Adversarial Gaussian Process Regression in Sensor Networks

Yi Li, Yevgeniy Vorobeychik, Xenofon Koutsoukos
Vanderbilt University

INTRODUCTION

- Consider machine learning models for anomaly detection based on Gaussian process regression.
- Define stealthy attacks and investigate the feasibility of designing undetectable attacks with catastrophic potential damage.
- Design resilient anomaly detectors for stealthy attacks based on the game theoretical framework.

ANOMALY DETECTION

- Sensor network:

$$< y_1, y_2, \dots, y_n >$$

- A collection of predictors:

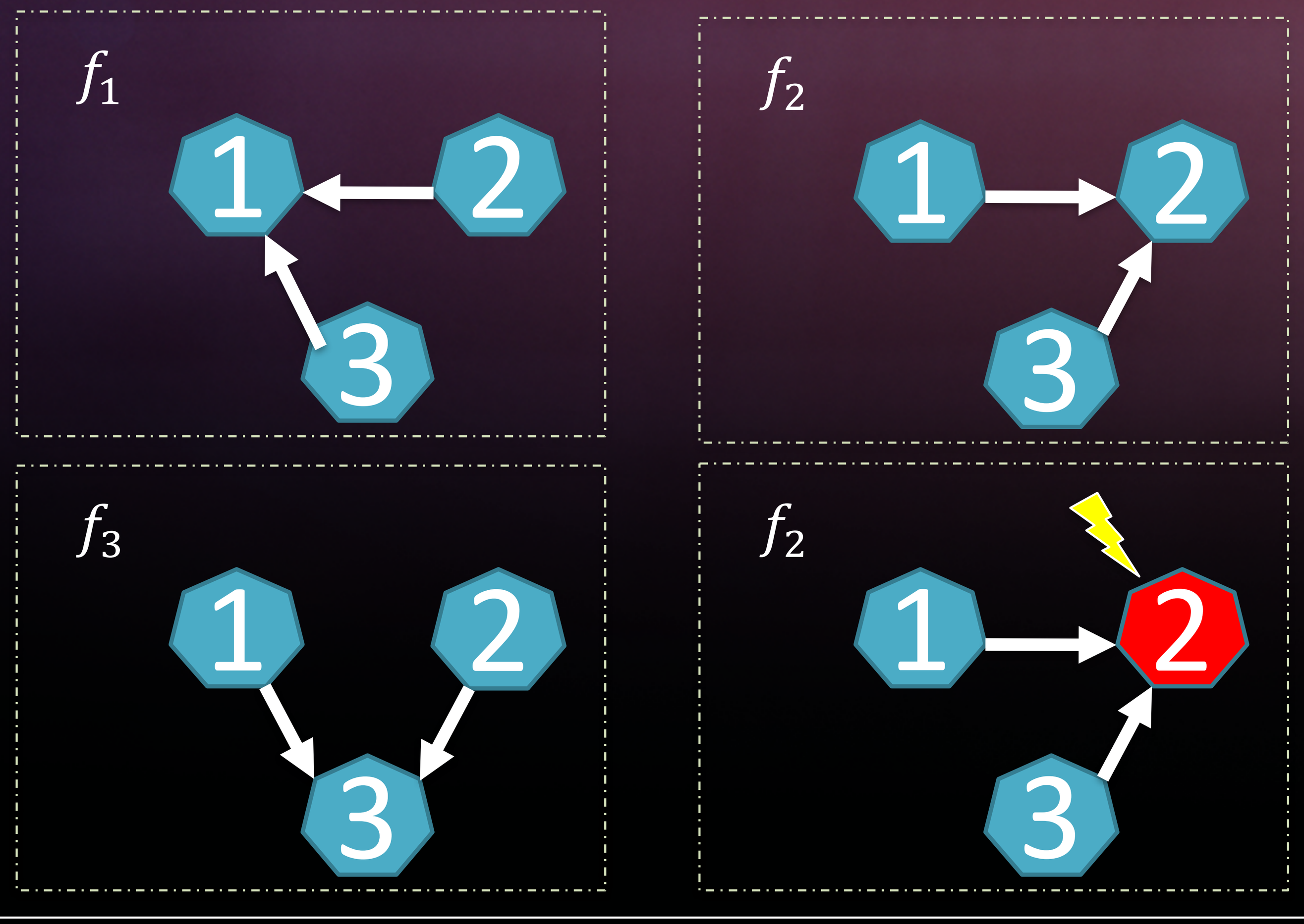
$$f_i(\tilde{y}_{-i}) \sim N(\mu_i(\tilde{y}_{-i}), \sigma_i(\tilde{y}_{-i}))$$ ← predictions are Gaussian

f: gaussian process regression $\tilde{y}_{-i}$: the readings of the sensors other than i

- Anomaly behaviors

$$A = \Phi^{-1}\left(1 - \frac{\alpha_i}{2}\right)\sigma_i(\tilde{y}_{-i})$$

$$\exists i, \tilde{y}_i \notin (u_i(\tilde{y}_{-i}) - A, u_i(\tilde{y}_{-i}) + A)$$ ← α_i confidence interval



STEALTHY ATTACKS

- Find undetectable attacks via optimization approaches:

$$\underset{\Delta\tilde{y}}{argmin} \,/\, \underset{\Delta\tilde{y}}{argmax}\, \Delta\,\tilde{y}_s$$

Objective: maximizing the deviation of the reading of the targeted sensor

$$s.t.$$
$$\forall i,$$
$$(\tilde{y}_i + \Delta\tilde{y}_i) > u_i(\tilde{y}_{-i}) - A$$
$$(\tilde{y}_i + \Delta\tilde{y}_i) < u_i(\tilde{y}_{-i}) + A$$
$$|\tilde{y}_i + \Delta\tilde{y}_i| < D_i$$

Stealthy attack: avoiding being detected via modifying the readings of correlated sensors

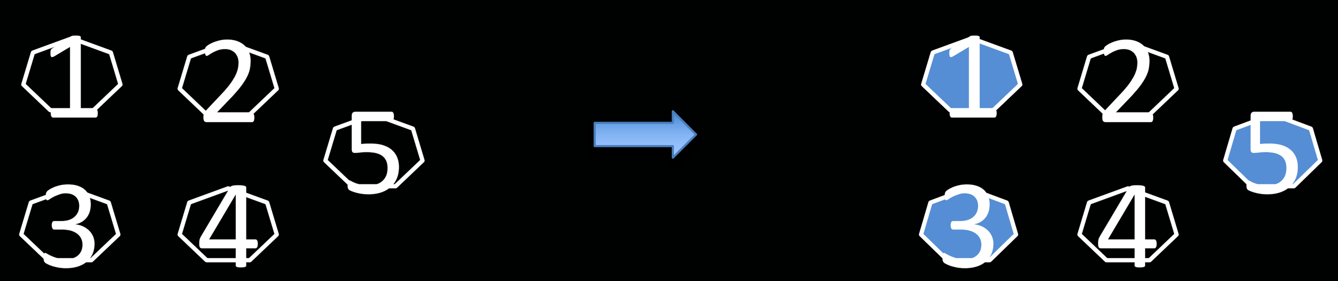
$$||\Delta\tilde{y}||_0 \le H$$

Budget: the number of changeable sensors

Issue: non-linear, non-convex, solved via local linear approximation and feasible direction searching.

RESILIENT ANOMALY DETECTORS

- The action space of the defender:
  - Sensor placement pattern



  - Tolerances: the confidence level, $< \alpha_1, \dots, \alpha_n >$
- The objective of the defender:

$$\max_{\theta,\alpha} \lambda\, G(\theta,\alpha) + (1-\lambda) S(\theta,\alpha)$$

θ: sensor placement pattern, α: $< \alpha_1, \dots, \alpha_n >$, λ: trade-off parameter

  - the predictor's non-false alarm rate, $G(\theta, \alpha)$
  - The impact of attacks, $S(\theta, \alpha)$
- Finding optimal (θ, α)
  - The Defender moves first
  - Stackelberg game equilibrium

Solved via random greedy algorithm.

RESULT

- Data: Tennessee Eastman problem. The temperature, liquid level and pressure sensors among the reactor, the product separator and the stripper.
- Targeted sensor: Reactor pressure

Computational Cybersecurity in Compromised Environments
2018 Fall Workshop | September 17-19 | Atlanta, Georgia