Resiliency of Nonlinear Control Systems Against Stealthy Attacks

Amir Khazraei Miroslav Pajic

CPSL@Duke

Department of Electrical and Computer Engineering

Pratt School of Engineering

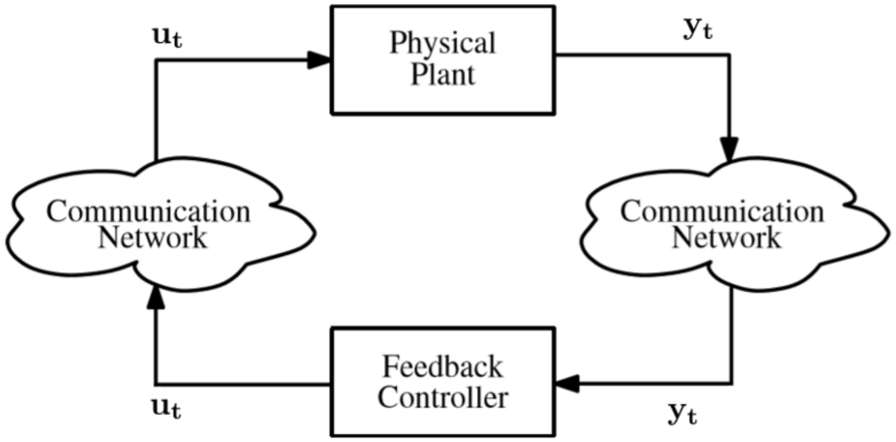
Duke University

Duke

PRATT SCHOOL of ENGINEERING

Motivation

How to design resilient control systems?



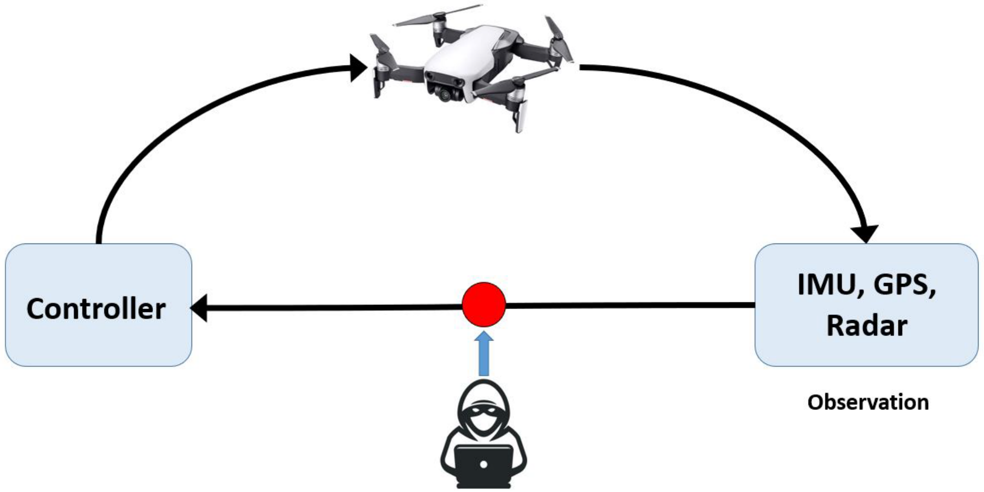
- Authenticating all the transmitted data
- Intermittently authenticating the transmitted data

We need to know which class of systems are vulnerable to cyber attacks

Introduction

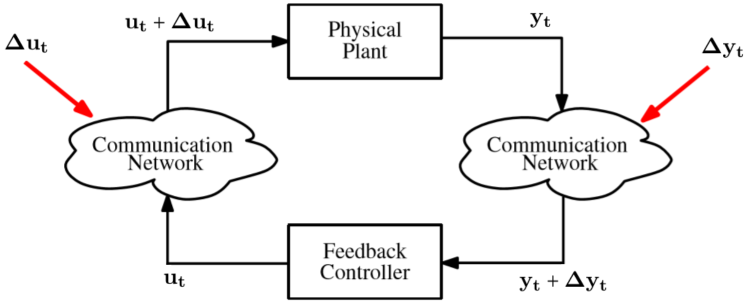
What is the impact of stealthy adversarial attacks on nonlinear control systems?

- We model the sensor attacks on nonlinear control systems and formalize the attacker's goal
- We consider the notion of stealthiness independent of any existing intrusion detector
- We derive the condition for the existence of impactful yet stealthy attacks

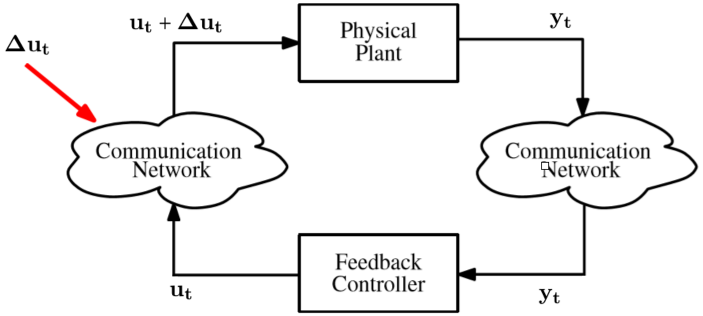


Related Works

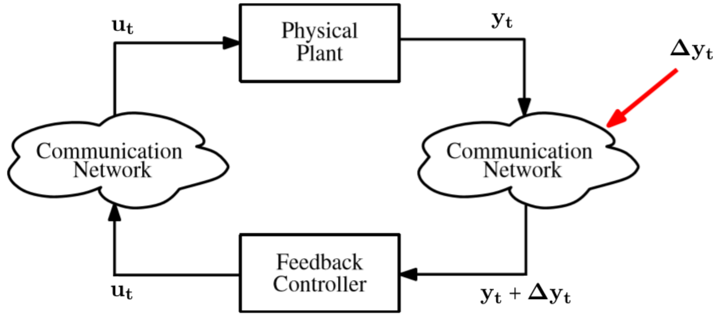
Control Theoretic VA of CPS



- Replay Attack [1]
- Covert Attack [2]



Zero Dynamic Attack [3]



False Data Injection Attack [4]

- **The previous works only consider systems with LTI model**
- **The notion of stealthiness is only limited to a specific class of intrusion detectors**

[1] Y. Mo and B. Sinopoli, “Secure control against replay attacks”. In 2009 47th Annual Allerton Conference on Communication, Control, and Computing.
[2] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure." IEEE Control Systems Magazine
[3] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson. Revealing stealthy attacks in control systems. In 2012 50th Annual Allerton Conference on Communication, Control, and Computing.
[4] Y. Mo and B. Sinopoli, “*False data injection attacks in control systems*,” in First Workshop on Secure Control Systems, 2010

System Architecture

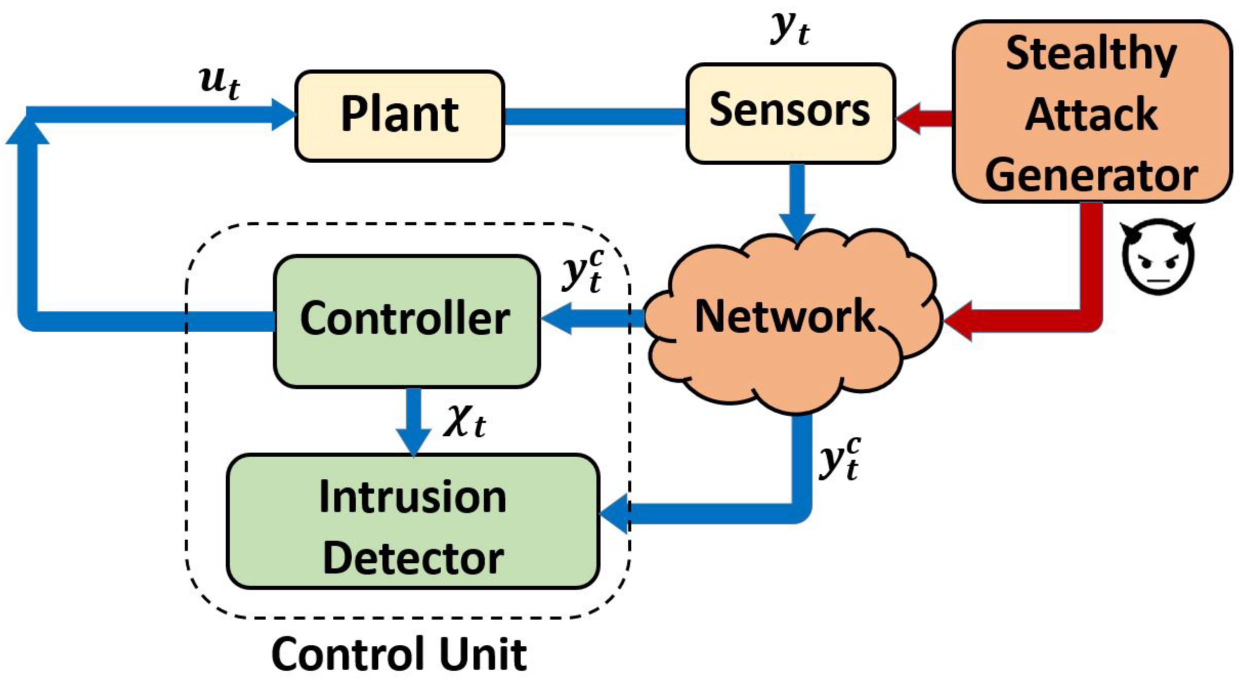
Duke PRATT SCHOOL of ENGINEERING

Plant: $x_{t+1} = f(x_t, u_t) + w_t$

Sensor model: $y_t = h(x_t) + v_t$

Controller: $\chi_t = f_c(\chi_{t-1}, y_t^c)$

$u_t = h_c(\chi_t, y_t^c)$



$$\boldsymbol{X} = \begin{bmatrix} x \\ \chi \end{bmatrix}, \ \boldsymbol{W} = \begin{bmatrix} w \\ v \end{bmatrix}, \ \boldsymbol{U} = \begin{bmatrix} u \\ w \end{bmatrix}$$

$\boldsymbol{X}_{t+1} = F(\boldsymbol{X}_t, \boldsymbol{W}_t)$ Closed-loop dynamic

$x_{t+1} = f_u(x_t, \boldsymbol{U}_t)$ Open-loop dynamic

Intrusion Detector

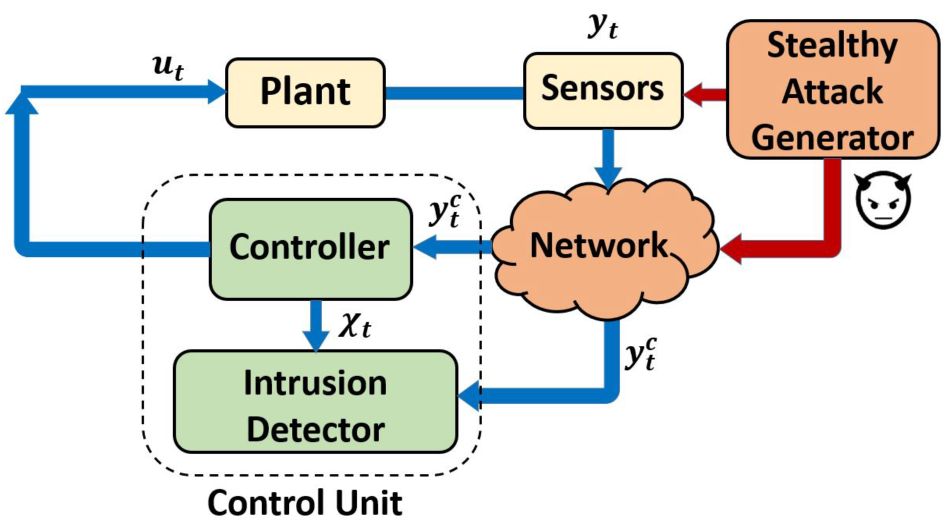
- H_0: Normal condition (the ID receives $\mathrm{Y} = Y_{-\infty}{:}\,Y_t$ with distribution $\boldsymbol{P}$) $\qquad Y_t = \begin{bmatrix} y_t^c \\ \chi_t \end{bmatrix}$

- H_1: Abnormal behavior (the ID receives $Y^a = Y_{-\infty}^{-1}, Y_0^a{:}\,Y_t^a$ with distribution $\boldsymbol{Q}$) $\qquad Y_t^a = \begin{bmatrix} y_t^{c,a} \\ \chi_t^a \end{bmatrix}$

Intrusion Detector: $\mathcal{D}(\bar{Y}) \to \{0,1\}$

$$p^{FA} = \mathbb{P}(\mathcal{D}(\bar{Y}) = 1 | \bar{Y} \sim \boldsymbol{P})$$

$$p^{TD} = \mathbb{P}(\mathcal{D}(\bar{Y}) = 1 | \bar{Y} \sim \boldsymbol{Q})$$



It is desired for the system: $p^{FA} < p^{TD}$

Incremental Stability

$$x_{t+1} = f(x_t, d_t) \qquad x_t \in \mathbb{X} \subseteq \mathbb{R}^n,\ d_t \in \mathbb{D} \subseteq \mathbb{R}^m,\quad t \geq 0$$

Definition 1: The system is incrementally exponentially stable (IES) in the set $\mathbb{X} \subseteq \mathbb{R}^n$ if there exist $\kappa > 1$ and $\lambda > 1$ such that

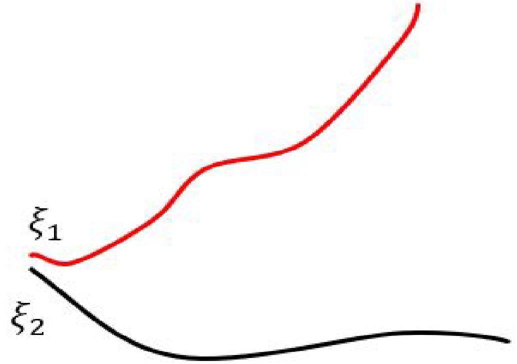
$$\|x(t, \xi_1, d) - x(t, \xi_2, d)\| \leq \kappa \|\xi_1 - \xi_2\| \lambda^{-t}$$

holds for all $\xi_1,\ \xi_2 \in \mathbb{X}$ and $d_t \in \mathbb{D}$ and $t \geq 0$. When $\mathbb{X} = \mathbb{R}^n$, the system is referred to as globally incrementally exponentially stable.

Definition 2: The system is incrementally unstable (IU) in the set $\mathbb{X} \subseteq \mathbb{R}^n$ if for all $\xi_1 \in \mathbb{X}$ and any $d_t \in \mathbb{D}$ there exists a ξ_2 such that for any $M > 0$

$$\|x(t, \xi_1, d) - x(t, \xi_2, d)\| \geq M$$

hold for all $t \geq t'$ for some $t' \geq 0$.



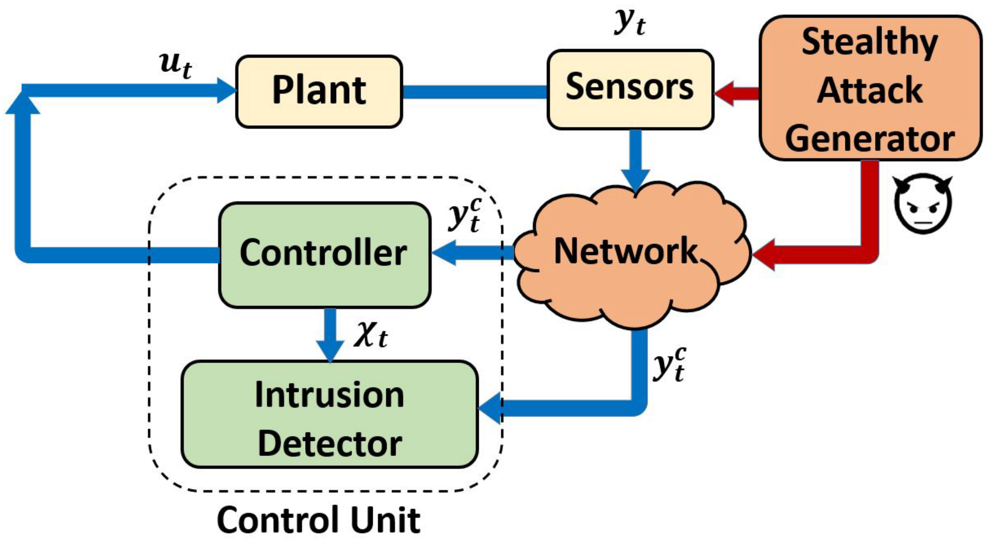
Attack Model

- The attacker has full knowledge of the system, its dynamics and the states x_t^a and the input u_t^a.

- The attacker has the required computation power to calculate suitable attack signals to inject a subset of sensors, while planning ahead as needed

- The attacker can also compromise the sensor measurements

$$y_t^{c,a} = y_t^a + a_t$$



Stealthiness Definition

Definition 3: An attack sequence

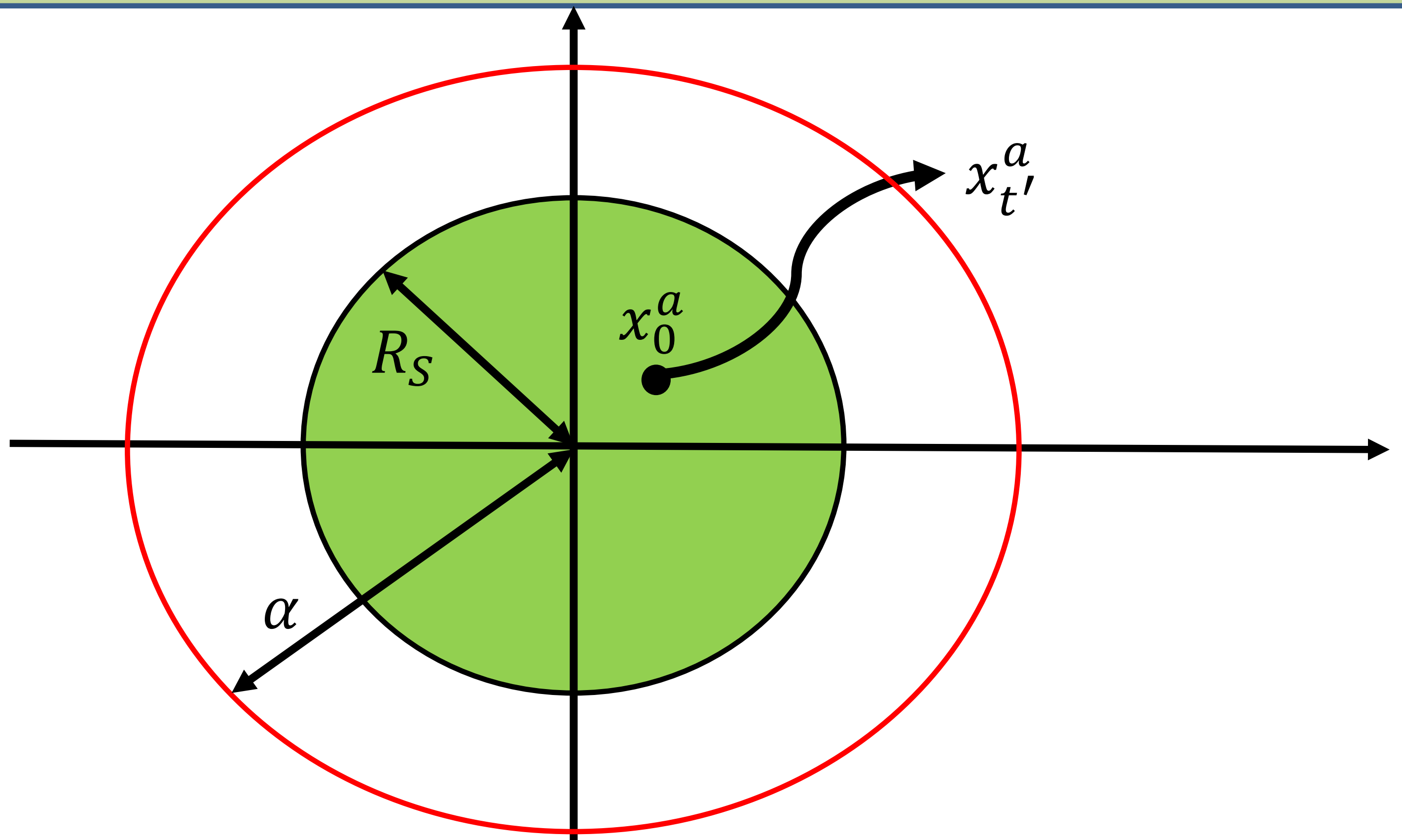
- is **strictly stealthy** if there exists no detector that satisfies $\boldsymbol{p_t^{FA}} < \boldsymbol{p_t^{TD}}$, for any $t \geq 0$,
- is **$\boldsymbol{\epsilon}$-stealthy** if for a given $\epsilon > 0$, there exists no detector such that $\boldsymbol{p_t^{FA}} < \boldsymbol{p_t^{TD}} - \epsilon$ for any $t \geq 0$.

Theorem 1: An attack sequence

- is **strictly stealthy** if and only if $\boldsymbol{KL}\big(\boldsymbol{Q}(Y_{-\infty}^{-1}, Y_0^a \colon Y_t^a) || \boldsymbol{P}(Y_{-\infty} \colon Y_t)\big) = 0$ for any $t \geq 0$,
- is **$\boldsymbol{\epsilon}$-stealthy** if it satisfies $\boldsymbol{KL}\big(\boldsymbol{Q}(Y_{-\infty}^{-1}, Y_0^a \colon Y_t^a) || \boldsymbol{P}(Y_{-\infty} \colon Y_t)\big) \leq \log(\frac{1}{1-\epsilon^2})$ for any $t \geq 0$.

Formalizing the Attacker's Goal

Definition 4: Attack sequence $\{a_0, a_1, \dots\}$ is an (ϵ, α)-successful attack if there exists $t' \geq 0$ such that $\|x_{t'}^a\| \geq \alpha$ for all $t \geq t'$ and the attack is ϵ-stealthy for all $t \geq 0$. When such a sequence exists for a system, the system is called (ϵ, α)-attackable.



Vulnerability Analysis of Nonlinear Control Systems

Theorem 2: The system is (ϵ, α)-attackable for arbitrarily large α and arbitrarily small ϵ, if the closed-loop dynamics is incrementally exponentially stable (IES) in the set S and the open loop dynamics is incrementally unstable in the set S.

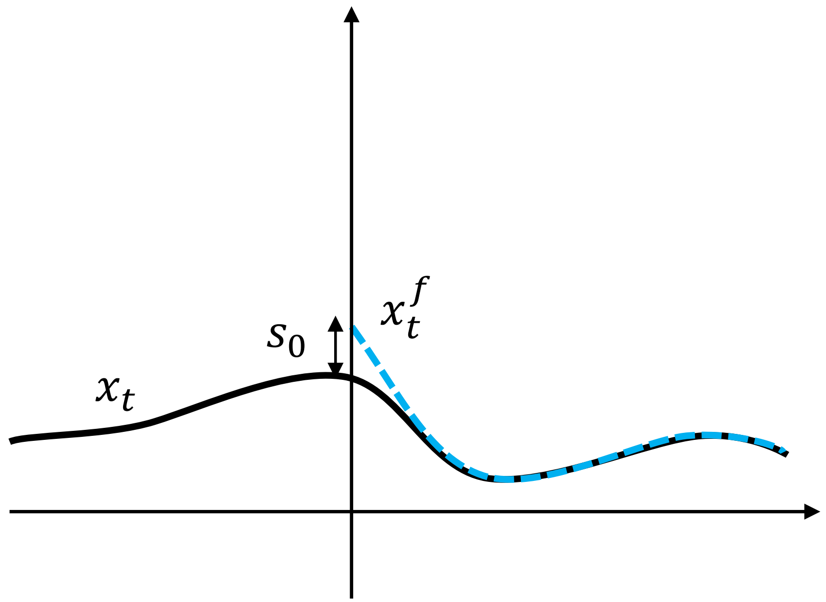
$$\begin{cases} s_{t+1} = f(x_t^a, u_t^a) - f(x_t^a - s_t, u_t^a) \\ a_t = h(x_t^a - s_t) - h(x_t^a) \end{cases}$$

$$y_t^{c,a} = y_t^a + a_t = h(x_t^a - s_t) + v_t$$

$$x_t^f \triangleq x_t^a - s_t \qquad \boldsymbol{X}^f = \begin{bmatrix} x^f \\ \chi^a \end{bmatrix}$$

incrementally exponentially stability of closed-loop system

$$\left\|\boldsymbol{X}_t^f - \boldsymbol{X}_t\right\| \leq \kappa \|s_0\| \lambda^{-t}$$



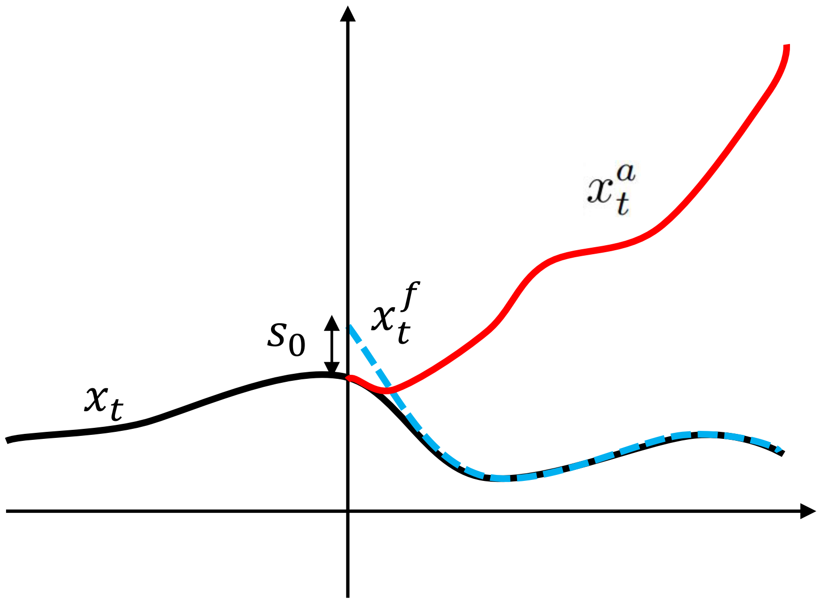
Vulnerability Analysis of Nonlinear Control Systems

incremental exponentially stability of closed-loop system $\Rightarrow$ $\|\boldsymbol{X}_t^f - \boldsymbol{X}_t\| \leq \kappa \|s_0\| \lambda^{-t}$

$$\boldsymbol{KL}\big(\boldsymbol{Q}(Y_0^a\!:Y_t^a)||\boldsymbol{P}(Y_0\!:Y_t)\big) \leq \frac{\kappa\|s_0\|}{1-\lambda^2}\left(\lambda_{max}(\Sigma_w^{-1}) + L_h \lambda_{max}(\Sigma_v^{-1})\right)$$

incremental instability of open-loop system $\Rightarrow$ $\|x_t^a - x_t^f\| \geq M$

$$\|x_t^a\| \geq \underbrace{M - \kappa\|s_0\| - R_S}_{\alpha}$$



Attack on LTI Systems

Duke
PRATT SCHOOL *of* ENGINEERING

$$x_{t+1} = Ax_t + Bu_t + w_t, \qquad y_t = Cx_t + v_t$$

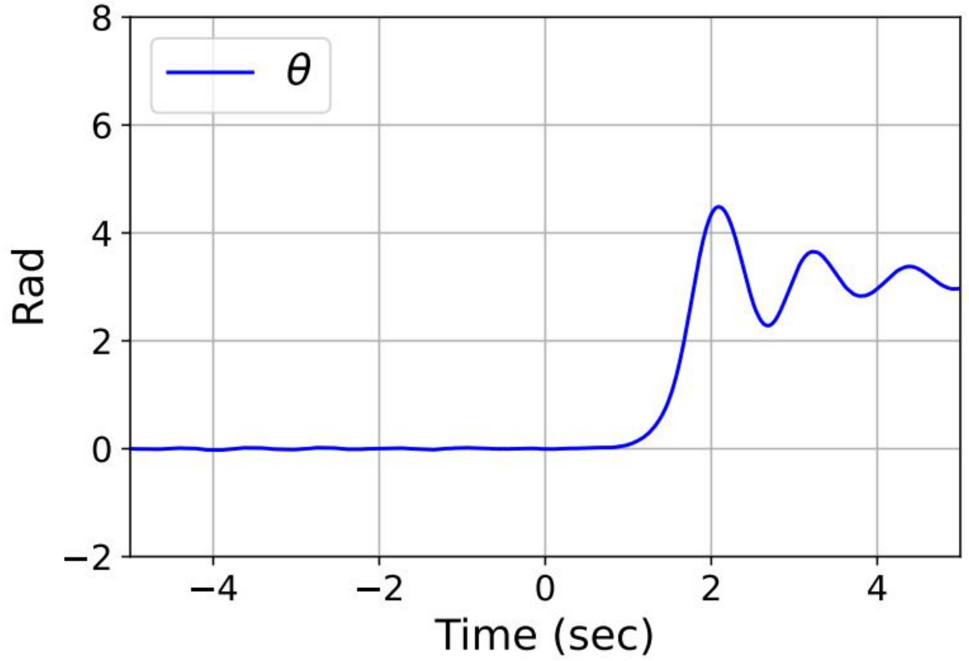
$$\chi_t = A_c\chi_{t-1} + B_c y_t^c, \qquad u_t = C_c\chi_t$$

Corollary 1: The LTI system is (ϵ, α)-attackable for arbitrarily large α and arbitrarily small ϵ, if the closed-loop dynamics is asymptotically stable and the matrix A is unstable.

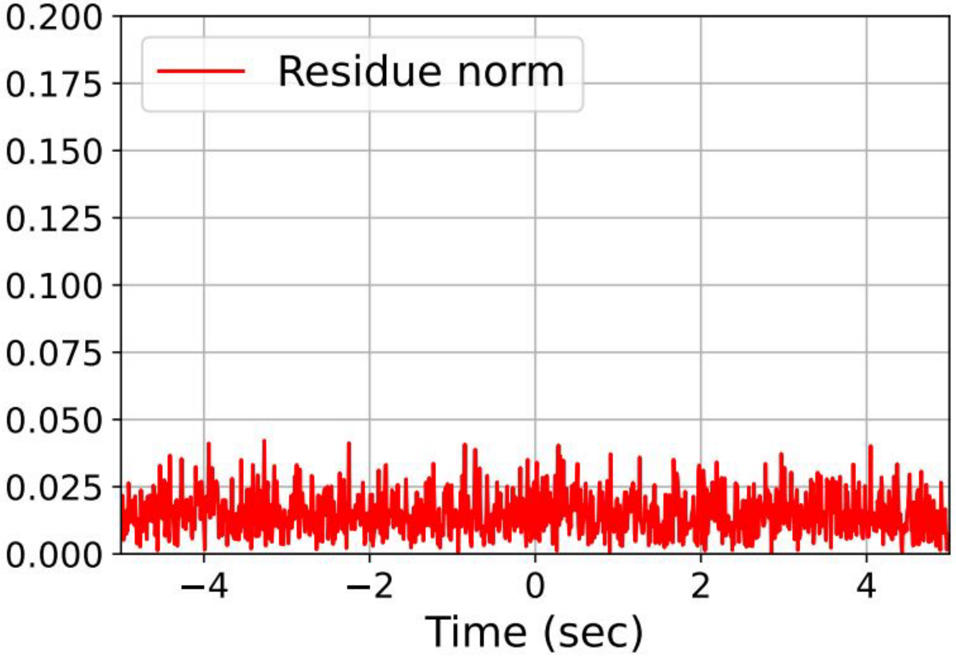
$$s_{t+1} = f(x_t^a, u_t^a) - f(x_t^a - s_t, u_t^a) = Ax_t^a + Bu_t^a - A(x_t^a - s_t) - Bu_t^a = As_t$$

$$a_t = h(x_t^a - s_t) - h(x_t^a) = C(x_t^a - s_t) - C(x_t^a) = -Cs_t$$

Case Study : Inverted Pendulum

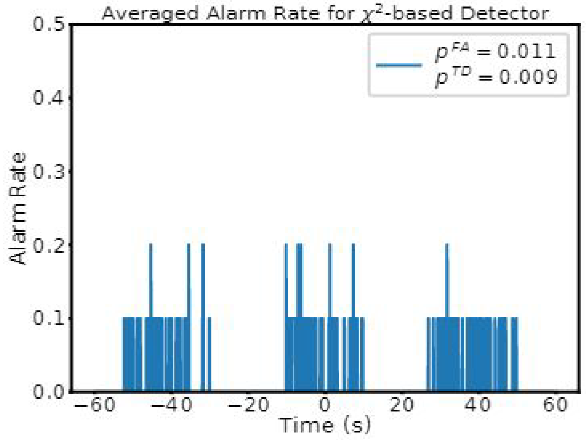


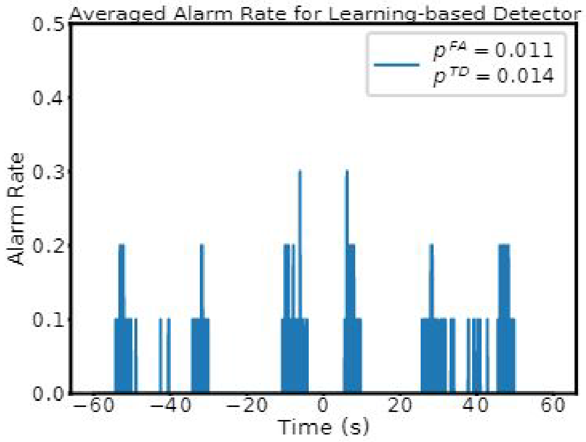
The angle of the pendulum of pod



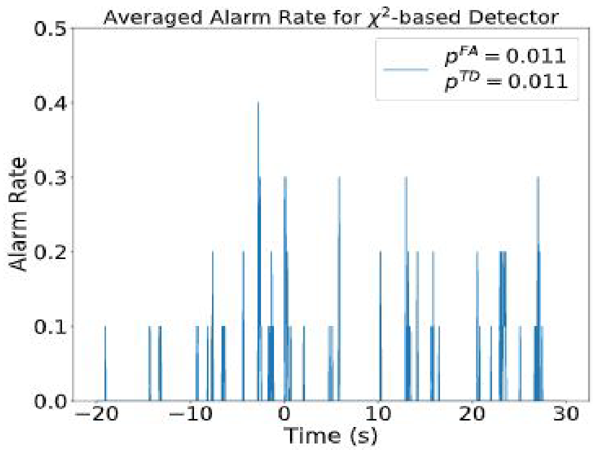
The norm of the residue

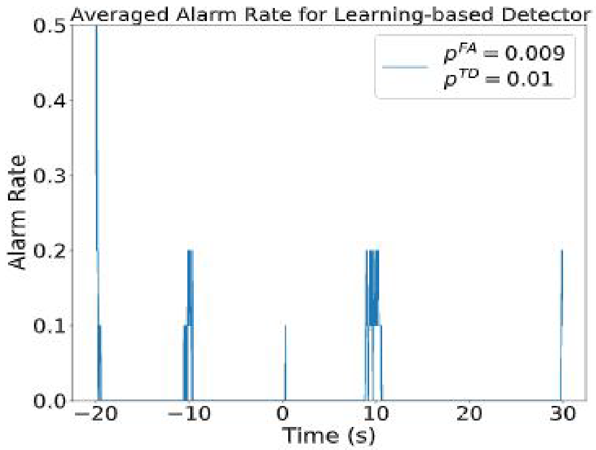
Case Study: UAV Quadrotor – Following a moving vehicle





Case Study: UAV Quadrotor – Landing on a moving vehicle





Conclusion

- We considered the resiliency under sensing attacks for nonlinear control systems.
- We introduced the notion of ϵ -stealthiness in a general form.
- We derived sufficient conditions for an effective yet ϵ-stealthy attack sequence to exists.

Thank you