Confidentiality of Linear Systems with Quadratic Output

Zeyad M. Manaa, Nathan van de Wouw, Michelle S. Chong

Dynamics and Control, Eindhoven University of Technology, The Netherlands. Emails: {z.manaa, n.v.d.wouw, m.s.t.chong}@tue.nl.

1 Introduction

Modern cyber-physical systems rely on networked sensors to monitor physical processes. An adversary with read and write access to sensor data can infer internal states and launch more dangerous, undetectable attacks that disrupt system performance while remaining undetected [1]. We adopt the adversary perspective, which is formulated as an estimation problem for a linear plant with a quadratic sensor output map, stabilized by a dynamic controller. We assume the closed-loop system has inherent stability properties. Yet the attack-free closed loop is unobservable at equilibrium due to the quadratic output map, which precludes observer designs based on local observability. We therefore ask whether an adversary can still estimate the plant and controller states. Indeed, this can be achieved by inducing observability by manipulating the sensor output while preserving closed-loop stability to avoid detection by existing abnormality detectors in the system.

2 Problem formulation

We consider a linear plant with scalar quadratic output $\Sigma_p:\ \dot{x}_p = A_p x_p + B_p u,\ y = x_p^\top Q_p x_p$, controlled by the dynamic output-feedback controller $\Sigma_c:\ \dot{x}_c = A_c x_c + B_c y,\ u = C_c x_c + D_c y$, where $x_p \in \mathbb{R}^{n_p}$, $x_c \in \mathbb{R}^{n_c}$, and $u, y \in \mathbb{R}$. Let $z := [x_p^\top\ x_c^\top]^\top \in \mathbb{R}^n$, the closed-loop dynamics are

$$\dot{z} = Az + Bh(z) =: f(z), \qquad h(z) = z^\top Q z, \tag{1}$$

with $Q = \begin{bmatrix} Q_p & 0 \\ 0 & 0 \end{bmatrix}$, $A = \begin{bmatrix} A_p & B_p C_c \\ 0 & A_c \end{bmatrix}$, $B = \begin{bmatrix} B_p D_c \\ B_c \end{bmatrix}$. We assume $z = 0$ is locally exponentially stable equilibrium, i.e., A is Hurwitz. The linearization of (1) is $F = \partial f/\partial z|_0 = A$ and $H = \partial h/\partial z|_0 = 0$, and is therefore unobservable at the origin. We assume the adversary knows (A, B, Q), can read and additively manipulate the sensor output such that $\tilde{y} = y + a$, with $a \in \mathbb{R}$, but has no access to the control input u. With the manipulated sensor measurements, the controller now sees $\tilde{y}$ instead of y and becomes $\dot{x}_c = A_c x_c + B_c \tilde{y},\ u = C_c x_c + D_c \tilde{y}$. In turn, the attacked version of (1) can be written as

$$\dot{z} = Az + B\big(h(z) + a\big) =: \bar{f}(z, a), \qquad h(z) = z^\top Q z. \tag{2}$$

Figure 1 illustrates the problem setup.

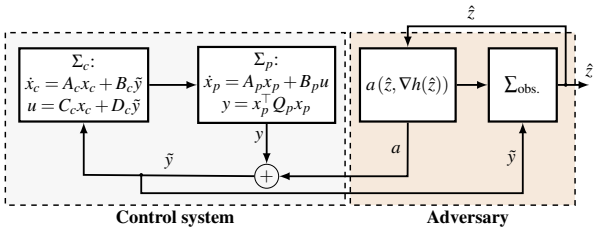

Figure 1: Problem setup. The block $\Sigma_{\text{obs.}}$ denotes the adversary's observer that will be designed.

In short, the adversary's objectives are:

Objective 1. *Design the injected signal a using the measured output $h(z)$ and the gradient $\nabla h(\hat{z}) = 2Q\hat{z}$ computed from the observer estimate $\hat{z}$ (from Objective 2), such that it renders system* (2) *locally observable at $z = 0$, i.e., the linearization of* (2) *is observable.*

Objective 2. *Design a state observer to get an estimate $\hat{z}$ of the state z of* (2)*, so that the origin is a locally exponentially stable equilibrium of the estimation error system $(z(t) - \hat{z}(t))$, with rate $\alpha > 0$.*

Objective 3. *Design the attack signal a such that the origin remains a locally exponentially stable equilibrium of the closed-loop system* (2)*.*

3 Approach

To achieve Objective 1, we design the attack as a projection of the estimated gradient using the observer state $\hat{z}$. Let $\pi \in \mathbb{R}^{n_p}$ be a projection direction and $\bar{\pi} := \big[\pi^\top,\ 0^\top\big]^\top \in \mathbb{R}^n$. Since $\nabla h(\hat{z}) = 2Q\hat{z}$, the attack is therefore $a(\hat{z}) := \bar{\pi}^\top \nabla h(\hat{z}) = \bar{H}(\pi)\hat{z}$ where $\bar{H}(\pi)[2\pi^\top Q_p \quad 0]$. Linearizing (2), at the origin yields the pair $(\bar{F}(\pi), \bar{H}(\pi))$ where $\bar{F}(\pi) = A + B\bar{H}(\pi)$. We design π such that $(\bar{F}(\pi), \bar{H}(\pi))$ is observable, thereby inducing local observability. We do so by applying the Hautus test to characterize the inadmissible set Π_{unobs} for which $(\bar{F}(\pi), \bar{H}(\pi))$ is unobservable, and then selecting $\pi \notin \Pi_{\text{unobs}}$.

Next, to achieve Objectives 2 and 3, we propose the observer

$$\Sigma_{\text{obs.}}: \dot{\hat{z}} = A\hat{z} + B(\hat{z}^\top Q\hat{z} + 2\bar{H}(\pi)\hat{z}) - L(\hat{z}^\top Q\hat{z} + 2\bar{H}(\pi)\hat{z} - \tilde{y}).$$

where L is the observer gain to be designed and π selected as discussed above. Recall that A is Hurwitz, and that the adversary's model as stated in section 2. In addition, assume $\sigma(A_p) \cap \sigma(A_c) = \emptyset$, and $B_p C_c \neq 0$ where $\sigma(\cdot)$ denotes the set of eigenvalues of a matrix. Under these assumptions, we design π and L such that $(\bar{F}(\pi), \bar{H}(\pi))$ is observable, and such parameters always exist under the recalled assumptions. The gain L then assigns the observer error eigenvalues to $\{s \in \mathbb{C} : \text{Re}(s) \leq -\alpha\}$, ensuring local exponential convergence with rate α.

4 Conclusion

We proposed a systematic sensor-attack design that induces local observability in closed-loop systems with linear dynamics and quadratic outputs while preserving stability. A Luenberger-type observer then guarantees local exponential convergence of plant and controller state estimates. Future work will address broader classes of nonlinearities.

References

[1] S. M. Dibaji, M. Pirani, D. B. Flamholz, A. M. Annaswamy, K. H. Johansson, and A. Chakrabortty, "A systems and control perspective of CPS security," *Annu. Rev. Control*, vol. 47, 2019.