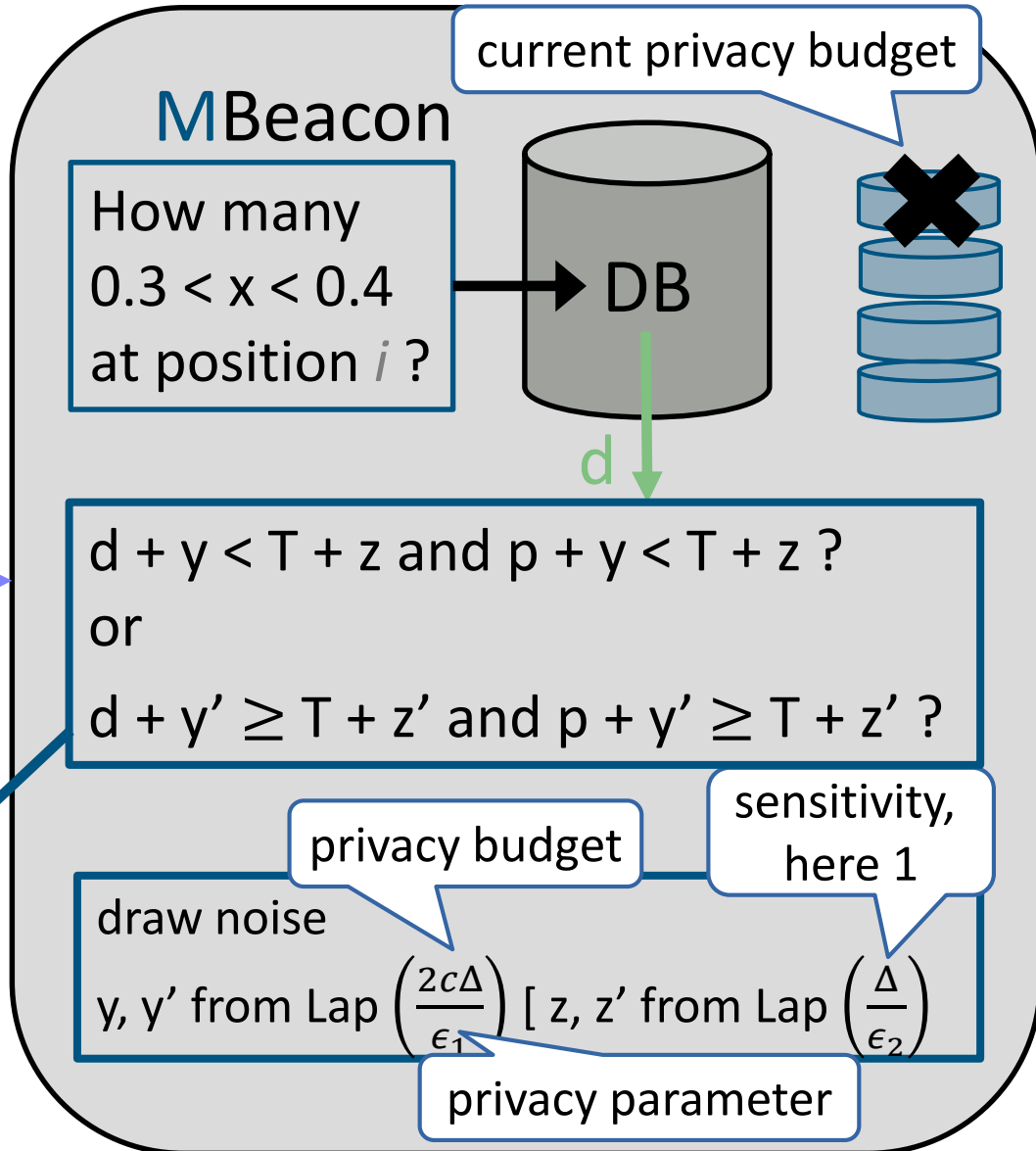
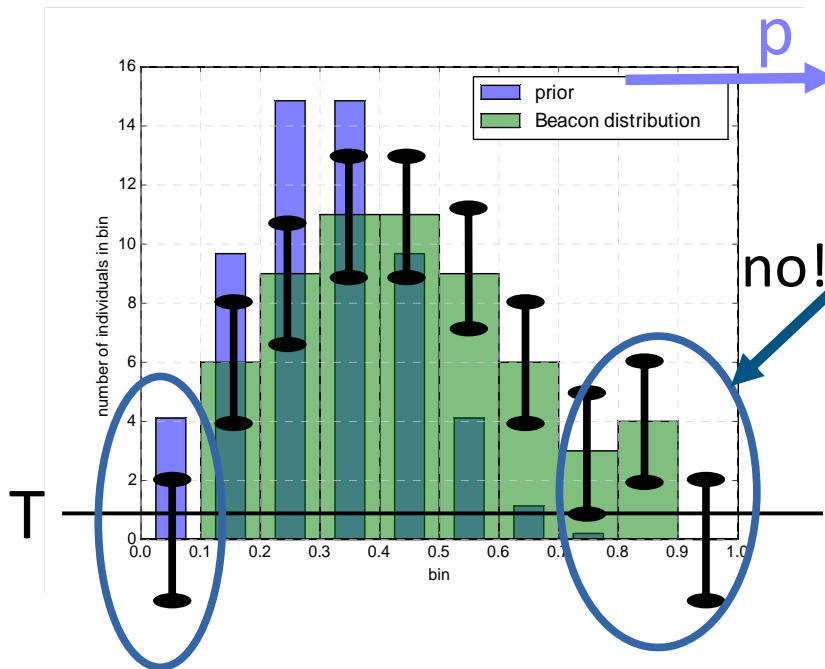
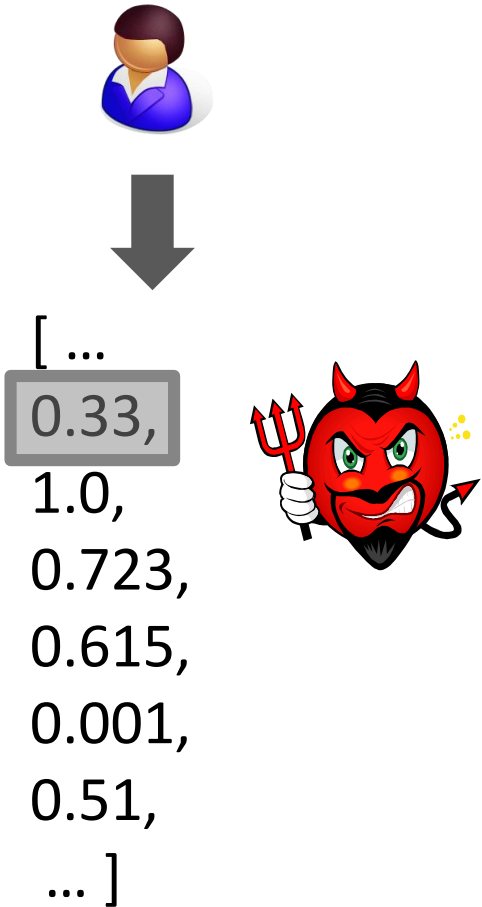


Is there 0.33 at position i ?

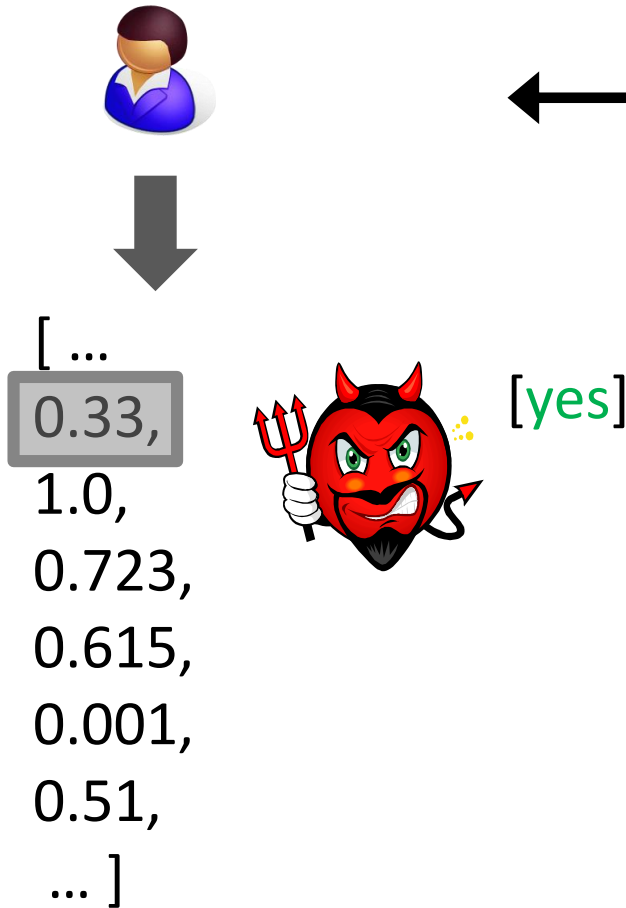


Our Defense in Detail: SVT²: Noisy Comparison

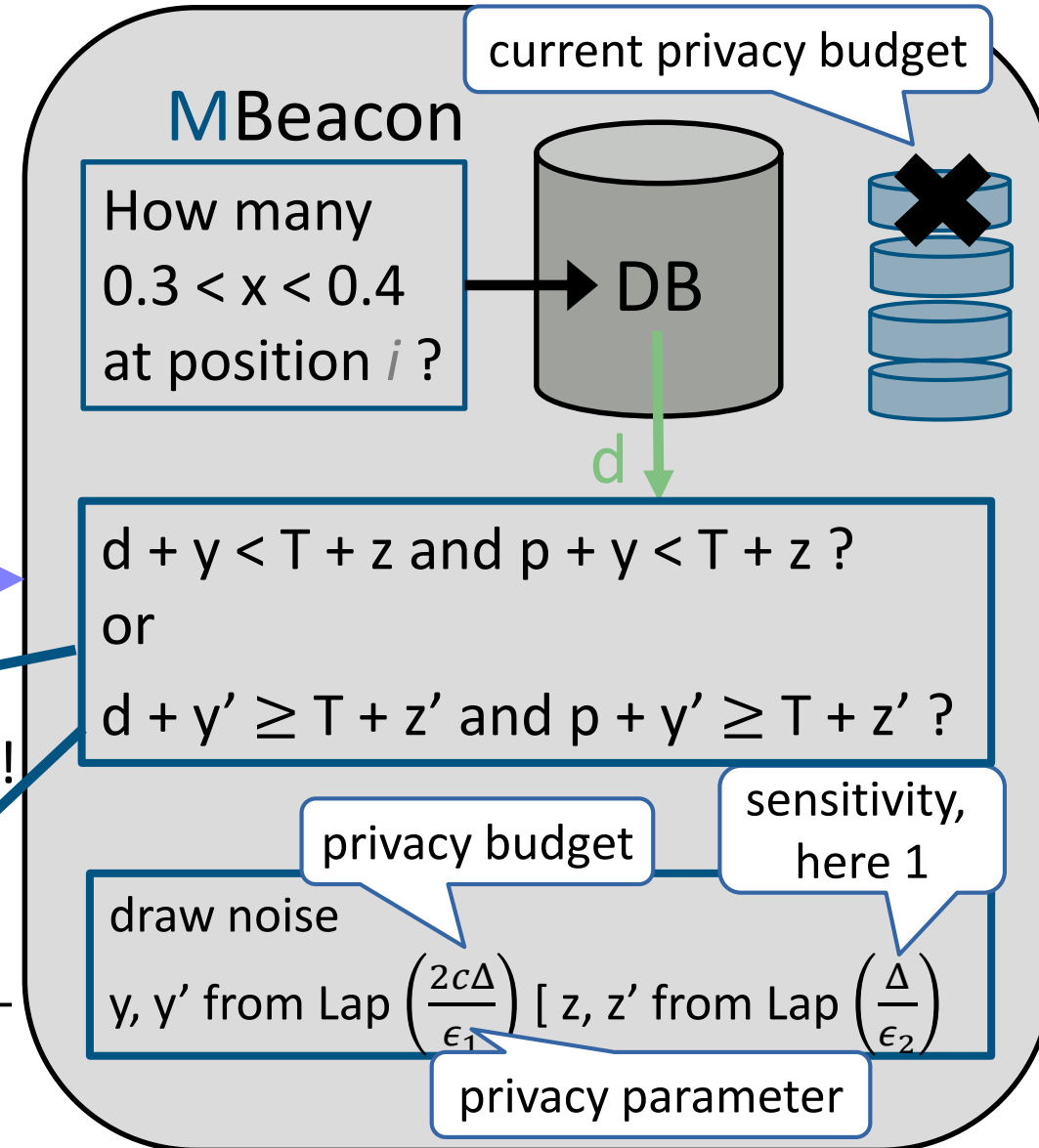
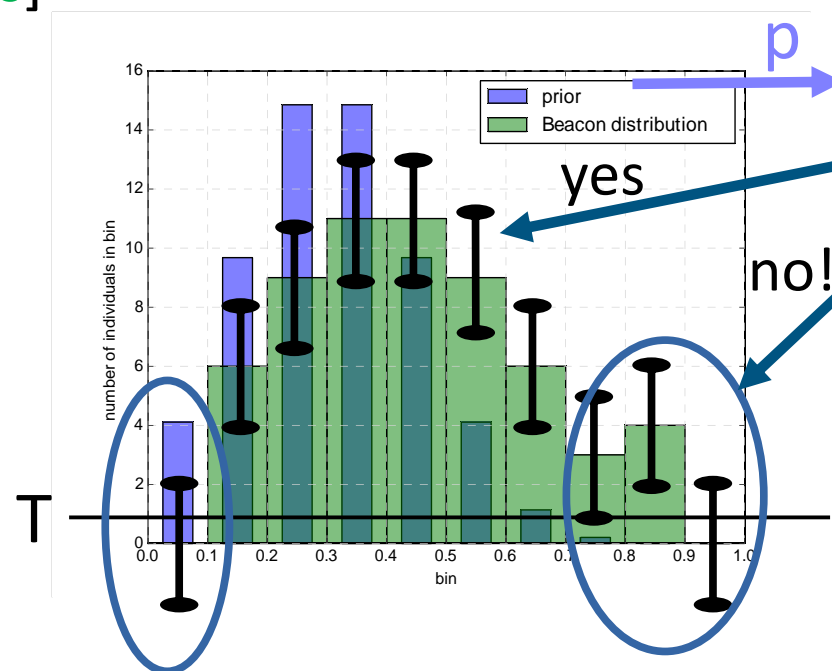
Is there 0.33 at position i ?





Our Defense in Detail: SVT²: Output of Answer







Is there 0.33 at position i ?
 probably **yes**



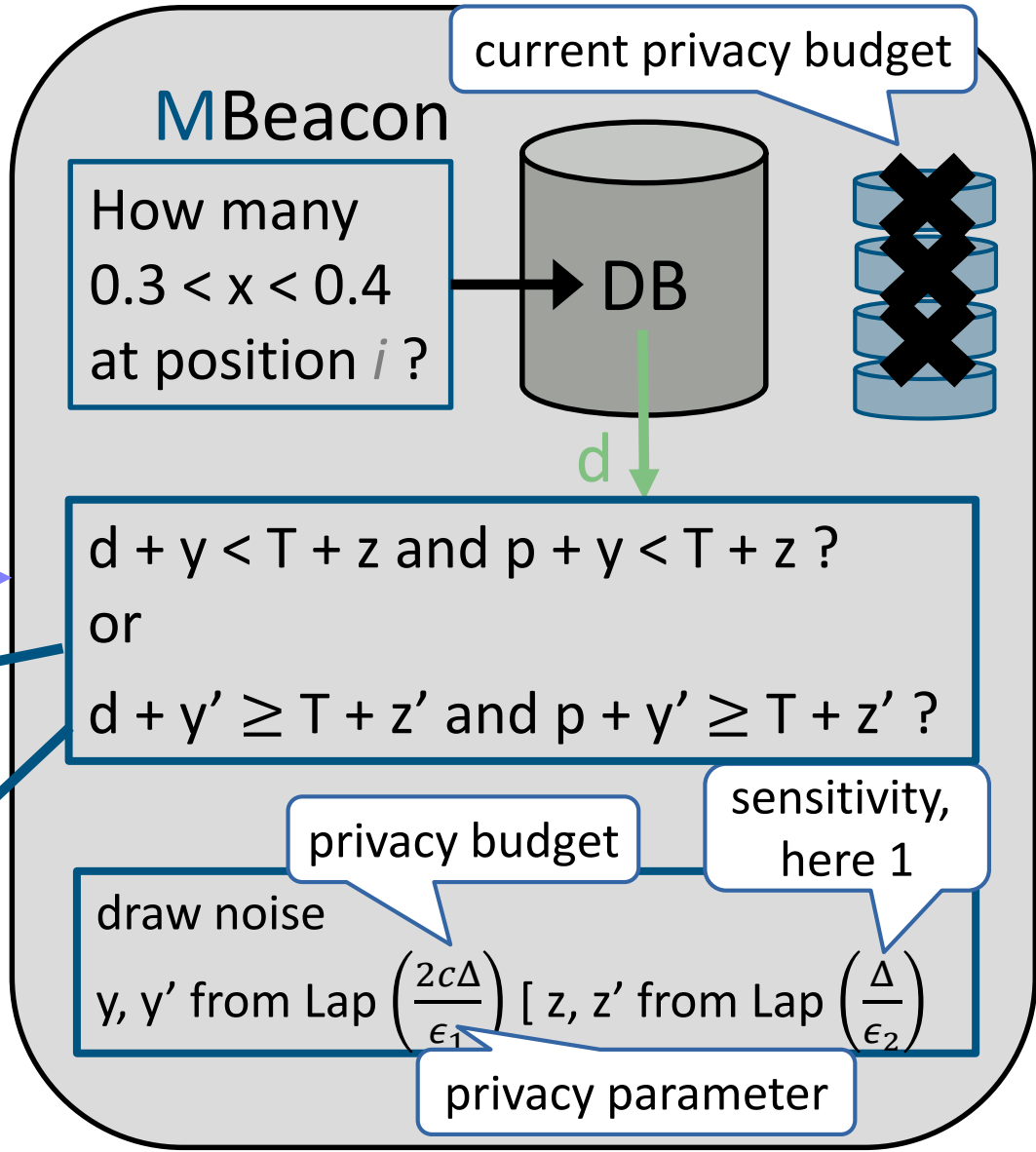
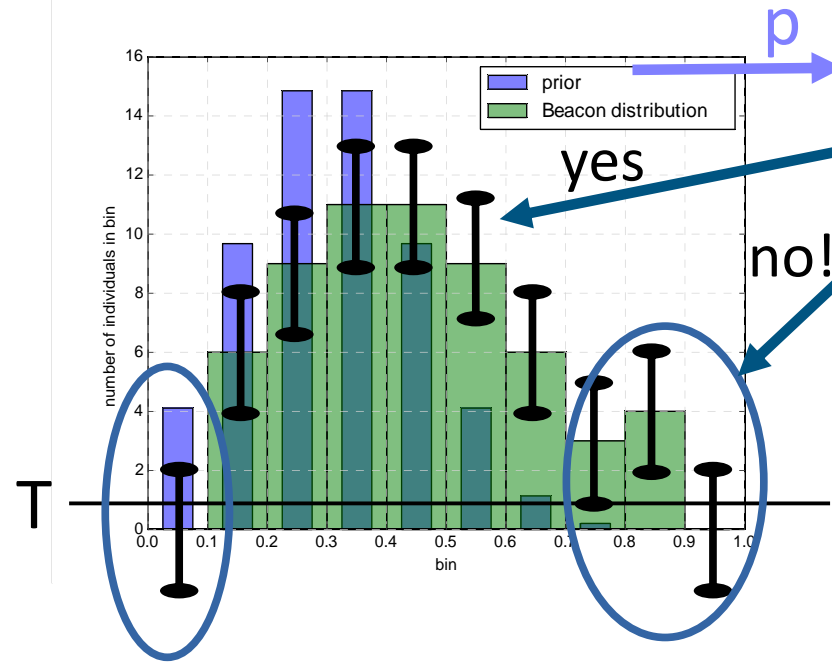
Our Defense in Detail: SVT²: Output of Answer



 [...
 0.33,
 1.0,
 0.723,
 0.615,
 0.001,
 0.51,
 ...]





Is there 0.33 at position i ? 
 probably yes 
 ...



[yes, yes, no, no, yes]



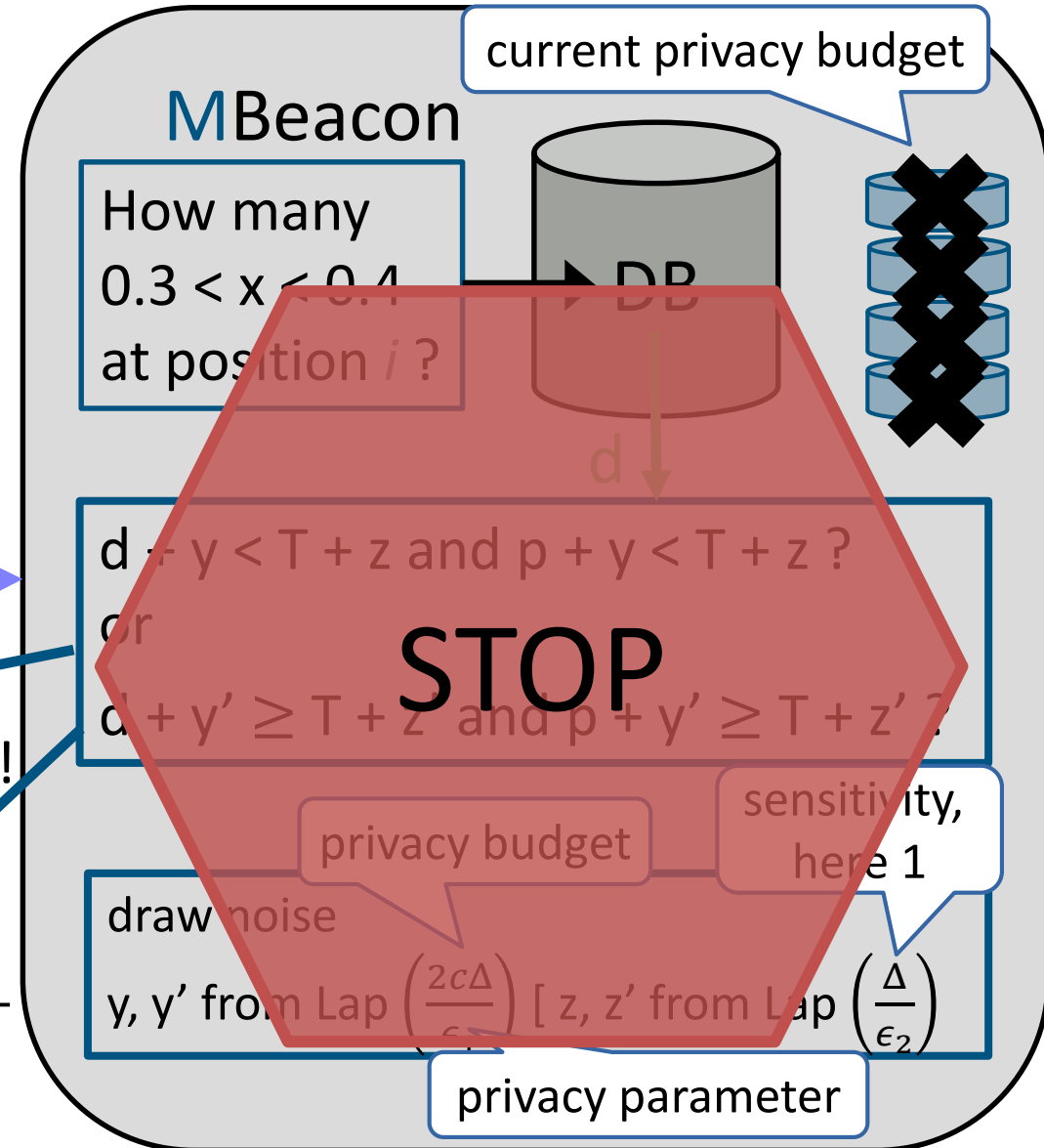
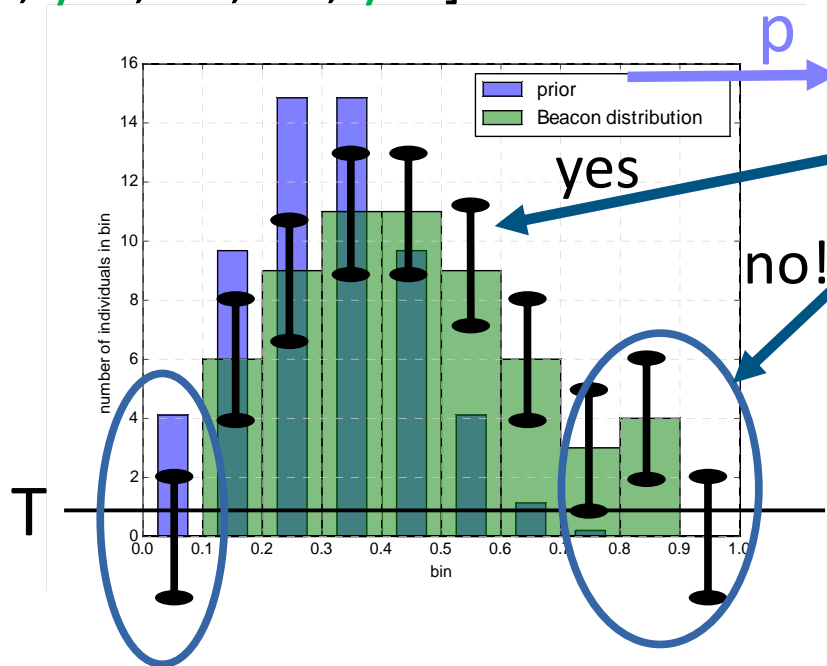
Our Defense in Detail: SVT²: Privacy Budget Depleted



 [...
 0.33,
 1.0,
 0.723,
 0.615,
 0.001,
 0.51,
 ...]



Is there 0.33 at position i ?
 ← probably **yes**
 ...
 ←
 →

[yes, yes, no, no, yes]



what is considered not to consume privacy budget

sparse vector technique (SVT):

$$d + y < T + z?$$

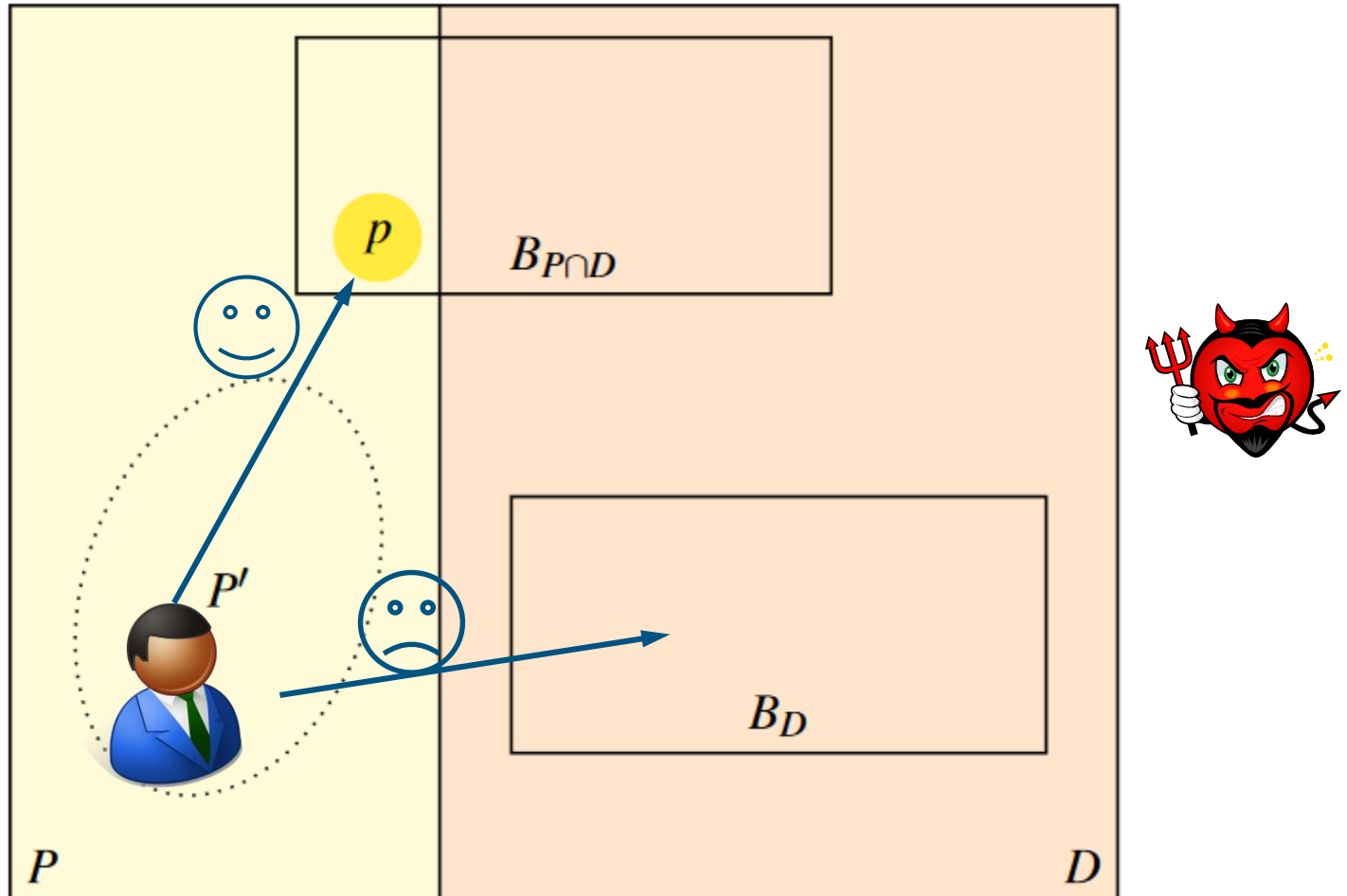
our SVT²:

$$d + y < T + z \text{ and } p + y < T + z ?$$

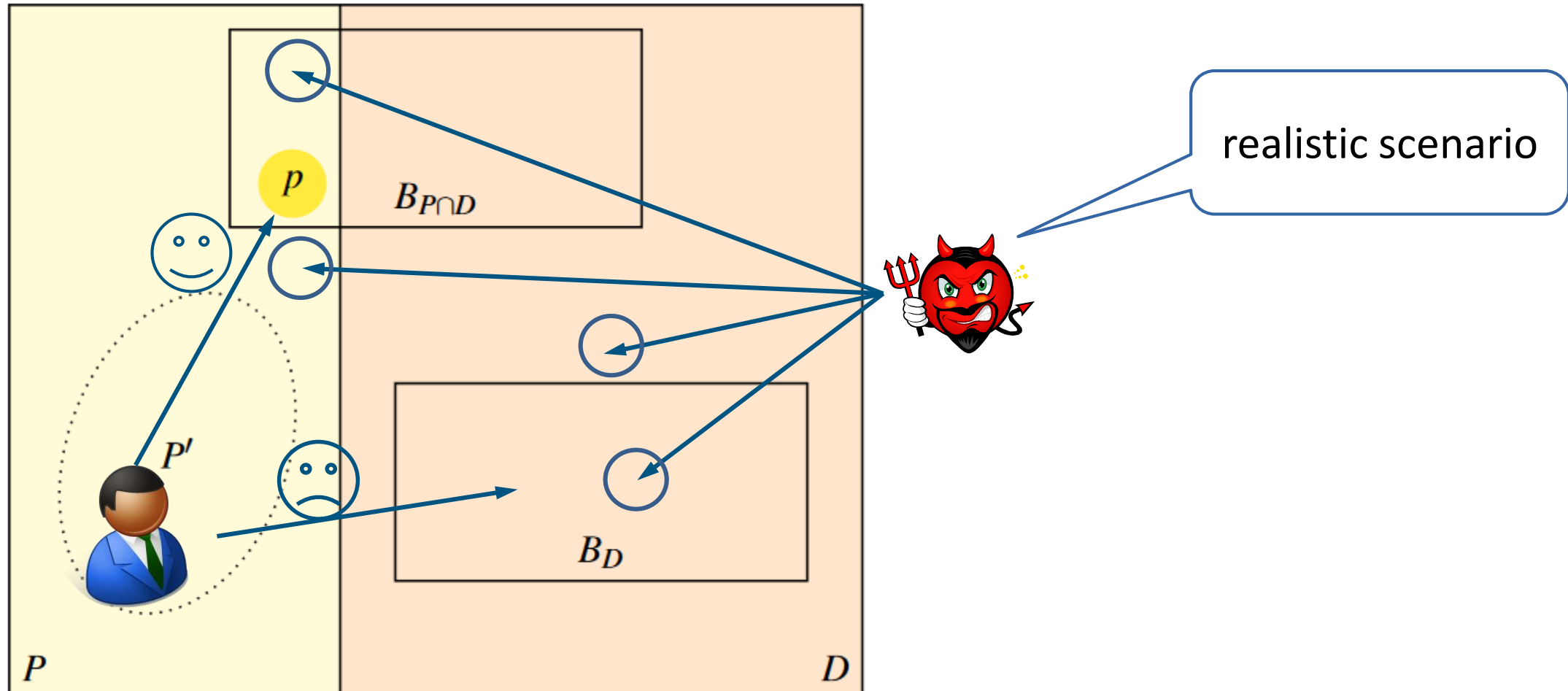
or

$$d + y' \geq T + z' \text{ and } p + y' \geq T + z' ?$$

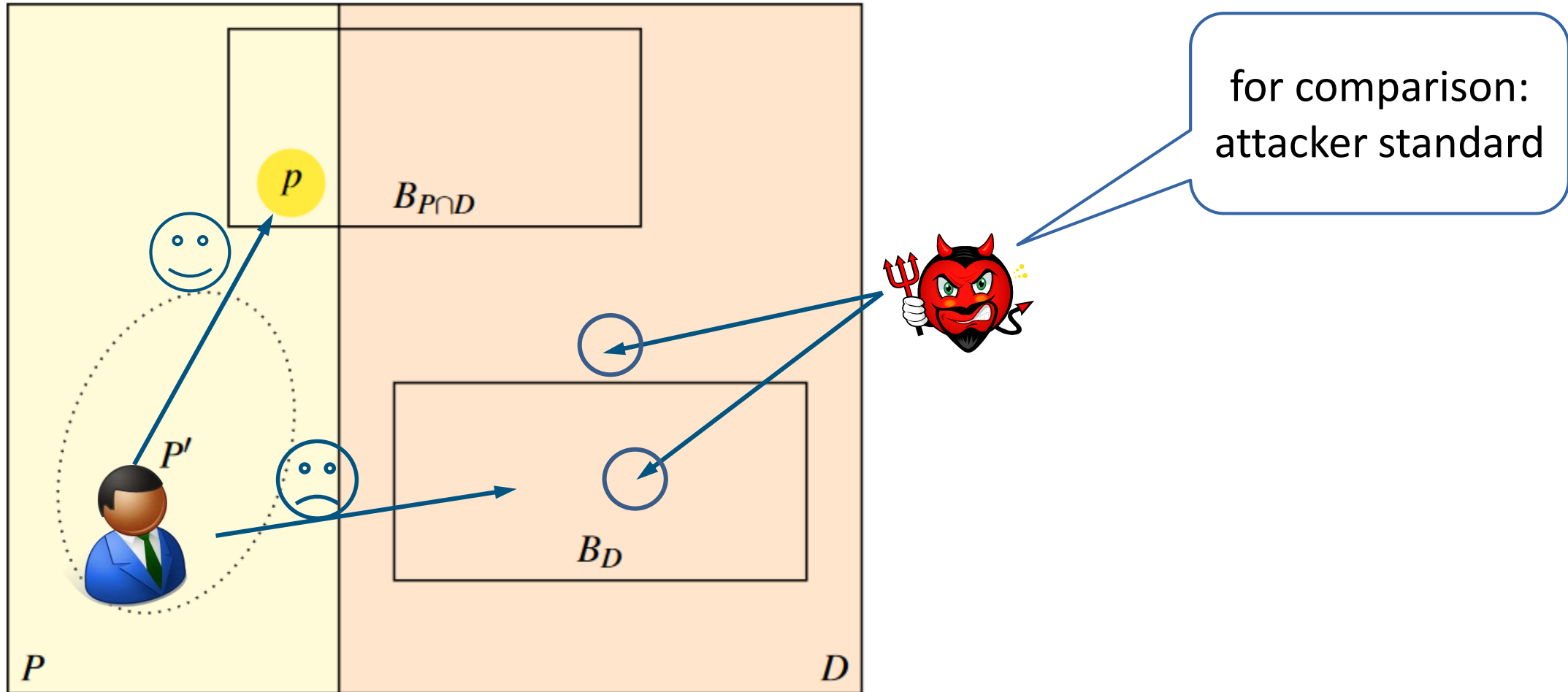
Measuring Utility: Researcher



Measuring Utility: Attacker



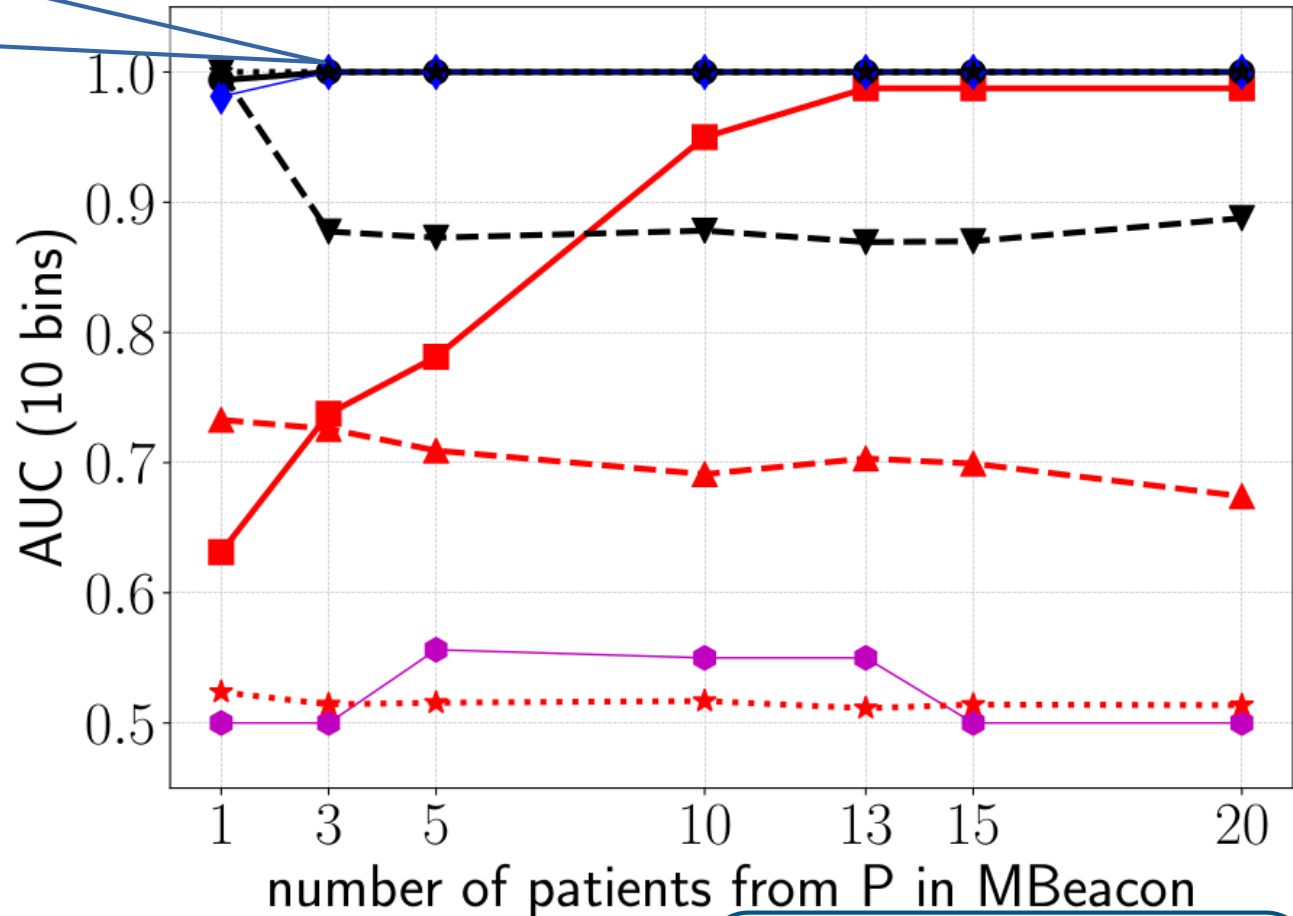
Measuring Utility: Attacker



Defense Evaluation

researcher,
no protection:
high utility

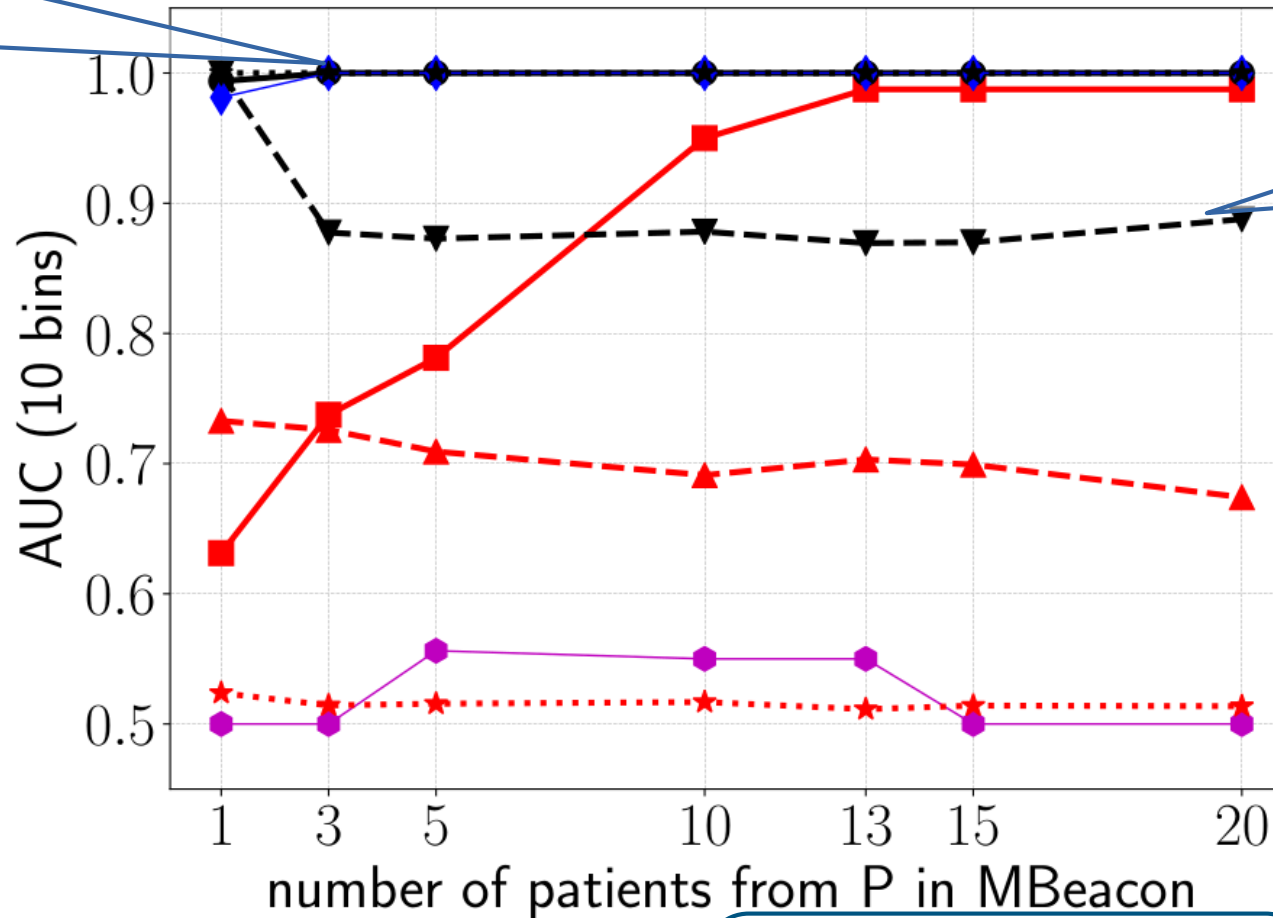
$$q, \frac{\epsilon}{c}=0.05, T=1, c=600000, \delta=1e-06$$



- SVT, researcher
- SVT, researcher (1000 q.)
- ▲- SVT, attacker
- ★ SVT, attacker standard
- unpr., researcher
- ◆ unpr., researcher (1000 q.)
- ▼- unpr., attacker
- ★ unpr., attacker standard

Defense Evaluation

$q, \frac{\epsilon}{c}=0.05, T=1, c=600000, \delta=1e-06$



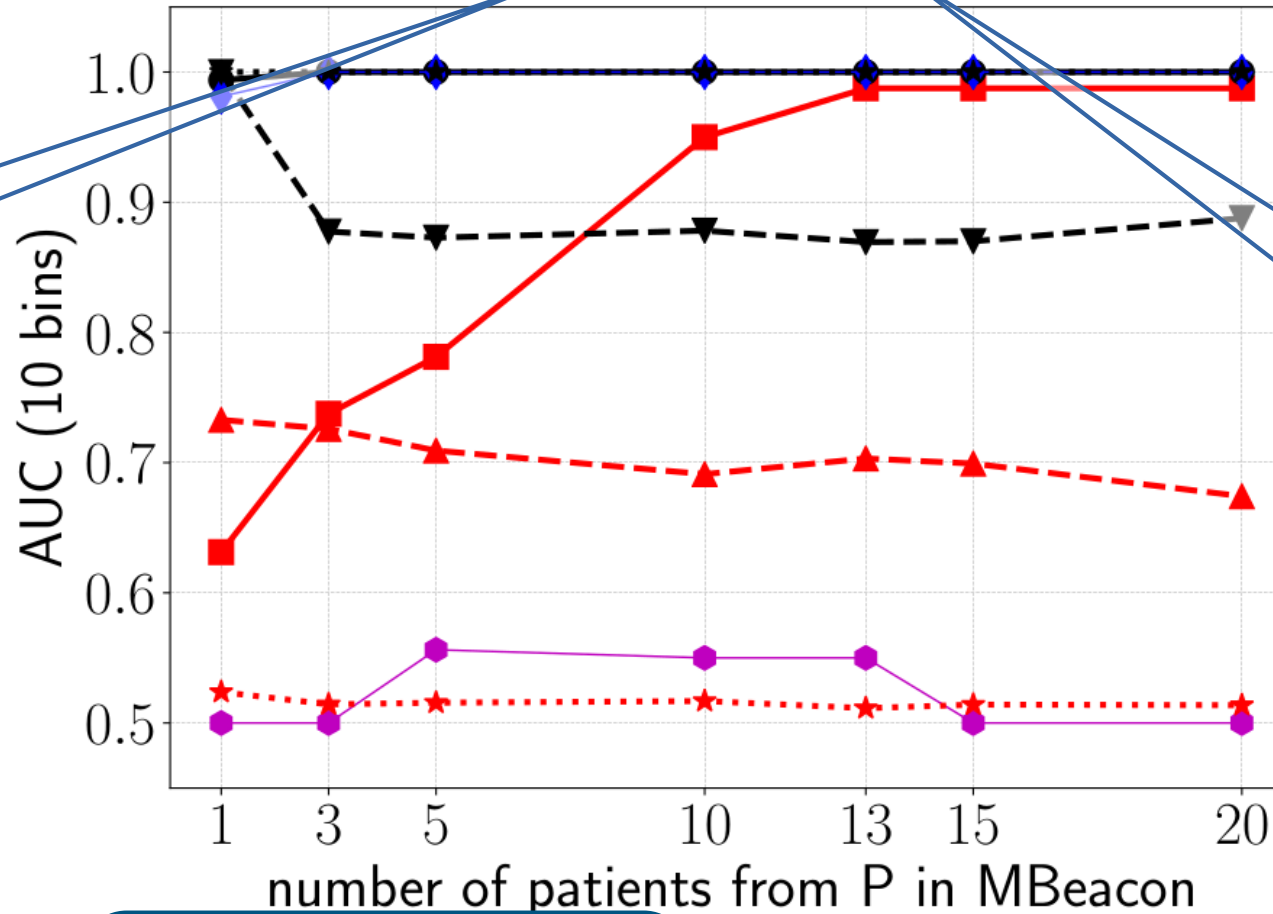
researcher,
no protection:
high utility

attacker,
no protection:
high privacy risk

- SVT, researcher
- unpr., researcher
- SVT, researcher (1000 q.)
- ◆ unpr., researcher (1000 q.)
- ▲- SVT, attacker
- ▼- unpr., attacker
- *· SVT, attacker standard
- *· unpr., attacker standard

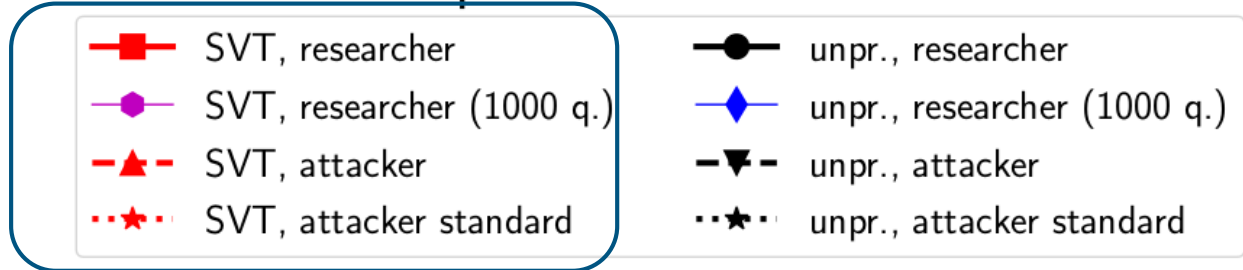
Defense Evaluation

$q, \frac{\epsilon}{c} = 0.05, T=1, c=600000, \delta=1e-06$

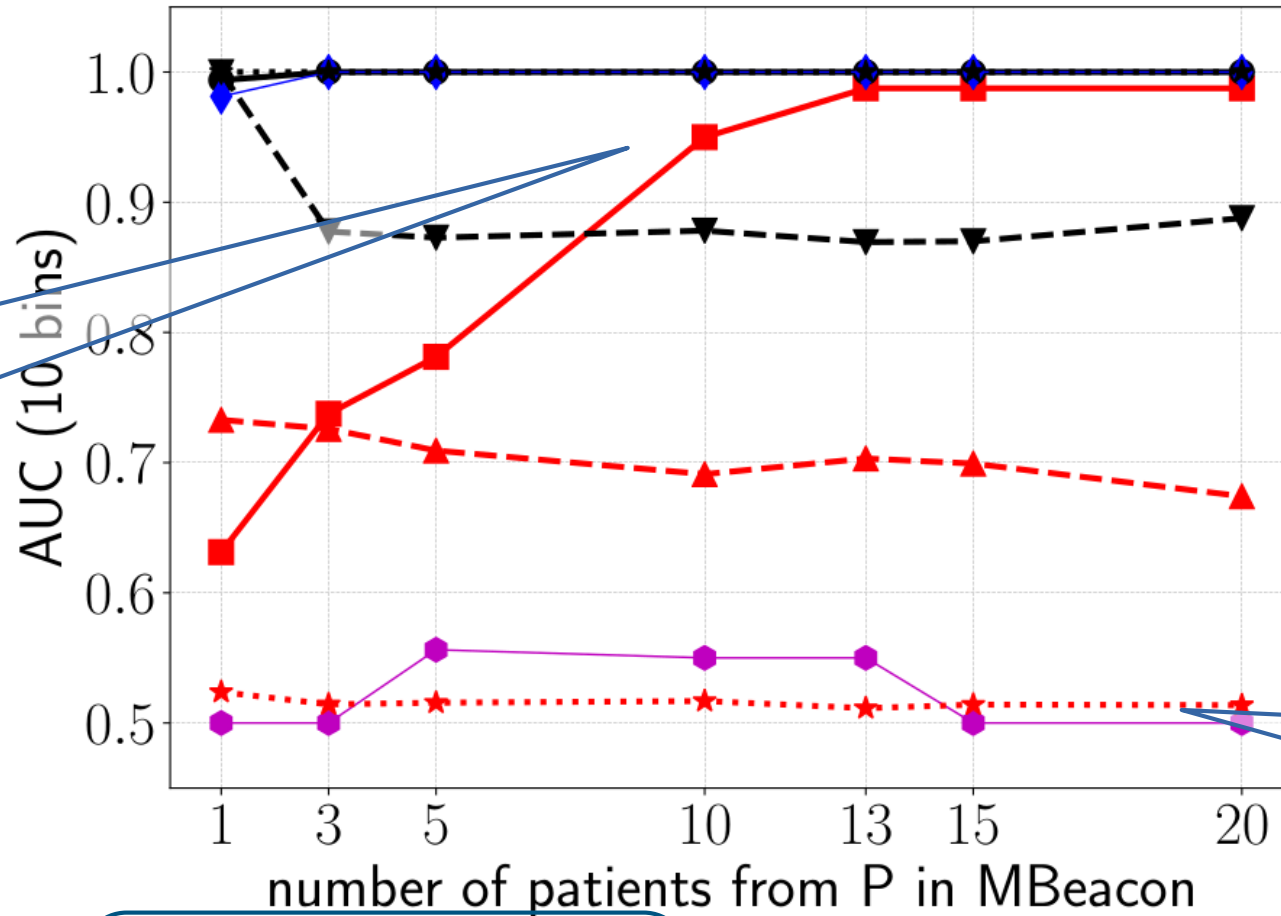


privacy parameter

privacy budget: 20% of all queries sufficient in our simulation

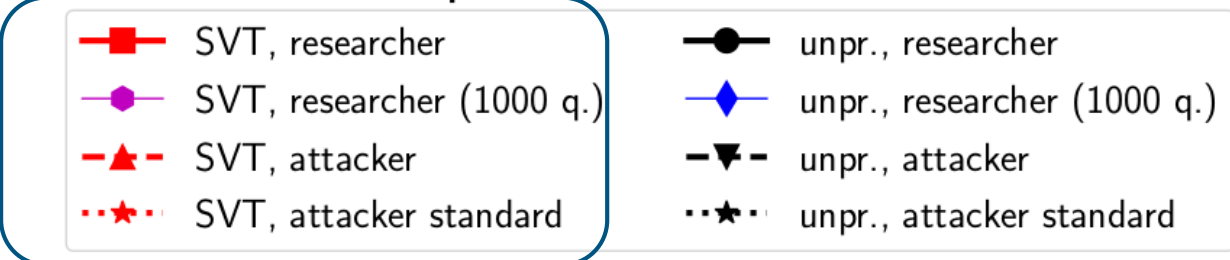


$q, \frac{\epsilon}{c}=0.05, T=1, c=600000, \delta=1e-06$

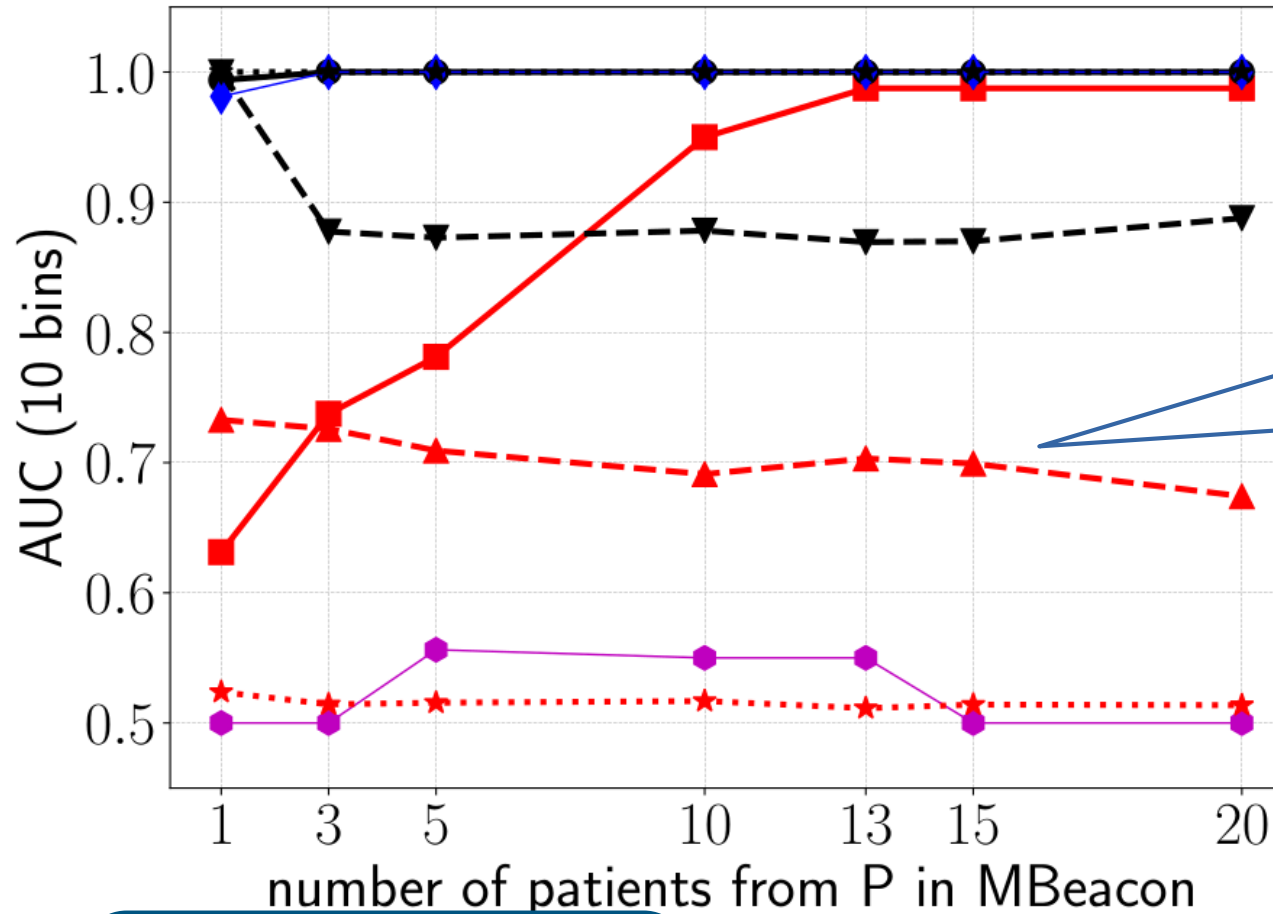


researcher,
with protection:
utility acceptable

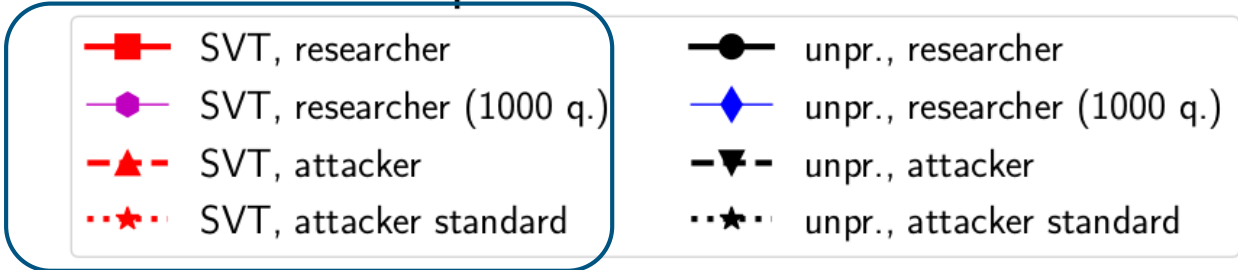
attacker,
with protection:
no privacy risk



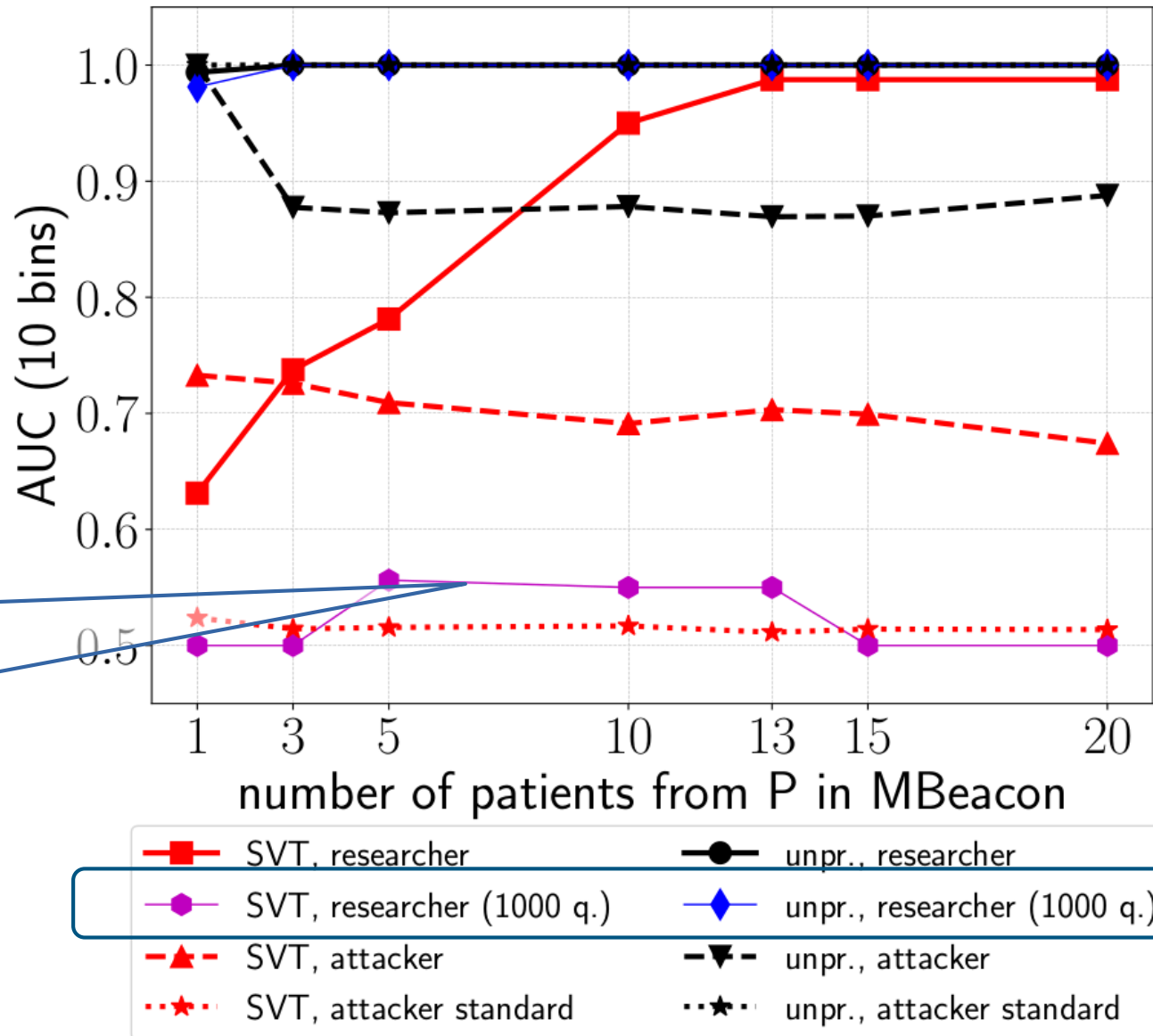
$q, \frac{\epsilon}{c}=0.05, T=1, c=600000, \delta=1e-06$



attacker, with protection realistic scenario: decreased privacy risk



$q, \frac{\epsilon}{c}=0.05, T=1, c=600000, \delta=1e-06$



researcher,
with protection:
more queries
necessary

MBeacon: Privacy-Preserving Beacons for DNA Methylation Data

Inken Hagestedt, Yang Zhang, Mathias Humbert, Pascal Berrang, Haixu Tang, XiaoFeng Wang, Michael Backes

CISPA Helmholtz Center for Information Security, Swiss Data Science Center, ETH Zurich & EPFL, Indiana University Bloomington

