Learning-based Attacks in Cyber-Physical Systems

MJ Khojasteh, PhD
Center for Autonomous Systems and Technologies (CAST)
California Institute of Technology

NASA's Jet Propulsion Laboratory (JPL)
April 28 2020

Cloud robots and automation systems

Security

We need to address physical security in addition to cyber security

News reports

Port of San Diego suffers cyber-attack, second port in a week after Barcelona

Hacker jailed for revenge sewage attacks

Job rejection caused a bit of a stink

HACKERS REMOTELY KILL A JEEP ON THE HIGHWAY—WITH ME IN IT

News reports

The Stuxnet outbreak

A worm in the centrifuge

The Economist

An unusually sophisticated cyber-weapon is mysterious but important

Computer virus Stuxnet a 'game changer,' DHS official tells Senate

CNN

"It has changed the way we view the security threat"

Proof of concept, Symantec

Zero Days Documentary (2016)

The man in the middle

A fictitious plant for the controller

A malicious controller for the plant

Mathematical formulation

- Linear dynamical system

$$X_{k+1} = aX_k + U_k + W_k$$

$$\{W_k\} \text{ are i.i.d. } \mathcal{N}(0, Var[W])$$

- The controller, at time k, observes Y_k and generates a control signal U_k as a function of all past observations Y_1^k.

$Y_k = X_k$ Under normal operation

$Y_k = V_k$ Under attack

- The attacker feeds a malicious input $\tilde{U}_k$ to the plant.

- How can the controller detect that the system is under attack?

Anomaly detection

- The controller is armed with a detector that tests for anomalies in the observed history Y_1^k.

$$X_{k+1} = aX_k + U_k + W_k \qquad \{W_k\} \text{ are i.i.d. } \mathcal{N}(0, Var[W])$$

- Under legitimate system operation $(Y_k = X_k)$ we expect

$$Y_{k+1} - aY_k - U_k(Y_1^k) \sim \quad \text{i.i.d.} \quad \mathcal{N}(0, Var[W])$$

- The detector performs the variance test

$$Var[W] = \mathbb{E}[W^2]$$

Anomaly detection

- Under legitimate system operation we expect

$$Y_{k+1} - aY_k - U_k(Y_1^k) \sim \quad \text{i.i.d.} \quad \mathcal{N}(0, Var[W])$$

- The controller performs a threshold-based detection

$$\frac{1}{T}\sum_{k=1}^{T}\left[Y_{k+1} - aY_k - U_k(Y_1^k)\right]^2 \in (Var[W] - \delta, Var[W] + \delta).$$

- What kind of attacks can we detect?

The man in the middle attack types

Stuxnet

Replay attack

Y. Mo, B. Sinopoli (2009)

Statistical-duplicate attack

$$X_{k+1} = aX_k + U_k + W_k$$

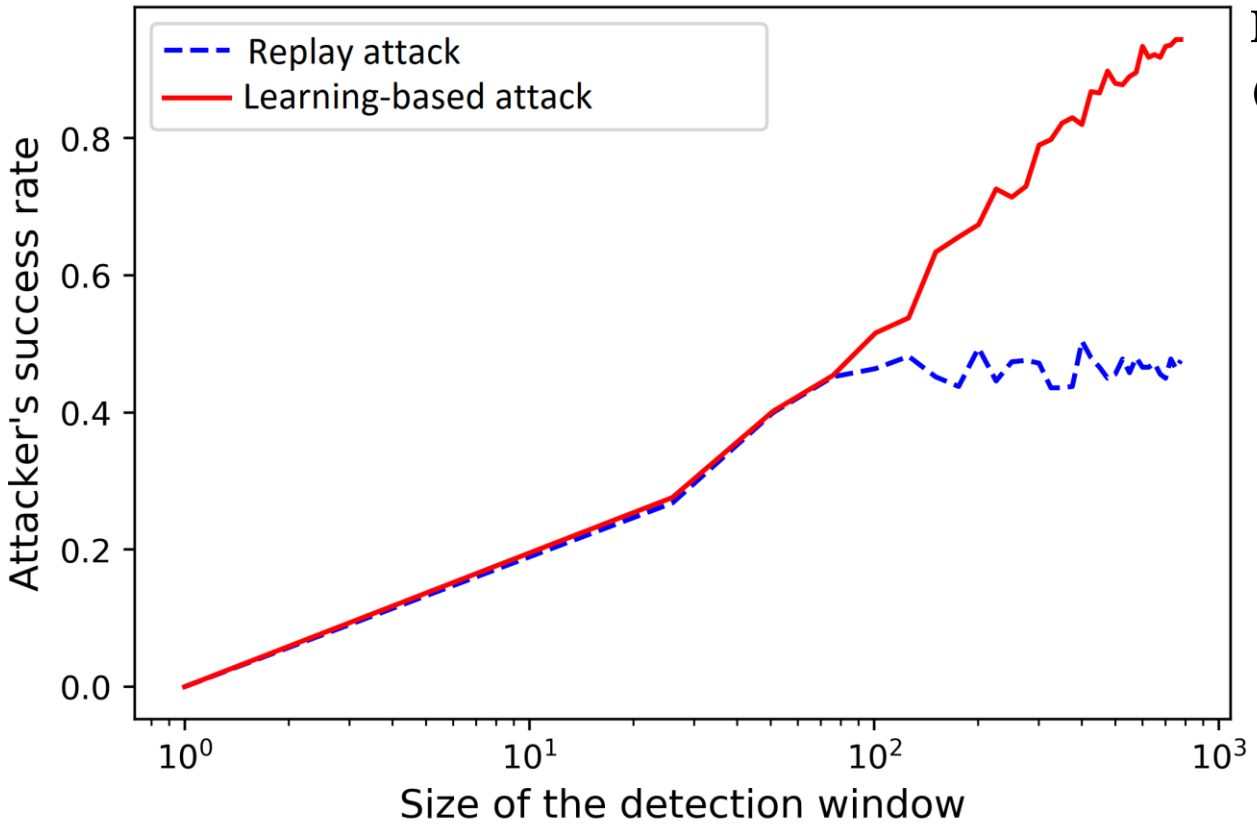
B. Satchidanandan,
P. R. Kumar (2017)
R. S. Smith (2011)

Learning-based attack

$$X_{k+1} = aX_k + U_k + W_k$$

MJ Khojasteh et al. (2020)

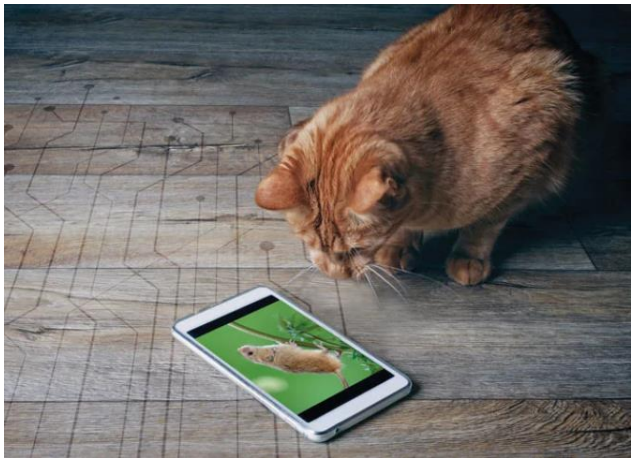
Comparison with a replay attack



MJ Khojasteh et al. (2020)

Stuxnet

Defense against learning-based attack

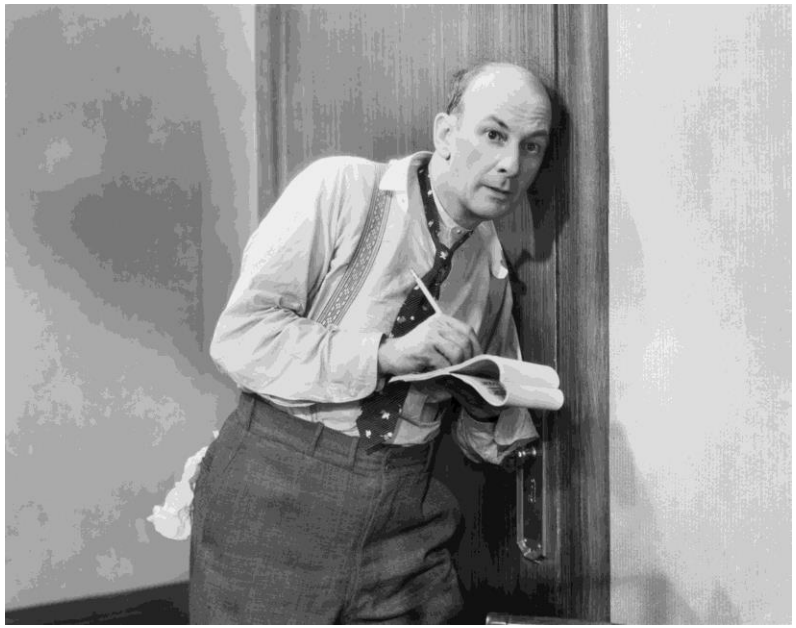


$$X_{k+1} = aX_k + U_k + W_k.$$

- The attacker has access to both X_k and U_k and knows the distribution of W_k and of the initial condition X_0, but it should learn the open loop gain a of the plant.

Two phases of the learning-based attack

Learning (exploration) phase



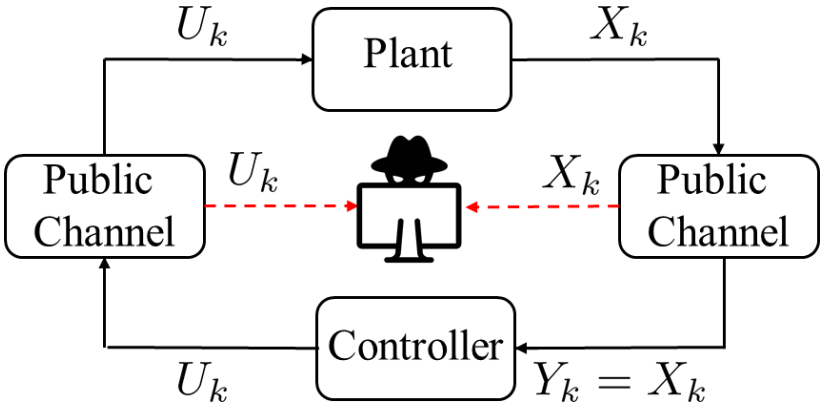
Eavesdropping and learning

Hijacking (exploitation) phase

Hijacking the system

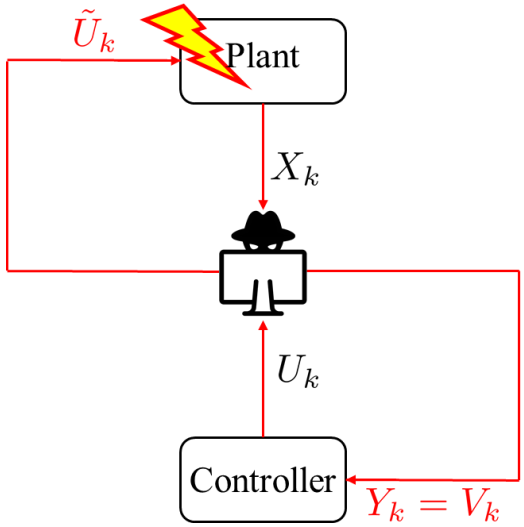
Two phases of the learning-based attack

Learning (exploration) phase



Eavesdropping and learning

Hijacking (exploitation) phase



Hijacking the system

Privacy-enhancing signal

Impede the learning process of the attacker

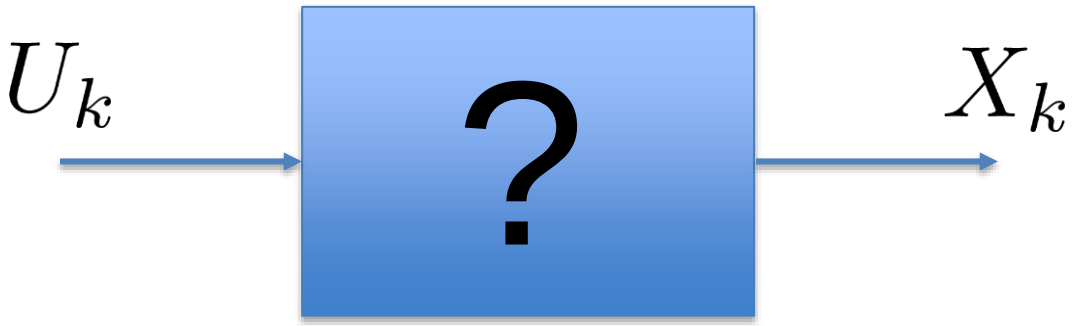
Privacy-enhancing signal

$$U_k = \bar{U}_k + \Gamma_k$$

Nominal control policy

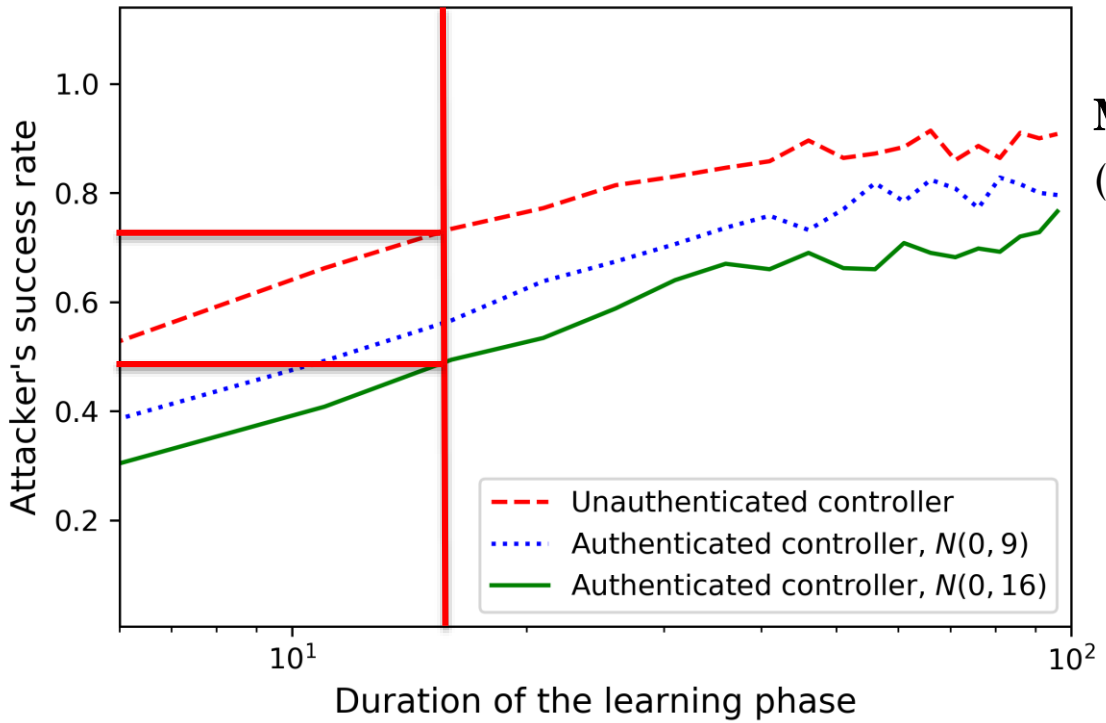
Privacy-enhancing signal

- Injecting a strong noise may in fact speed up the learning process



- Carefully crafted privacy-enhancing signals provide better guarantees on the deception probability

Privacy-enhancing signal

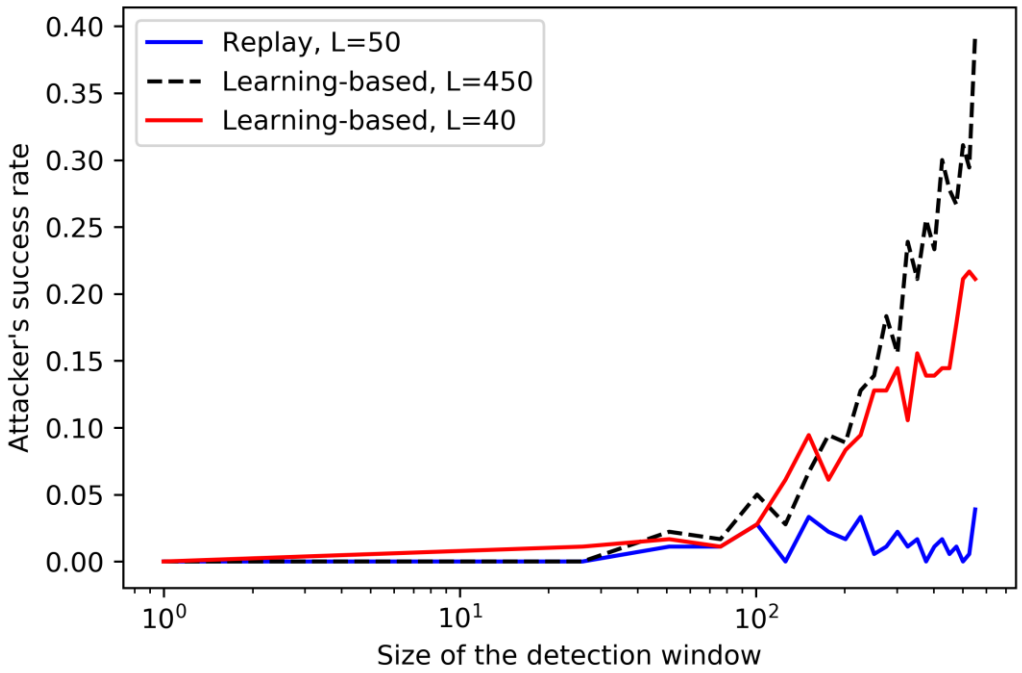


MJ Khojasteh et al. (2020)

Learning-based attack: vector systems

$$A \longrightarrow \mathbf{A} = \begin{bmatrix} ? \end{bmatrix}$$

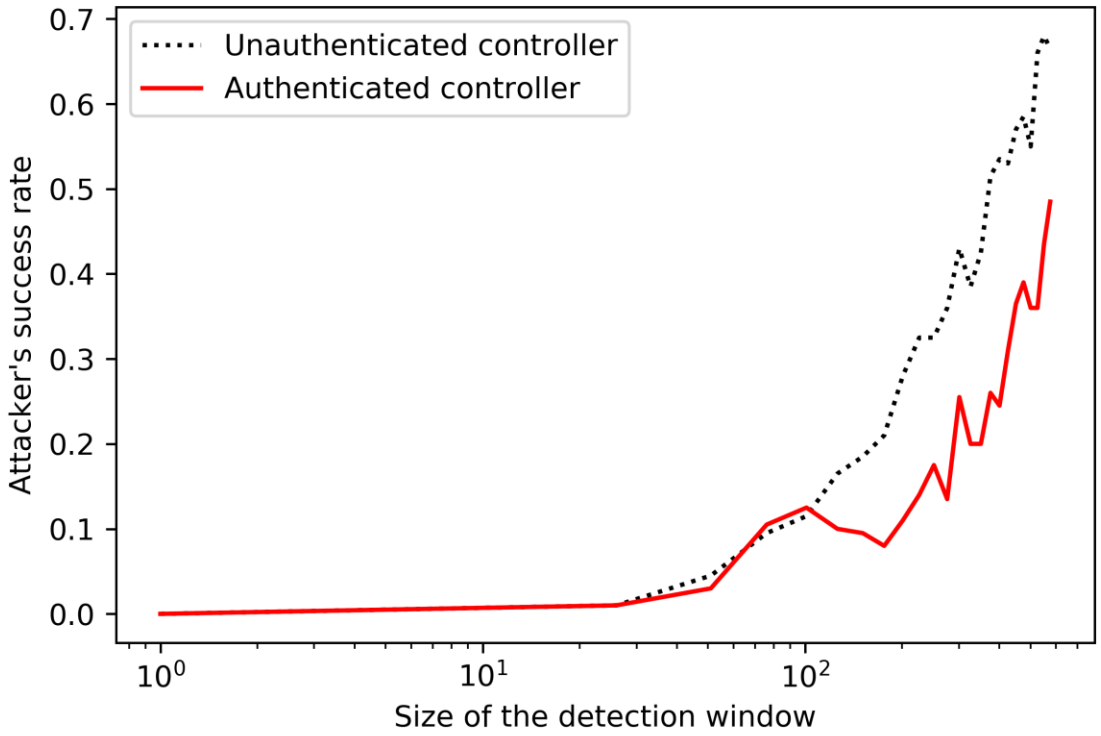
$$\mathrm{Var} \longrightarrow \mathrm{Cov}$$



MJ Khojasteh et al. (2020)

Stuxnet

Defense against vector learning-based attack

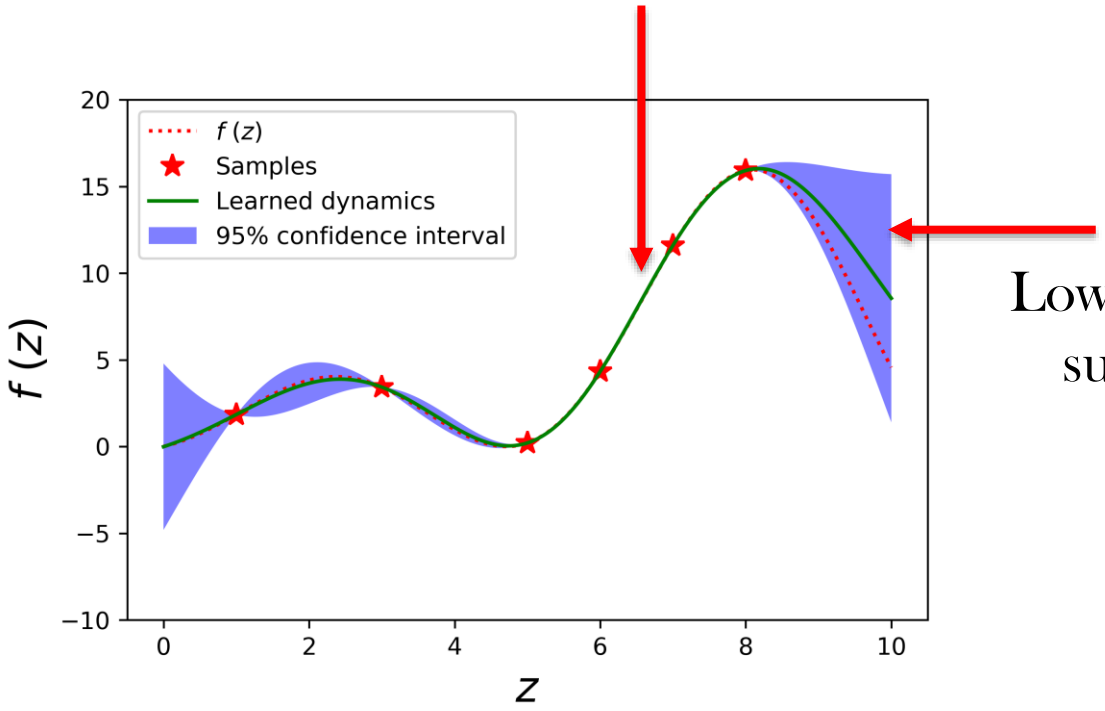


Nonlinear learning-based attack

$A \rightarrow f(X,U) \in$ Reproducing Kernel Hilbert Space (RKHS)

Linear regression → Bayesian learning: Gaussian processes (GP)

Vulnerable region



Lower attacker's success rate

References

- Khojasteh MJ, Khina A, Franceschetti M, Javidi T.
  Authentication of cyber-physical systems under learning-based attacks.
  IFAC-PapersOnLine. 2019 Jan 1; 52(20): 369-74.

- Khojasteh, M.J., Khina, A., Franceschetti, M. and Javidi, T.
  Learning-based attacks in cyber-physical systems.
  arXiv preprint arXiv:1809.06023, 2020.