CS 598 WSI, LECTURE 22

- Deployment considerations
- Adversarial Attacks in Vision
- Universal Adversarial Perturbation
- Robust UAP for wireless systems
- Defense.

Deploying ML-based Wireless

ML algorithm, evaluate performance.

⇓

accuracy, error, ...

test dataset.

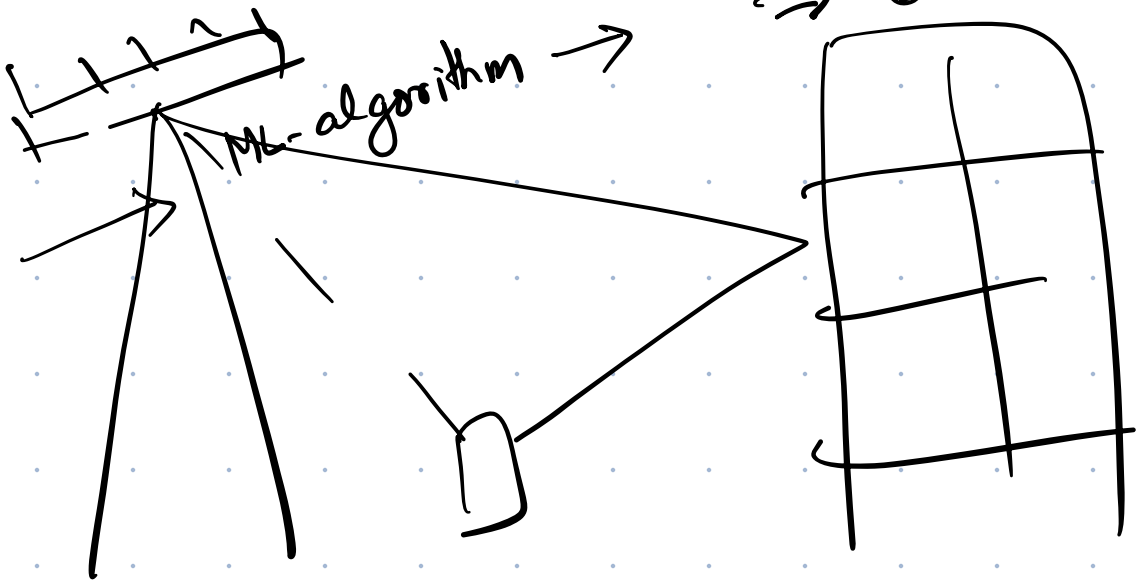
→ Interpretability / Explainability.

→ Latency / compute resources.

→ Convergence / performance "stability"

→ Generalizability / Robustness.

⇒ & distribution shifts.



Adversarial Attacks

what
Why
How.

Classification, Animal classification

$\|n\|_p$

"same" to human eyes

Output: panda

dog

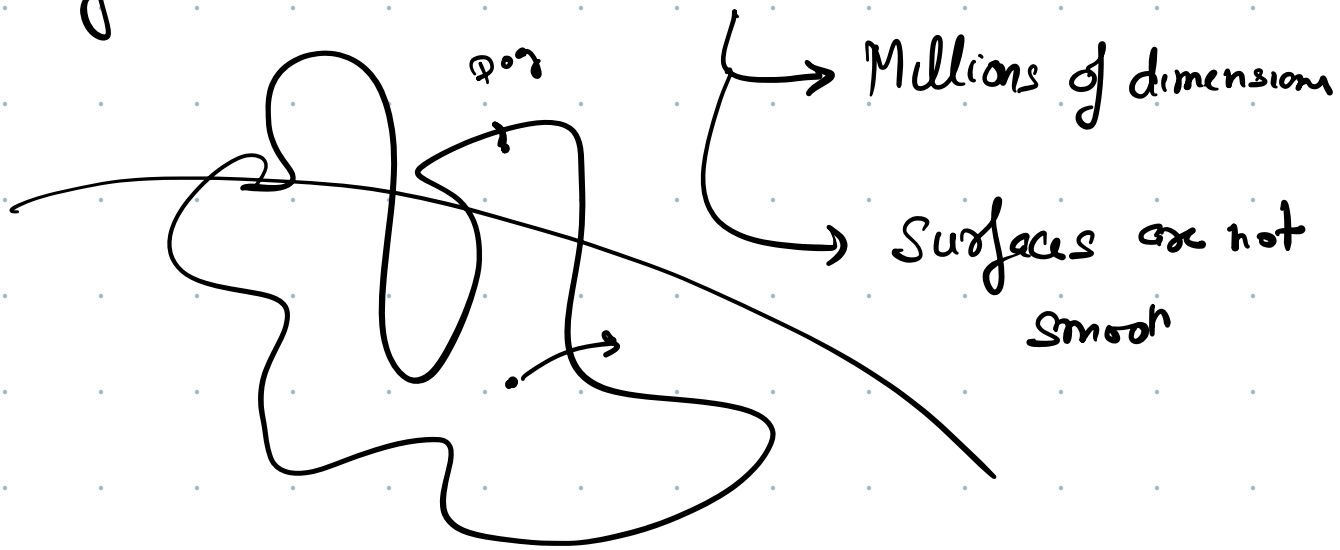
Adversarially modified image

small noise

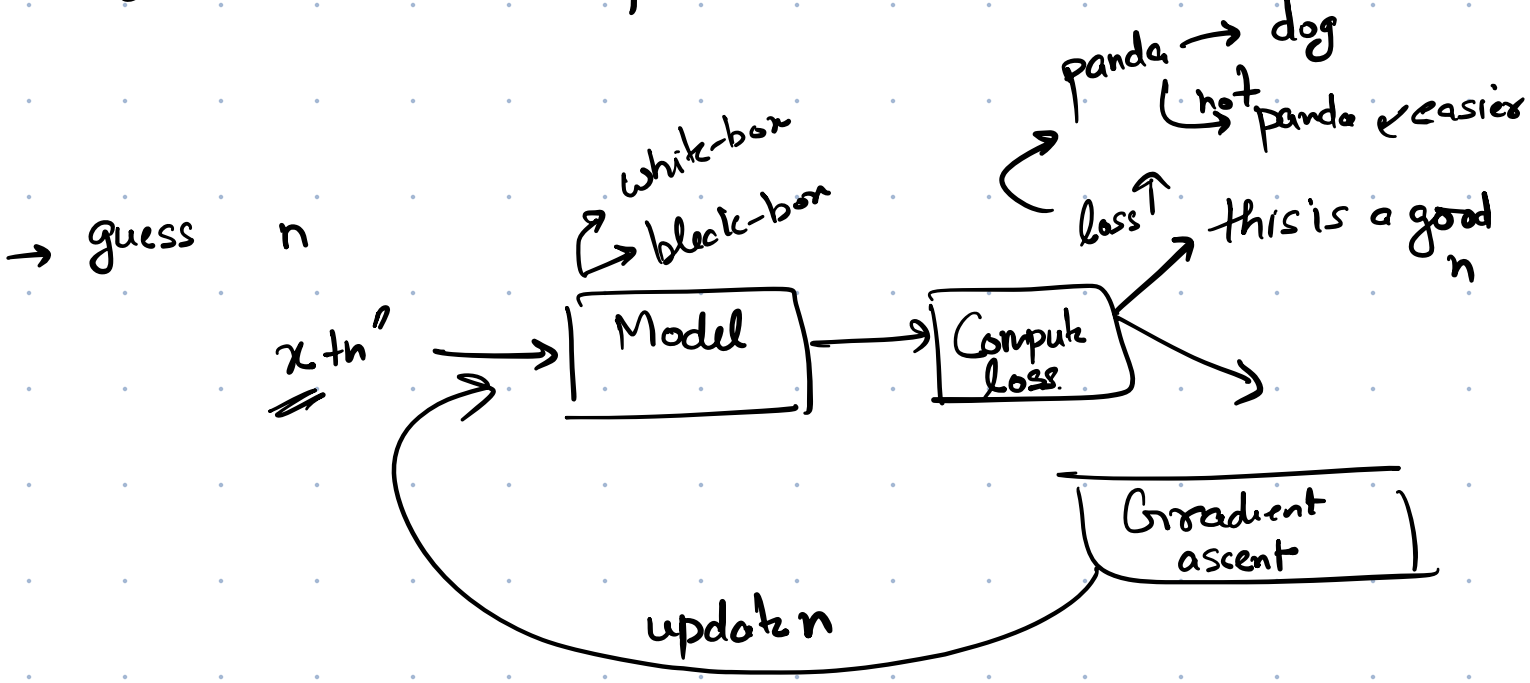
$x + n = ?$

designing this noise "carefully craft"

Why does this work?



How do I find 'n'?



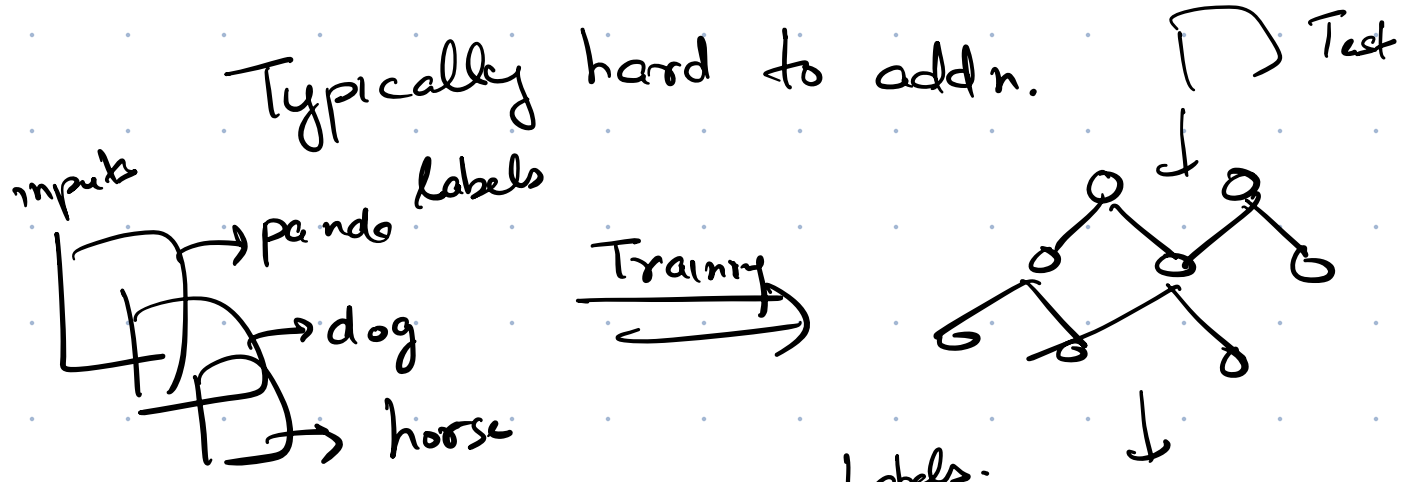
⟹ You know x

⟹ You can add n

Why is Wireless Different

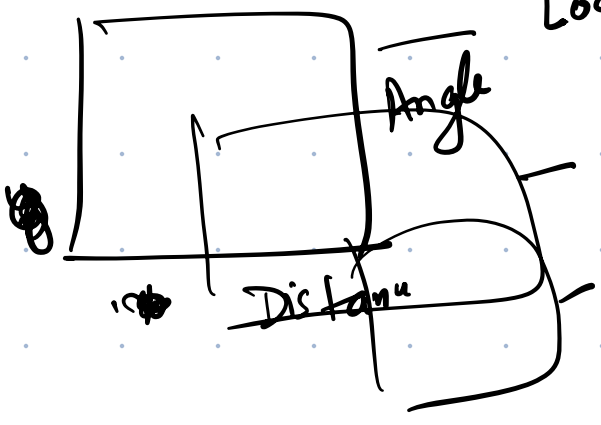
You typically do not know (x)

Typically hard to add n.



Wireless

Wireless signal



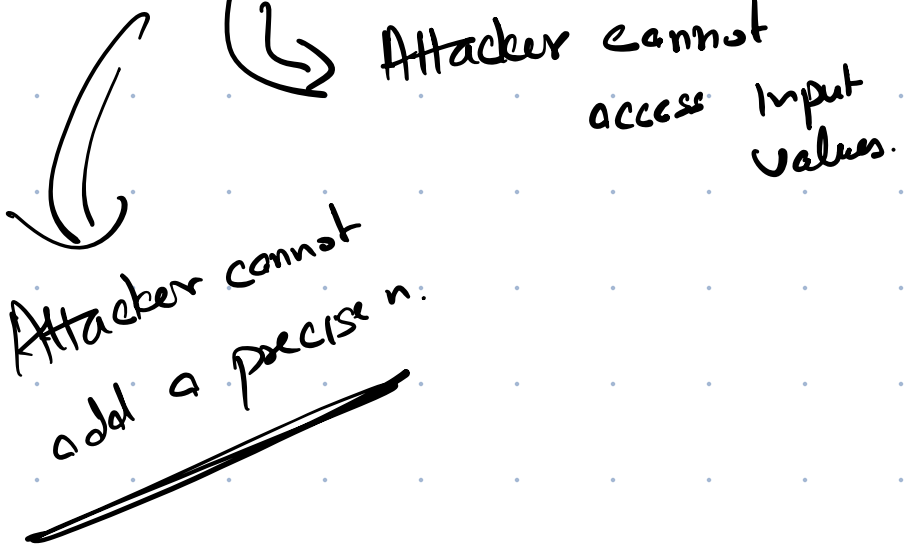
Labels.

Location (sensing)

Channel (communication)

Modulation

Wireless signal is privat by default.


Attacker cannot access input values.
Attacker cannot add a precision.

Universal Adversarial Perturbation

Don't know the input x

⇓

Find a n that works for $\forall x$

⇓

" " " realistic/probable values of x.

Sample/Choose x from a distribution

$x + n$ → Model → Loss → Gradient ascent

Average

update n

Attacker cannot add a precise n!

$n' = h_a n + n_a$

$\Downarrow$

(Reciprocity)

estimate h_a

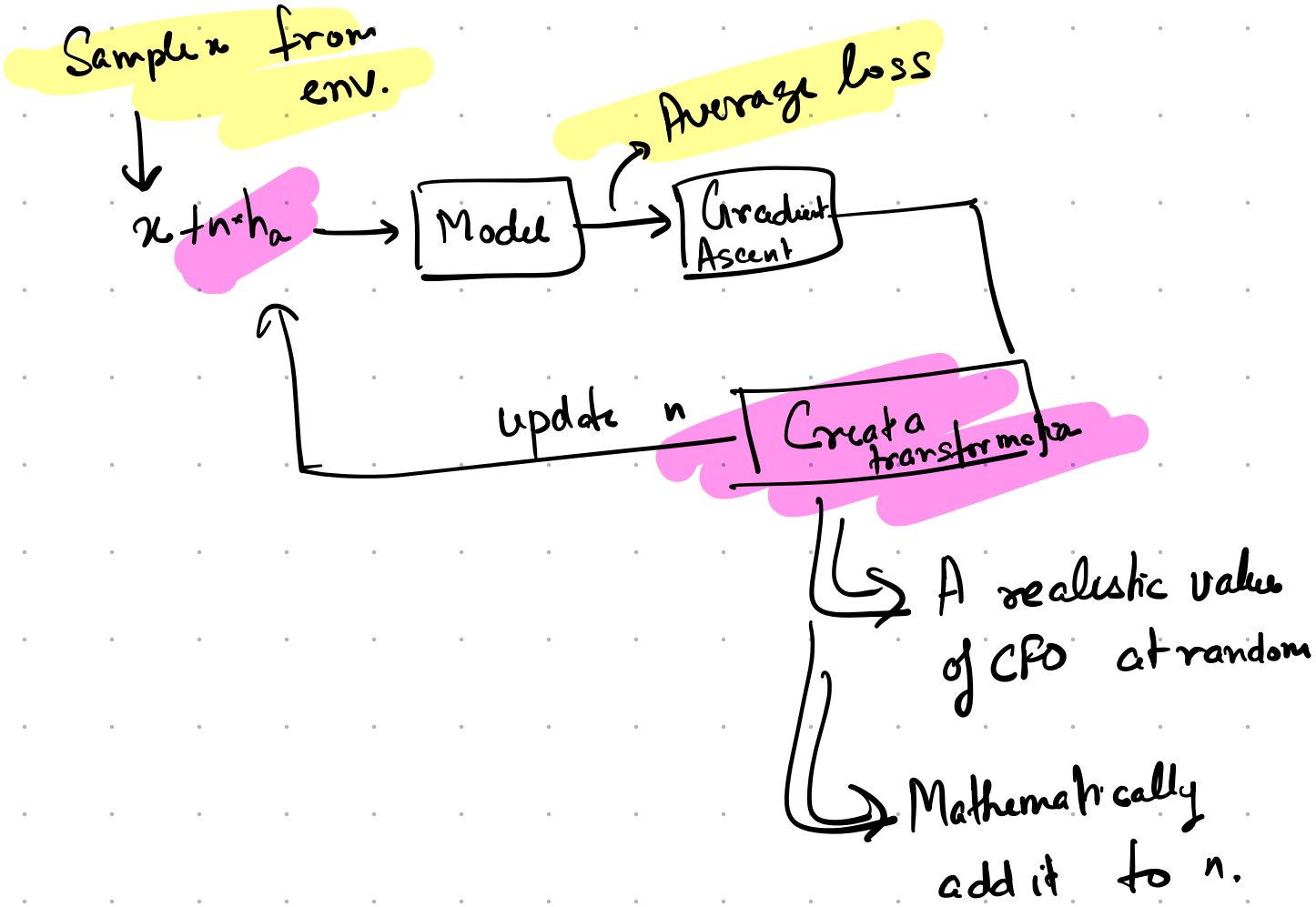
$(h_a)_{\text{attack} \to \text{BS}} \simeq (h_a)_{\text{BS} \to \text{attacker}} \alpha$

Timing & freq. offsets

- h_a
- CFO
- Packet detection delays.

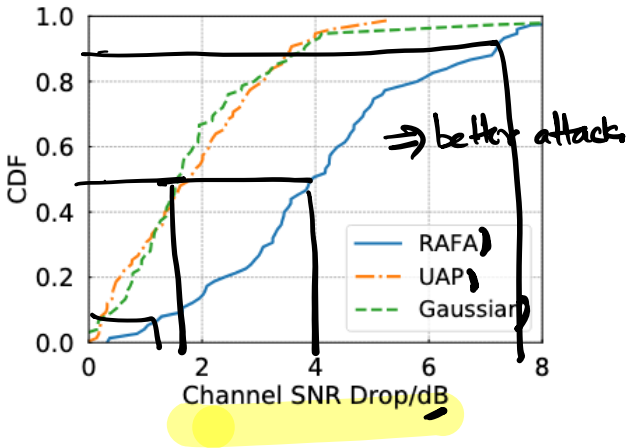
$\Rightarrow$ They can be mathematically modelled.

Robust UAP Design

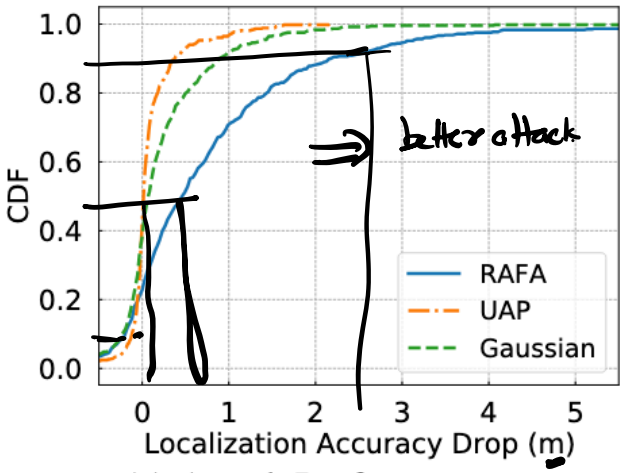


Attack Evaluation

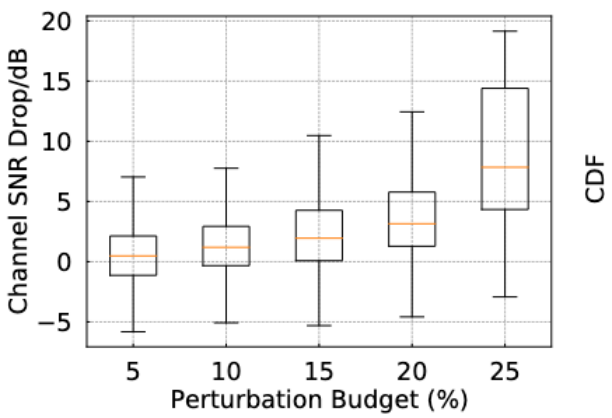
Jan:



(a) Attack Performance



(a) Attack Performance

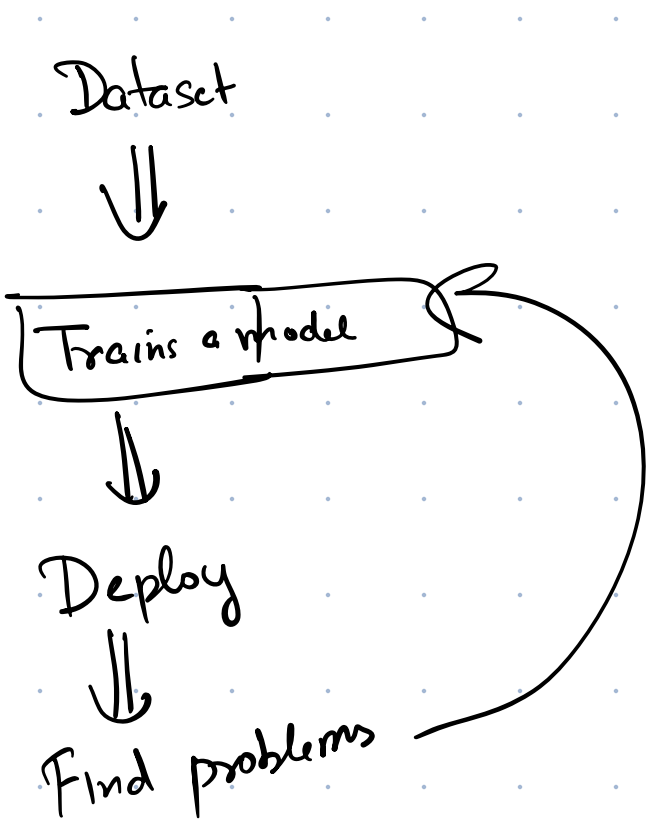


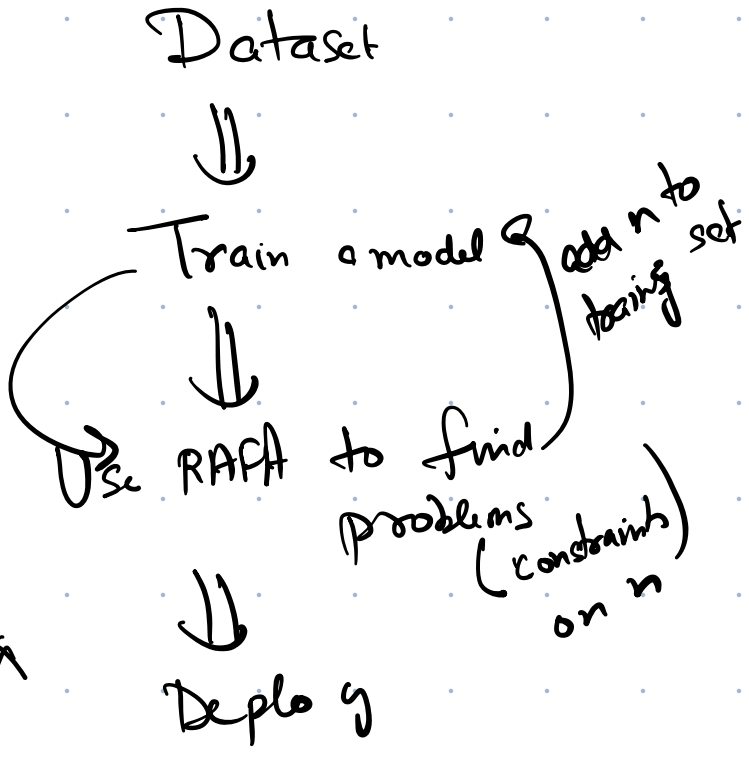
(c) Perturbation Budget Analysis

Budget (%)	10	20	30	40	50
Power Ratio-FIRE (%)	1.5	4.7	8.6	14.7	12.0
Power Ratio-DLoc (%)	1.3	8.6	20.2	34.6	51.8

Table 2: *Power Ratio vs Budget*

Potential Defense





Eventual goal: Design models that come with certificates.