

Problem Space

Feature Space

Phishing website



Modified website,
evaded the detector



Unimplementable

Original [#1]

isMailto	PctNull Hyperlinks	PctExt Hyperlinks	...
0	1.0	0.0	...

Perturbed [#2]

1	1.0	0.0	...
---	-----	-----	-----

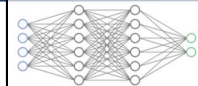
Perturbed & Infeasible [#3]

0.5	1.0	0.2	...
-----	-----	-----	-----



Evasion
Attack

Phishing
Detector



PHISHING

BENIGN

BENIGN

$$x_{\#3} \not\models \forall x: x.isMailto \in \{0,1\}$$

Structure-constraints **violation**

$$x_{\#3} \not\models \forall x: x.pctNullHyperlinks = 1.0 \rightarrow x.pctExtHyperlinks = 0.0$$

Relation-constraints **violation**
(can be expressed by DCs)