Physical Watermarking and Authentication in Cyber Physical Systems

Yilin Mo (Caltech), Sean Weerakkody (CMU), and Bruno Sinopoli (CMU)

Department of Electrical and Computer Engineering
Carnegie Mellon University

TRUST Conference

October 10, 2013
Cyber-physical systems refer to the embedding of information, communication technology, and control into physical spaces with the goal of making them safer, more efficient and reliable. Such systems are becoming pervasive, thanks to the advances in sensing, computing and networking.
Cyber-Physical Systems

Smart Grid
Smart Vehicles
Smart Buildings
Smart Manufacturing
Secure Cyber-Physical Systems

- Cyber-physical systems (CPS) create new attack opportunities.
  - unsheltered systems
  - authenticity in information technology (IT) vs. CPS

- Attacks on CPS can have disastrous consequences.
Example: Stuxnet Attack

*Stuxnet Attack Strategy*

1. Infect centrifuges in enrichment plant.

2. Record dynamics of normal operation.

3. “Man in the middle attack,” replay previous dynamics.

4. Insert destabilizing input.
The System Model

Suppose we have system dynamics as follows:

- \( x_{k+1} = Ax_k + Bu_k + w_k \quad \text{for} \quad x_k \in \mathbb{R}^n, u_k \in \mathbb{R}^p, w_k \sim N(0, Q) \)
- \( y_k = Cx_k + v_k \quad \text{for} \quad y_k \in \mathbb{R}^m, v_k \sim N(0, R) \)

A Linear Quadratic Gaussian controller is implemented.

\[
J = \lim_{T \to \infty} \frac{1}{2T + 1} E \left[ \sum_{k=-T}^{T} x_k^T W x_k + u_k^T U u_k \right]
\]

\[
u_k = u_k^* = L \hat{x}_k, \quad L = (B^T S B + U)^{-1} B^T S A
\]

Kalman Filter

\[
\hat{x}_{k+1|k} = A \hat{x}_k + B u_k, \quad \hat{x}_k = \hat{x}_{k|k-1} + K z_k
\]

\[
z_k = y_k - C \hat{x}_{k|k-1}, \quad K = P C^T (C P C^T + R)^{-1}
\]
The Attack Model

\[ u^a_k \rightarrow \text{Plant} \rightarrow \text{Sensors} \]

\[ u_k^* \rightarrow z^{-1} \rightarrow \text{Estimator} \rightarrow \text{Virtual System} \]

\[ \hat{x}_k \rightarrow \text{LQG Controller} \]

\[ y_k^v \sim y_k \]

\[ u_k \rightarrow \text{Failure Detector} \rightarrow z_k \]
Proposed Approach: Watermarking

1) $y_k$ with optimal inputs $u_k^*$

2) Inject input $u_k = u_k^* + \zeta_k$

3) $y_k$ with sub-optimal input $u_k$

4) Binary Detector: $g(z_k) < \eta$
The Attack Model: Watermark

\[ u_k^a \rightarrow \text{Plant} \rightarrow \text{Sensors} \]

\[ u_k \rightarrow z^{-1} \rightarrow u_{k-1} \]

\[ u_k^* \rightarrow \text{LQG Controller} \rightarrow \hat{x}_k \rightarrow \text{Estimator} \rightarrow z_k \]

\[ \zeta_k \rightarrow \text{Failure Detector} \]

\[ y_k^v \sim y_k \]
Watermark Design

Watermark Design Properties, \( u_k = u_k^* + \zeta_k \)

- Assume \( \zeta_k \) is a zero-mean stationary Gaussian process with
\[
\Gamma(d) \triangleq \text{cov}(\zeta_k \zeta_{k+d}^T) = \mathbb{E}[\zeta_k \zeta_{k+d}^T].
\]

- Attacker knows \( \Gamma(d) \)!

Tradeoff: Cost versus Detection Ability

- Larger \( \Gamma(d) \) increases the probability of detection, while also increasing the cost of control
- Small \( \Gamma(d) \) reduces the probability of detection while reducing the cost of control
- \( J = J^* + \Delta J \), \( \Delta J \) is linear in the auto-covariance \( \Gamma(d) \)
Detector Design

Residue Vector Properties

Use superscript $c$ to denote compromised system.

- $H_0$: $z_k \sim N_0(0, CPC^T + R)$, normal operation
- $H_1$: $z_k^c \sim N_1(\mu_k^c, CPC^T + R + \Sigma)$, under stealthy attack

$$\mu_k^c \triangleq -C \sum_{i=-\infty}^{k} [(A + BL)(I - KC)]^{k-i} B \zeta_i^c, \quad \Sigma = \phi(\Gamma(0), \Gamma(1), \Gamma(2), \ldots)$$

$\phi$ is linear

Use Neyman Pearson Detector

Maximize probability of detection $\beta_k^c$, for given probability of false alarm $\alpha$.

$$g(z_k) = z_k^T \bar{P}^{-1} z_k - (z_k - \mu_k^c)^T (\bar{P} + \Sigma)^{-1} (z_k - \mu_k^c)$$

$$\bar{P} = CPC^T + R$$

\[ H_0 \leq H_1 \]

$\eta_k$
Watermark Parameter Design

Desired Optimization: *Maximize Asymptotic Detection*

\[
\max_{\Gamma(d)} \lim_{k \to \infty} \beta_k^c
\]

subject to \( \Delta J \leq \delta \)

Challenge: difficult to obtain expression for \( \beta_k^c \)

Possible Metric: *Kullback-Liebler Distance:*

\[
E[D_{kl}(N_1||N_0)] = \text{tr}(\Sigma \bar{P}^{-1}) - \frac{1}{2} \log[\text{det}(I+\Sigma \bar{P}^{-1})]
\]

KL distance is convex in \( \Gamma(d) \). Can not perform concave maximization

Objective Relaxation:

\[
\text{tr}(\Sigma \bar{P}^{-1}) \leq E[D_{kl}(N_1||N_0)] \leq \text{tr}(\Sigma \bar{P}^{-1}) - \frac{1}{2} \log[\text{tr}(\Sigma \bar{P}^{-1})]
\]

Bounds are monotonically increasing in \( \text{tr}(\Sigma \bar{P}^{-1}) \)
Watermark Parameter Design

**Optimization Problem**

\[
\max_{\Gamma(d)} \quad \text{tr}(\Sigma \bar{P}^{-1}) \\
\text{subject to} \quad \Delta J \leq \delta
\]

Challenge: Infinitely many optimization variables

**Bochner’s Theorem**

\(\Gamma(d)\) is the auto-covariance function of a stationary Gaussian process \(\{\zeta_k\}\), if and only if there exists a unique positive definite Hermitian measure \(\nu\) such that

\[
\Gamma(d) = \int_{-1/2}^{1/2} e^{2\pi j d \omega} d \nu(\omega)
\]

**Alternative Expressions:**

Partition \([0, \frac{1}{2}]\) into disjoint intervals \(I_1, \ldots, I_q\) of maximal length \(\sigma\).

\[
\Gamma(d) = \lim_{\sigma \to 0} 2Re \left[ \sum_{i=1}^{q} e^{2\pi j d \omega_i} \nu(I_i) \right], \quad \omega_i \in I_i
\]
Watermark Parameter Design

Optimization Problem

\[
\begin{align*}
\max_{\Gamma(d)} & \quad \text{tr}(\Sigma \bar{P}^{-1}) \\
\text{subject to} & \quad \Delta J \leq \delta \\
\end{align*}
\]

\[
\text{tr}(\Sigma \bar{P}^{-1}) = \lim_{\sigma \to 0} \sum_{i=1}^{q} \text{tr}[F_2(\omega_i, \nu(I_i))C^T \bar{P}^{-1}C], \quad \Delta J = \lim_{\sigma \to 0} \sum_{i=1}^{q} F_1(\omega_i, \nu(I_i))
\]

Alternative Formulation

\[
\psi = \max_{H, \omega} \text{tr}(F_2(\omega, H)C^T \bar{P}^{-1}C), \quad H_*, \omega_* \text{ maximizers} \\
\text{subject to} \quad F_1(\omega, H) \leq \delta, \quad 0 \leq \omega \leq 0.5, \quad H \geq 0
\]

\(F_1, F_2\) are linear in \(H \rightarrow \text{tr}(\Sigma \bar{P}^{-1}) \leq \psi\) for \(\Delta J \leq \delta\)

Suppose \(\nu(I_i) = H_* I_{\{\omega_* \in I_i\}} + \bar{H}_* I_{\{-\omega_* \in I_i\}}\) \(\xrightarrow{\text{yields}}\) \(\Delta J = \delta, \quad \text{tr}(\Sigma \bar{P}^{-1}) = \psi\)
Input Generation

Result: $\Gamma^*(d) = 2\text{Re} [\exp(2\pi j d \omega^*) H^*]$, Note that: $H^* = hh^H$

Watermark Generation

1) $\xi_0 \sim N(0, I)$,

2) $\xi_{k+1} = A_\omega \xi_k$,

3) $\zeta_k = C_h \xi_k$,

$\Gamma^*(d) = C_h A_\omega^d C_h^T$

$C_h = \sqrt{2}[h_r \ h_i]$

$A_\omega = \begin{bmatrix} \cos(2\pi \omega^*) & -\sin(2\pi \omega^*) \\ \sin(2\pi \omega^*) & \cos(2\pi \omega^*) \end{bmatrix}$

Problem: Attacker knows one $\zeta_k$, he can determine all $\zeta_k$

Solution: Suboptimal approach, add randomness at each step

2) $\xi_{k+1} = \rho A_\omega \xi_k + \lambda_k$, $\lambda_k \sim N(0, (1 - \rho^2)I)$, $0 \leq \rho \leq 1$
Simulation: Power versus Size

Asymptotic Probability of Detection

Probability of False Alarm

- **Optimal**
- **Sub-Optimal, \( \rho = 0.9 \)**
- **IID**
Simulation: Improvement over IID

% Improvement of Asymptotic Detection

Probability of False Alarm

Optimal

Sub-Optimal $\rho = 0.9$
Simulation: Power vs Cost
Summary and Conclusions

• Reviewed some security challenges in cyber-physical systems.

• Considered strong attack model.

• Proposed watermarking technique to detect attacks.

• Analyzed and discussed design of parameters for watermarking schemes.

Future Work: Develop and analyze suboptimal approach
Thank You!