CERIAS
The Center for Education and Research in Information Assurance and Security

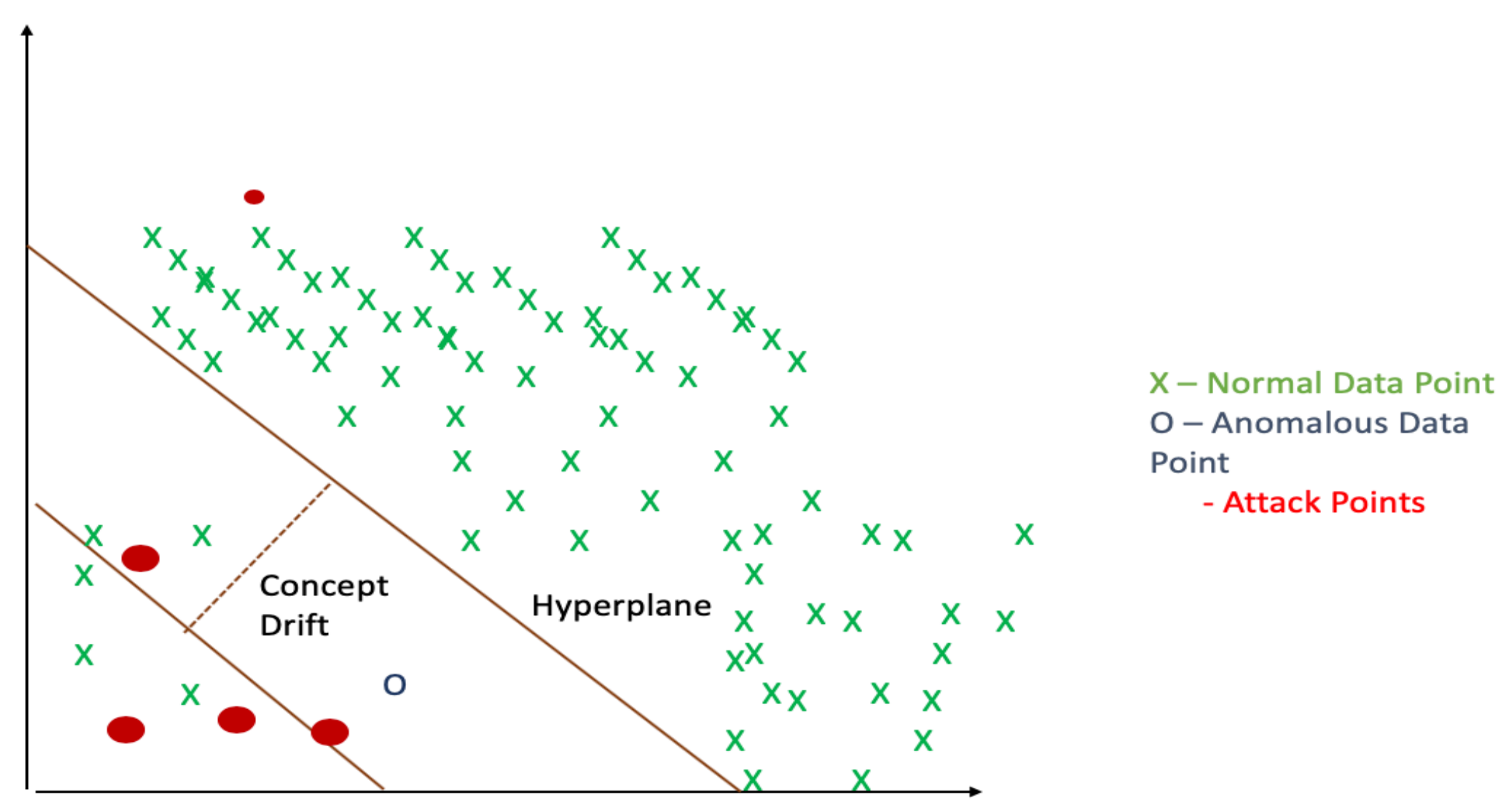
Poisoning Attacks Against SVM based Anomaly Detection Techniques

Radhika Bhargava & Chris Clifton
Department Of CERIAS and Computer Science, Purdue University

PROBLEM

Adversary wants to hide from anomaly detection

- Adversary is unable to change their own data and still have the attack achieve its goals
- Adversary can create fake points.
- How do we estimate the risk poised by such an attack?



ATTACK MODEL

Adversary's goal -

- Attacker wants to perform a targeted, integrity violation.

Adversary's knowledge –

- "The enemy knows the system".
- The adversary has knowledge of the training algorithm.
- Partial or complete information about the training set, such as its distribution.

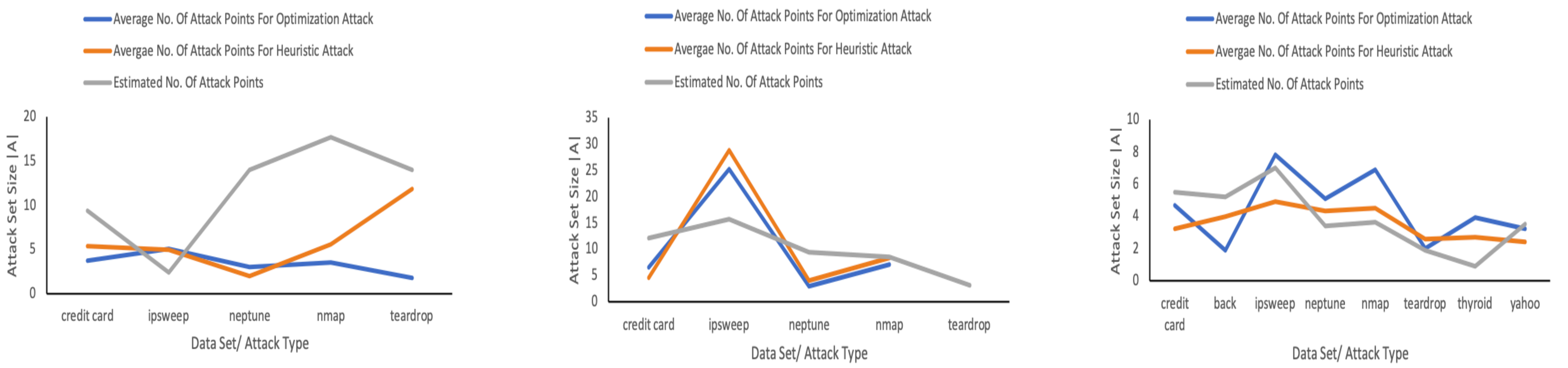
Adversary's capability -

- The adversary can poison the dataset.

Attack Strategy –

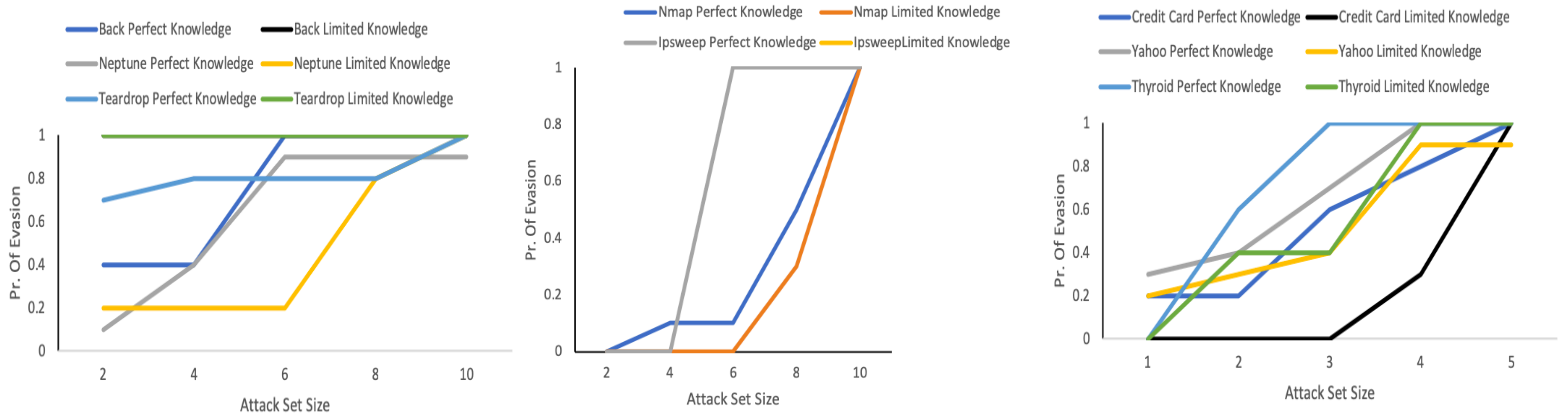
- Make the neighborhood of the anomaly point a denser so that it "looks-like" a normal data point.

RESULTS



a b c

Expected Poison Points needed for Adversary to defeat SVM based Anomaly Detection (a) Linear Kernel (b) Polynomial Kernel (c) Gaussian Kernel



a b c

Probability of Evasion vs. Size of the Attack Set -(a) KDD Cup'99 Dataset dos attacks (b) KDD Cup'99 Dataset u2r\ attacks (c) Yahoo S5, Thyroid & Credit Card Anomaly Detection Dataset

Key Observations:

1. An adversary needs to control 0.01% of the training data set tor a targeted attack.
2. The evasion rate increases to 100%.

PURDUE UNIVERSITY
Discovery Park
CERIAS