CERIAS

The Center for Education and Research in Information Assurance and Security

Model Order Reduction of Cyber-Physical Systems Considering Stealthy Attacks

Minhyun Cho, Suriyan Anandavel, Soughwan Hwang, and Inseok Hwang

(mhcho, sanandav, hwang214, ihwang@purdue.edu)

Motivation

- **Model Order Reduction**
  - Model order reduction (MOR) techniques have been used to reduce the complexity of mathematical models, especially in systems governed by differential equations
  - Used in control theory, system dynamics, simulation and machine learning
  - Different techniques have been developed: balanced truncation[1], (Time-domain) Moment matching methods[2], Krylov subspace methods[3], Hankel norm minimization methods[4]

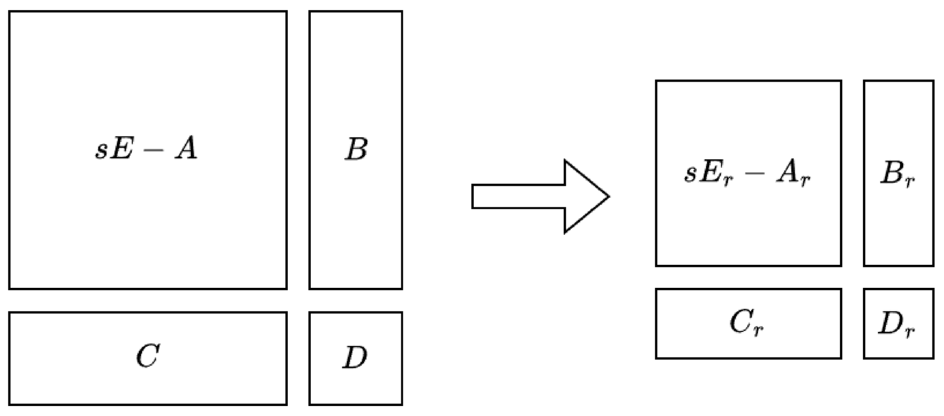

Figure 1. Reduction of a state-space model

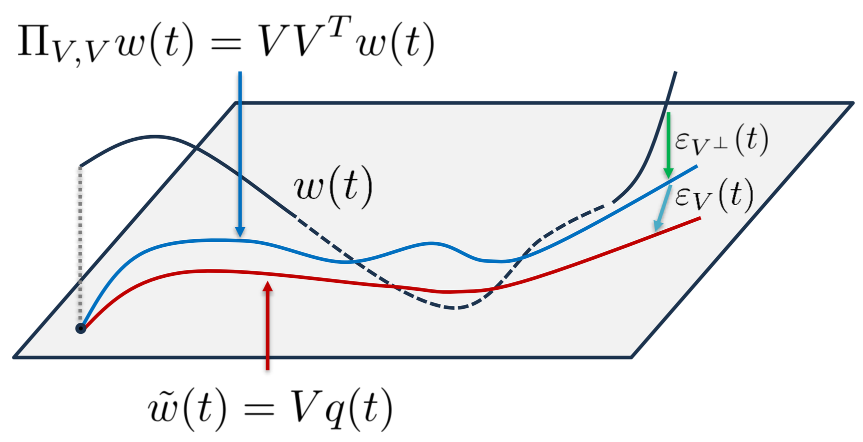

Figure 2. MOR interpreted as projection operator

- **Limitations of Previous Studies**
  - MOR methods primarily focused on preserving controllability, observability, and in some cases, stability, which are essential to for controller design, simulation and verification
  - No MOR method has been developed to maintain the vulnerability, which is needed to synthesize cyberattack models and analyze cyber-physical vulnerabilities (CPVs)
- **Objectives**
  - Develop a MOR technique that can preserve CPVs to enable vulnerability analysis and corresponding exploits of cyber-physical systems (CPSs)
  - Enhance the effectiveness and scalability of CPV analysis on large-scale systems

Problem Formulation

- **Problem Statement**
  - Investigate stealthy actuator and sensor attacks by leveraging nonminimum-phase zeros and unstable poles, and quantify the dimension of the vulnerable subspace of an original CPS
  - Use the extended pole-zero technique, i.e., Routh's criterion, and optimization-based MOR to preserve unstable zeros and poles, ensuring the dimension of the vulnerable subspace is retained as possible in the reduced order model (ROM)
- **Problem Formulation**
  - Consider SISO LTI system that is vulnerable to zero-stealthy attacks

$$P(s) = k\frac{s^{n-r} + c_{n-r-1}s^{n-r-1} + \cdots + c_1 s + c_0}{s^n + d_{n-1}s^{n-1} + \cdots + d_1 s + d_0}$$

  - Minimal realization of the system and a dynamic output feedback controller in Byrnes-Isidori normal form

$$\mathcal{P}: \quad \dot{x} = \underbrace{\begin{bmatrix} (\boldsymbol{\Omega}_r + \mathbf{e}_r\phi_1^T) & \mathbf{e}_r\phi_2^T \\ \phi_3\mathbf{e}_1^T & \Phi \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \underbrace{\begin{bmatrix} b\mathbf{e}_r \\ \mathbf{0}_{n-r} \end{bmatrix}}_{B} u, \quad y = \underbrace{\begin{bmatrix} \mathbf{e}_1^T & \mathbf{0}_{n-r}^T \end{bmatrix}}_{C^T} x$$

$$\mathcal{C}: \quad \dot{x}_c = \underbrace{\begin{bmatrix} (\boldsymbol{\Omega}_q + \mathbf{e}_q\psi_1^T) & \mathbf{e}_q\psi_2^T \\ \psi_3\mathbf{e}_1^T & \Psi \end{bmatrix}}_{A_c} \begin{bmatrix} x_{c1} \\ x_{c2} \end{bmatrix} + \underbrace{\begin{bmatrix} f\mathbf{e}_q \\ \mathbf{0}_{m-q} \end{bmatrix}}_{B_c} y, \quad u = \underbrace{\begin{bmatrix} \mathbf{e}_1^T & \mathbf{0}_{m-q}^T \end{bmatrix}}_{C_c^T} x_c + D_c y$$

  - Attack signal injected as $u' \leftarrow u + a_a, \quad y' \leftarrow y + a_s$
- **Preliminaries**
  - **Definition 1 (Stealthy Attack):** The non-zero attack signal $a(t) = [a_a(t),\, a_s(t)]$ is said to be ε -stealthy for the output if $||y(t) - y_{af}(t)|| = ||y_a(t)|| \leq \varepsilon, \forall t \geq t_0$ is satisfied. In particular, the attack is called zero-stealthy, or undetectable when the threshold becomes $\varepsilon = 0$.
  - **Definition 2 (Vulnerability to Stealthy Attack):** The system induced by the attack is said to be susceptible to a given ε -stealthy attack satisfying $||y_a(t)|| \leq \epsilon,\ t \in [t_0, t^*]$ if the attack causes an impact, $||x_a(t^*)|| \geq \rho,\ \exists t^* \geq t_0$ for a given threshold ρ and some time t^*, given the initial condition $x_a(t_0) = 0$ and the attack $a(t)$ for $t \in [t_0, t^*]$

References

1. M. G. Safonov and R. Y. Chiang, "A Schur Method for Balanced Model Reduction," in 1988 American Control Conference, USA: IEEE, Jun. 1988, pp. 1036–1040.
2. A. Padoan, "On model reduction by least squares moment matching," in 2021 60th IEEE Conference on Decision and Control (CDC), Dec. 2021, pp. 6901–6907.
3. H. K. F. Panzer, T. Wolf, and B. Lohmann, "H2 and H∞ error bounds for model order reduction of second order systems by Krylov subspace methods," in 2013 European Control Conference (ECC), Jul. 2013, pp. 4484–4489.
4. K. Glover, "A Tutorial on Hankel Norm Approximations," 1989.

Main Results

- **Characterization of Stealthy Attacks**

Proposition 1 (Zero-dynamics Attack): Suppose that an actuator attack

$$\dot{\eta}_a = \Phi\eta_a, \qquad a_a = -\frac{1}{b}\phi_2^T\eta_a$$

with an initial condition $\eta_a(t_0) = \eta_{a0}$ is injected from the time $t_0 > 0$ into the closed-loop. If η_{a0} is sufficiently small, then the attack becomes stealthy. Moreover, if Φ has at least one eigenvalue whose real part is positive (i.e., the plant has an unstable zero, or, is non-minimum phase) and the initial condition η_{a0} excites the unstable mode, then the attack is disruptive.

Proposition 2 (Pole-dynamics Attack): Suppose that a sensor attack

$$\dot{\eta}_s = (A + BD_cC)\eta_s, \qquad a_s = -C\eta_s,$$

with an initial condition $\eta_s(t_0) = \eta_{s0}$, where t_0 denotes the time when the attack is injected into the communication channel between the system output and the controller input. The generated attack becomes ε-stealthy if the initial condition η_{s0} is set sufficiently small. Furthermore, if the matrix $A + BD_cC$ has at least one eigenvalue with a positive real part, i.e., the system has unstable poles, or the system is not output feedback stabilizable or the controller has zero feedforwards by using an observer-based controller. Then, we can claim that the system is regarded to be vulnerable to the given stealthy attack.

- **Extended pole-zero method (MOR Technique):**
  - Minimizing the L^2-norm of the error signal between the original model and a reduced model

$$J = \int_0^\infty ||e(t)||_2^2 = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} E(s)E(-s)ds,$$

$$E(s) = P(s) - \hat{k}\frac{\hat{U}(s)}{\hat{V}(s)} = P(s) - \hat{k}\frac{\hat{U}^s(s)\hat{U}^u(s)}{\hat{V}^s(s)\hat{V}^u(s)},$$

$$\hat{P}_\ell(s) = \hat{k}\frac{(s^{\ell-1} + \hat{c}_{\ell-2}s^{\ell-2} + \cdots + \hat{c}_0)}{s^\ell + \hat{d}_{\ell-1}s^{\ell-1} + \cdots + \hat{d}_0} \triangleq \hat{k}\frac{\hat{U}(s)}{\hat{V}(s)},$$

Illustrative Examples

- **Results**
  - Our MOR method preserves the vulnerability of two CPSs—the linearized elevator-pitch dynamics of an LTV A-7A Corsair II aircraft with second-order actuator dynamics and the linearized pitch dynamics of a quadrotor—while performing the reduction. The reduction preserves the dimension of the vulnerable subspace in both CPSs to the extent allowed by the reduction order.
  - The proposed method is expected to reduce the computational complexity involved in the computation of simulation, and vulnerability exploits.

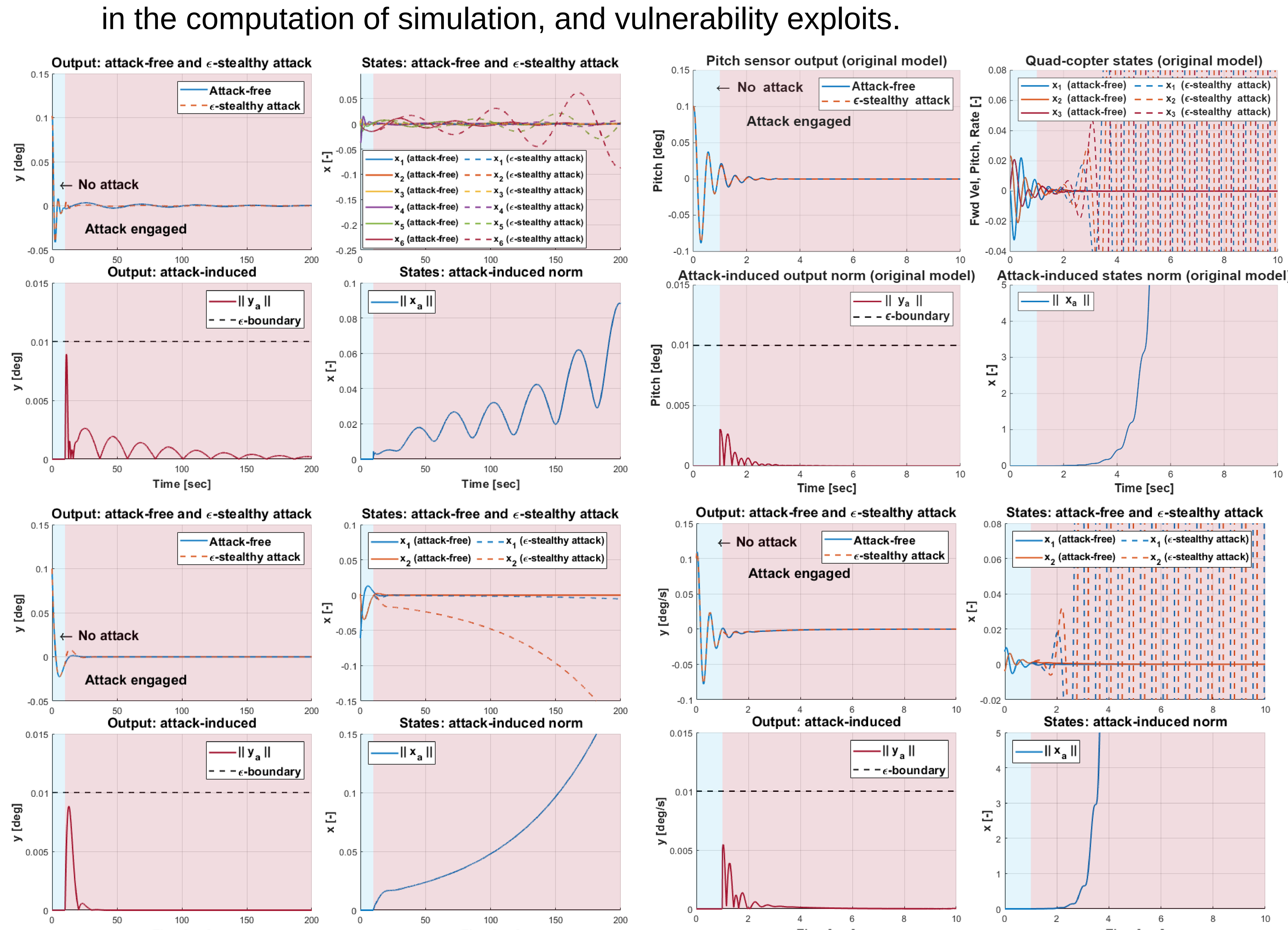

Fig 3. **Top Left:** Stealthy zero-dynamics attack on the original 6th-order A-7A Corsair II system. **Top Right:** Stealthy pole-dynamics attack on the original 3rd-order quadrotor system. **Bottom Left:** Stealthy zero-dynamics attack on the reduced 2nd-order A-7A Corsair II system. **Bottom Right:** Stealthy pole-dynamics attack on the reduced 2nd-order quadrotor system. For each subfigure: outputs (top left), attack-induced output (bottom left), states (top right), and attack-induced state norm (bottom right).

PURDUE UNIVERSITY

CERIAS