高技术通讯 2022 年 第 32 卷 第 6 期:607-614

doi:10.3772/j.issn.1002-0470.2022.06.006

基于“平滑水印”的隐蔽性重放攻击检测技术①

史立明②* 刘 斌③** 胡 勇***

(*武汉科技大学冶金自动化与检测技术教育部工程研究中心 武汉 430081)

(**湖北省冶金过程系统科学重点实验室 武汉 430081)

(***北京控制工程研究所 北京 100190)

摘 要 重放攻击在稳定的信息物理系统中对卡方检测器具有隐蔽性。以损失一定的控制性能为代价,向最优控制量中添加水印信号,可以有效应对该问题。水印信号通常是一组独立同分布的高斯噪声序列。然而,很多实际的工业被控对象是慢过程,其自身的大惯性会极大削弱水印信号的效果。针对上述问题,本文提出了一种水印信号改进方法,该方法的核心思想是使用采样和插值算法对水印信号进行“平滑化”处理。平滑之后的水印信号呈现出“低频性”,在慢过程中依然有明显响应,从而使卡方检测器有效识别重放攻击。本文使用单容水箱液位控制系统进行仿真实验,实验结果表明,在适当的平滑周期下,“平滑水印”可以在降低控制性能损失的同时,有效提升攻击检测效果。

关键词 重放攻击;隐蔽性;慢过程;大惯性;平滑水印

0 引 言

网络攻击检测(network attack detection)是信息物理系统(cyber-physical system, CPS)安全研究领域的一个热门话题。检测方法基本上分为两类,一类是基于时间序列分析的异常检测方法,另一类是基于观测器的检测方法[1]。在随机系统(stochastic system)中,基于观测器的卡方检测器因精度良好且计算量小而被广泛使用[2]。卡方检测器受残差驱动,当残差的统计特征发生变化时,检测器发出警报。文献[3]提出了一种具有隐蔽性的重放攻击策略,该攻击方法不会引起残差统计特征发生变化,导致卡方检测器失效。文献[4]对这种隐蔽性进行了严密的数学论证。为应对该问题,文献[3]提出向最优控制量中添加独立同分布的高斯噪声序列(independent identical distribution Gaussian noise sequences,IIDGNS)水印信号,攻击者无法获知水印信号的相关信息,从而破坏了重放攻击的隐蔽性。文献[4]使用隐马尔可夫模型(hidden Markov model, HMM)生成平稳随机序列替换 IIDGNS 水印信号,提高了检测效果。由于水印信号是一组噪声序列,向控制系统中注入水印信号无法避免一定的性能损失。文献[3]和文献[4]将水印信号设计问题的数学模型归纳为平衡检测效果与控制性能的最优化问题,后续的相关研究多以此范式展开。文献[5]设计了一种周期性水印信号,在非持续性攻击场景中,可以在保证检测效果的同时,降低控制性能损失。文献[6]进一步优化了水印信号的周期序列,基于水印信号的攻击检测技术应用场景渐趋广泛。文献[7]提出了一种在系统参数未知情况下水印信号的在线设计方法。文献[8]提出了一种时变动态水印信号,以适应更加复杂的被控对象。文献[9]将水印信号应用到多信道攻击场景中。文献[10]将水印信号应用到监控与数据采集(SCADA)系统中

① 国家自然科学基金(61333008)资助项目。

② 男,1997 年生,硕士生;研究方向:网络控制系统;E-mail: 2293197092@qq.com。

③ 通信作者,E-mail: liubin@wust.edu.cn。

(收稿日期:2021-03-16)

的完整性攻击(integrity attack)检测中。文献[11]设计了一种具有时变频率的正弦水印信号,可以改变输出信号的频率分布。文献[12]设计了一种使用水印信号对传感器数据进行加密的数据加密技术。文献[13]将水印加密技术应用到 CPS 中的隐形双通道虚假数据注入攻击的检测中。文献[14]设计了一种可以补偿性能损失的正弦水印信号。

水印信号可以有效破坏重放攻击的隐蔽性,但因自身的噪声属性,其应用场景依然存在限制。文献[6]发现 IIDGNS 水印信号在慢过程中表现较差,并提出了被控对象对水印信号的"敏感性"的概念,但并没有对这种敏感性做过多讨论,且目前也缺乏相关研究。设计适用于慢过程的水印信号是一个难题。

文献[15]指出,在采样频率较高的慢过程中使用白噪声,难以得到任何响应。受此启发,针对慢过程被控对象,本文提出了一种对水印信号进行"平滑化"的改进方法。该方法通过采样和插值算法对 IIDGNS 水印信号进行平滑化处理。平滑后的水印信号呈现出"低频性",适应了慢过程的大惯性。理论分析表明,平滑化处理使水印信号的方差减小、相关性增大。向慢过程中注入"平滑水印",可以在降低系统性能损失的同时,有效提高攻击检测效果。本文使用单容水箱液位控制系统进行了仿真实验。实验结果表明,平滑水印在大惯性的水箱被控对象上有优于 IIDGNS 水印的良好表现。最后,本文通过仿真对比实验,对平滑水印的平滑段长度与被控对象的最小时间常数之间的关系进行了探讨,归纳出了如何选定平滑段长度的一般规律。

1 问题描述

1.1 控制系统

考虑被控对象为线性时不变系统,其离散状态空间模型如下:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}\cdot\boldsymbol{x}_k + \boldsymbol{B}\cdot\boldsymbol{u}_k + \boldsymbol{w}_k \\ \boldsymbol{y}_k &= \boldsymbol{C}\cdot\boldsymbol{x}_k + \boldsymbol{v}_k \end{aligned} \tag{1}$$

其中 $\boldsymbol{x}_k \in R^n$、$\boldsymbol{u}_k \in R^p$ 和 $\boldsymbol{y}_k \in R^m$ 分别代表系统状态变量、控制量和输出变量。过程噪声 $\{\boldsymbol{w}_k \sim N(0,\boldsymbol{Q})\}$ 和观测噪声 $\{\boldsymbol{v}_k \sim N(0,\boldsymbol{R})\}$ 均是独立同分布的高斯过程,且 $\boldsymbol{w}_k$ 和 $\boldsymbol{v}_k$ 相互独立,$\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 分别为相应的正定协方差矩阵。

假定被控对象具有能稳性和能检性。采用卡尔曼滤波器作为系统状态估计器,其表达式如下:

$$\begin{aligned} \hat{\boldsymbol{x}}_{k|k-1} &= \boldsymbol{A}\cdot\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{B}\cdot\boldsymbol{u}_{k-1} \\ \hat{\boldsymbol{x}}_k &= \hat{\boldsymbol{x}}_{k|k-1} + \boldsymbol{K}\cdot(\boldsymbol{y}_k - \boldsymbol{C}\cdot\hat{\boldsymbol{x}}_{k|k-1}) \\ \boldsymbol{z}_k &\triangleq \boldsymbol{y}_k - \boldsymbol{C}\cdot\hat{\boldsymbol{x}}_{k|k-1} \\ \boldsymbol{e}_k &\triangleq \boldsymbol{x}_k - \hat{\boldsymbol{x}}_k \end{aligned} \tag{2}$$

其中 $\hat{\boldsymbol{x}}_{k|k-1}$ 和 $\hat{\boldsymbol{x}}_k$ 分别为 $\boldsymbol{x}_k$ 的先验估计和最优估计,$\boldsymbol{z}_k$ 代表残差,$\boldsymbol{e}_k$ 代表估计误差。闭环系统稳定后,卡尔曼滤波增益收敛到稳态值 $\boldsymbol{K}$。

$$\boldsymbol{K} = \boldsymbol{P}\cdot\boldsymbol{C}^{\mathrm{T}}\cdot(\boldsymbol{C}\cdot\boldsymbol{P}\cdot\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{R})^{-1} \tag{3}$$

定义 $\boldsymbol{P} \triangleq \lim\limits_{k\to\infty}\boldsymbol{P}_k$,$\boldsymbol{P}_k$ 满足如下黎卡提差分方程:

$$\begin{aligned} \boldsymbol{P}_k &= \boldsymbol{A}\cdot\boldsymbol{P}_{k+1}\cdot\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Q} \\ &\quad - \boldsymbol{A}\cdot\boldsymbol{P}_{k+1}\cdot\boldsymbol{C}^{\mathrm{T}}\cdot(\boldsymbol{C}\cdot\boldsymbol{P}_{k+1}\cdot\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{R})^{-1} \\ &\quad \cdot\boldsymbol{C}\cdot\boldsymbol{P}_{k+1}\cdot\boldsymbol{A}^{\mathrm{T}} \end{aligned} \tag{4}$$

控制器采用无限时域线性二次高斯(linear-quadratic-Gaussian control, LQG)控制器[4],其性能指标由状态变量和控制量两部分组成,表达式如下:

$$J = \lim_{N\to\infty} E\left\{\frac{1}{N}\cdot\left[\sum_{k=0}^{N}(\boldsymbol{x}_k^{\mathrm{T}}\cdot\boldsymbol{W}\cdot\boldsymbol{x}_k + \boldsymbol{u}_k^{\mathrm{T}}\cdot\boldsymbol{U}\cdot\boldsymbol{u}_k)\right]\right\} \tag{5}$$

其中 $\boldsymbol{W}$ 和 $\boldsymbol{U}$ 是相应的正定权值矩阵。由分离定理可知,表达式(5)的最优解 $\boldsymbol{u}_k^*$ 是状态最优估计 $\hat{\boldsymbol{x}}_k$ 的线性反馈,可以分别求解卡尔曼滤波问题和 LQG 控制问题,且卡尔曼滤波为最优估计、LQG 控制为最优控制[16]。系统稳定后,控制率收敛到稳态值 L。

$$\boldsymbol{L} = -(\boldsymbol{B}^{\mathrm{T}}\cdot\boldsymbol{S}\cdot\boldsymbol{B} + \boldsymbol{U})^{-1}\cdot\boldsymbol{B}^{\mathrm{T}}\cdot\boldsymbol{S}\cdot\boldsymbol{A} \tag{6}$$

定义 $\boldsymbol{S} \triangleq \lim\limits_{k\to\infty}\boldsymbol{S}_k$,$\boldsymbol{S}_k$ 满足如下黎卡提差分方程:

$$\begin{aligned} \boldsymbol{S}_k &= \boldsymbol{A}^{\mathrm{T}}\cdot\boldsymbol{S}_{k+1}\cdot\boldsymbol{A} + \boldsymbol{W} \\ &\quad - \boldsymbol{A}^{\mathrm{T}}\cdot\boldsymbol{S}_{k+1}\cdot\boldsymbol{B}\cdot(\boldsymbol{B}^{\mathrm{T}}\cdot\boldsymbol{S}_{k+1}\cdot\boldsymbol{B} + \boldsymbol{U})^{-1} \\ &\quad \cdot\boldsymbol{B}^{\mathrm{T}}\cdot\boldsymbol{S}_{k+1}\cdot\boldsymbol{A} \end{aligned} \tag{7}$$

LQG 控制器的最优控制量为

$$\boldsymbol{u}_k^* = \boldsymbol{L}\cdot\hat{\boldsymbol{x}}_k \tag{8}$$

最优控制性能损失为

$$\begin{aligned} J^* &= trace(\boldsymbol{S}\cdot\boldsymbol{Q}) \\ &\quad + trace[(\boldsymbol{A}^{\mathrm{T}}\cdot\boldsymbol{S}\cdot\boldsymbol{A} + \boldsymbol{W} - \boldsymbol{S}) \\ &\quad \cdot(\boldsymbol{P} - \boldsymbol{K}\cdot\boldsymbol{C}\cdot\boldsymbol{P})] \end{aligned} \tag{9}$$

其中,$trace(\boldsymbol{A})$ 表示矩阵 $\boldsymbol{A}$ 的迹。

1.2 重放攻击及攻击检测

考虑系统受到重放攻击,攻击者监听从传感器到控制中心的网络信道,并记录传感器数据。如图1所示,攻击者在 s 时刻开始发动攻击,用历史传感器数据替换当前传感器数据,重放延迟 d 个步长。为简化分析,认为攻击时段和被重放片段之间没有重叠,数学表达式为

$$\boldsymbol{y}_k = \boldsymbol{y}_{k-d} \quad s \leqslant k < s + d \tag{10}$$

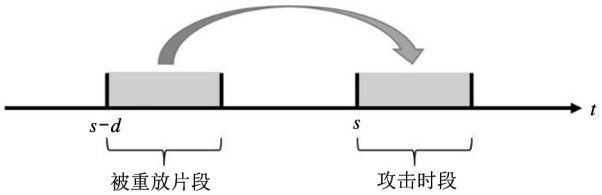

图1 重放攻击模型

如图2所示,攻击者通过通信网络向控制系统发动重放攻击。文献[4]对重放攻击的隐蔽性给出了严密的数学证明,并提出了向最优控制量中添加水印信号的攻击检测方法。图2中,$\Delta\boldsymbol{u}_k \in R^p$ 代表IIDGNS水印信号,记为 $\{\Delta\boldsymbol{u}_k \sim N(0,\boldsymbol{\Lambda})\}$,$\boldsymbol{\Lambda}$ 为水印信号的方差对角矩阵。添加水印信号之后的控制量为

$$\boldsymbol{u}_k = \boldsymbol{u}_k^* + \Delta\boldsymbol{u}_k \tag{11}$$

定义 $\boldsymbol{\Gamma} \triangleq (\boldsymbol{A} + \boldsymbol{B}\cdot\boldsymbol{L})\cdot(\boldsymbol{I} - \boldsymbol{K}\cdot\boldsymbol{C})$,系统在受到攻击时的残差量为

$$\boldsymbol{z}_k = \boldsymbol{z}_{k-d} - \boldsymbol{C}\cdot\boldsymbol{\Gamma}^{k-d}\cdot(\hat{\boldsymbol{x}}_{d+1|d} - \hat{\boldsymbol{x}}_{1|0}) - \boldsymbol{C}\cdot\sum_{i=d+1}^{k}\boldsymbol{\Gamma}^{k-i}\cdot\boldsymbol{B}\cdot(\Delta\boldsymbol{u}_i - \Delta\boldsymbol{u}_{i-d}) \tag{12}$$

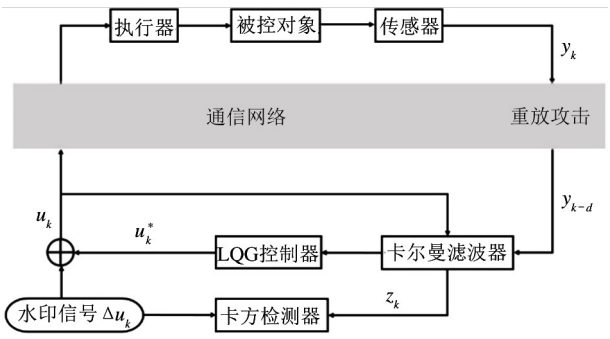

图2 添加水印信号检测重放攻击

假定 $\boldsymbol{\Gamma}$ 稳定,当 k 趋于无穷大时,式(12)等号右边的第2项收敛到0,第3项中 $\Delta\boldsymbol{u}_i$ 和 $\Delta\boldsymbol{u}_{i-d}$ 与 $\boldsymbol{z}_{k-d}$ 均不相关。卡方检测器受残差驱动,文献[4]中给出了卡方检测器的二元假设如下。

未受攻击时,$H_0:\boldsymbol{z}_k \sim N(\boldsymbol{0},\boldsymbol{\rho})$;受到攻击时,$H_1:\boldsymbol{z}_k \sim N(\boldsymbol{\mu}_{k-1},\boldsymbol{\rho}+\boldsymbol{\Sigma}_{k-1})$,其中,

$$\begin{aligned}\boldsymbol{\rho} &= \boldsymbol{C}\cdot\boldsymbol{P}\cdot\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{R}\\ \boldsymbol{\mu}_k &= -\boldsymbol{C}\cdot\sum_{i=s}^{k}\boldsymbol{\Gamma}^{k-i}\cdot\boldsymbol{B}\cdot\Delta\boldsymbol{u}_i\\ \boldsymbol{\Sigma}_k &= 2\cdot Cov[\boldsymbol{C}\cdot\sum_{i=s}^{k}\boldsymbol{\Gamma}^{k-i}\cdot\boldsymbol{B}\cdot\Delta\boldsymbol{u}_{i-d}]\end{aligned} \tag{13}$$

由于控制系统为线性系统,为简化计算,可以将重放攻击起始点设置为时间轴原点,即 $s = 0$。卡方检测器的检测指标为[4]

$$g(\boldsymbol{z}_k) = \boldsymbol{z}_k^{\mathrm{T}}\cdot\boldsymbol{\rho}^{-1}\cdot\boldsymbol{z}_k - (\boldsymbol{z}_k - \boldsymbol{\mu}_{k-1})^{\mathrm{T}}\cdot(\boldsymbol{\rho}+\boldsymbol{\Sigma})^{-1}\cdot(\boldsymbol{z}_k - \boldsymbol{\mu}_{k-1}) \tag{14}$$

其中,

$$\begin{aligned}\boldsymbol{\Sigma} &= \lim_{k\to\infty}\boldsymbol{\Sigma}_k\\ &= 2\cdot\sum_{i=0}^{\infty}\boldsymbol{C}\cdot\boldsymbol{\Gamma}^i\cdot\boldsymbol{B}\cdot\boldsymbol{\Lambda}\cdot\boldsymbol{B}^{\mathrm{T}}\cdot(\boldsymbol{\Gamma}^i)^{\mathrm{T}}\cdot\boldsymbol{C}^{\mathrm{T}}\end{aligned} \tag{15}$$

定义 η 为检测阈值,当 $g(\boldsymbol{z}_k) < \eta$ 时,接受假设 H_0,即系统运行于正常状态;当 $g(\boldsymbol{z}_k) \geqslant \eta$ 时,则接受假设 H_1,即系统受到攻击。卡方检测器的渐进检测率为

$$\beta \triangleq \lim_{k\to\infty}P(g(\boldsymbol{z}_k) \geqslant \eta) \tag{16}$$

文献[4]指出,β 是式(15)中 $\boldsymbol{\Sigma}$ 的单调增函数。从式(15)可知,$\boldsymbol{\Sigma}$ 仅与 $\boldsymbol{\Lambda}$ 有关。所以,对于IIDGNS水印信号,卡方检测器的渐进检测率 β 是仅关于水印信号方差 $\boldsymbol{\Lambda}$ 的单调增函数。

添加水印信号后系统的性能损失为[3]

$$J = J^* + trace[(\boldsymbol{B}^{\mathrm{T}}\cdot\boldsymbol{S}\cdot\boldsymbol{B} + \boldsymbol{U})\cdot\boldsymbol{\Lambda}] \tag{17}$$

J^* 为式(9)中的最优控制性能损失。定义 ΔJ 代表添加水印信号后产生的额外性能损失,有:

$$\Delta J \triangleq trace[(\boldsymbol{B}^{\mathrm{T}}\cdot\boldsymbol{S}\cdot\boldsymbol{B} + \boldsymbol{U})\cdot\boldsymbol{\Lambda}] \tag{18}$$

显然,ΔJ 也是关于 $\boldsymbol{\Lambda}$ 的单调增函数。

根据上述分析可知,当水印信号为独立同分布的高斯噪声序列时,卡方检测器的渐进检测率和系统额外性能损失只与水印信号的方差有关,且均是方差的单调增函数。提高水印信号的方差可以提高

攻击检测效果,但同时也会导致系统性能损失升高。通常,牺牲一定的控制性能来换取攻击检测效果是值得的。但往往很多工业被控对象是慢过程,在较高的采样频率下,对 IIDGNS 水印几乎无响应。提高水印信号的方差不仅会使系统性能损失增加,而且对攻击检测没有太大帮助。而降低采样频率又会导致整个系统的性能大幅下降。对于慢过程被控对象,IIDGNS 并不是理想的水印信号。

2　平滑水印设计

为解决慢过程对 IIDGNS 响应较弱的问题,本文将 IIDGNS 水印信号进行“平滑化”处理,使其呈现出“低频性”,以适应慢过程的大惯性。如图 3 所示,在连续的 2 个 IIDGNS 水印信号点之间均匀采样,生成若干新的信号点替换原始信号点,以达到平滑的目的。$\{\Delta \boldsymbol{u}_k\}$ 是原始水印信号,$\{\Delta \boldsymbol{u}'_n\}$ 代表平滑水印信号。平滑水印在各平滑段上服从均匀分布,记为 $\Delta \boldsymbol{u}'_n \sim U(\Delta \boldsymbol{u}_k, \Delta \boldsymbol{u}_{k+1})$,表示 $\Delta \boldsymbol{u}'_n$ 服从于 $(\Delta \boldsymbol{u}_k, \Delta \boldsymbol{u}_{k+1})$ 之间的均匀分布。为简化计算,使平滑水印在各平滑段上相互独立。具体平滑方法如下。

步骤 1　确定平滑段长度 q。

步骤 2　从初始时刻开始,取连续的 2 个信号点 $\Delta \boldsymbol{u}_k$ 和 $\Delta \boldsymbol{u}_{k+1}$。

步骤 3　在 $\Delta \boldsymbol{u}_k$ 和 $\Delta \boldsymbol{u}_{k+1}$ 之间均匀采样 q 个新的信号点,组成一段平滑水印。数学表达式为

$$\begin{aligned} &\Delta \boldsymbol{u}'_n = \Delta \boldsymbol{u}_k + (\Delta \boldsymbol{u}_{k+1} - \Delta \boldsymbol{u}_k) \cdot y_n \\ &(k-1)/2 \cdot q + 1 \leqslant n \leqslant (k+1)/2 \cdot q \end{aligned} \tag{19}$$

其中 $\{y_n\}$ 相互独立,且均服从于(0,1)之间的均匀分布,记为 $\{y_n \sim U(0,1)\}$。y_n 与 $\Delta \boldsymbol{u}_k$ 相互独立。

步骤 4　更新信号点为 $\Delta \boldsymbol{u}_{k+2}$ 和 $\Delta \boldsymbol{u}_{k+3}$,重复执行步骤 3。

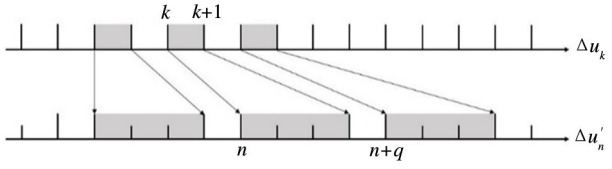

图 3　IIDGNS 水印信号的平滑化处理

平滑水印的均值和方差为

$$E\Delta \boldsymbol{u}'_n = E[\Delta \boldsymbol{u}_k + (\Delta \boldsymbol{u}_{k+1} - \Delta \boldsymbol{u}_k) \cdot y_n] = 0 \tag{20}$$

$$\begin{aligned} D\Delta \boldsymbol{u}'_n &= D[\Delta \boldsymbol{u}_k + (\Delta \boldsymbol{u}_{k+1} - \Delta \boldsymbol{u}_k) \cdot y_n] \\ &= E(\Delta \boldsymbol{u}_k + (\Delta \boldsymbol{u}_{k+1} - \Delta \boldsymbol{u}_k) \cdot y_n)^2 \\ &= E[(1-y_n)^2 \cdot \Delta \boldsymbol{u}_k^2 + y_n^2 \cdot \Delta \boldsymbol{u}_{k+1}^2 \\ &\quad + 2 \cdot (1-y_n) \cdot y_n \cdot \Delta \boldsymbol{u}_k \cdot \Delta \boldsymbol{u}_{k+1}] \\ &= E[(1-y_n)^2 \cdot \Delta \boldsymbol{u}_k^2] + E[y_n^2 \cdot \Delta \boldsymbol{u}_{k+1}^2] \\ &= 2/3 \cdot \boldsymbol{\Lambda} \end{aligned} \tag{21}$$

由上述计算可知,平滑水印 $\{\Delta \boldsymbol{u}'_n\}$ 在整个时间域上零均值同分布。平滑水印的均值为 0,在无穷时域上避免了在控制量中产生直流偏置。从式(18)可以得出,平滑水印导致的额外性能损失为

$$\begin{aligned} \Delta J' &= trace[(\boldsymbol{B}^{\mathrm{T}} \cdot \boldsymbol{S} \cdot \boldsymbol{B} + \boldsymbol{U}) \cdot (2/3 \cdot \boldsymbol{\Lambda})] \\ &= 2/3 \cdot \Delta J \end{aligned} \tag{22}$$

显然,采用该平滑化处理方法,可以明显降低系统性能损失。

由于平滑水印在各平滑段上相互独立,只需分析平滑水印在一个平滑段内的相关性。平滑水印的自协方差为

$$\begin{aligned} Cov(\Delta \boldsymbol{u}'_n, \Delta \boldsymbol{u}'_{n+m}) &= E[(\Delta \boldsymbol{u}_k + (\Delta \boldsymbol{u}_{k+1} - \Delta \boldsymbol{u}_k) \cdot y_n) \\ &\quad \cdot (\Delta \boldsymbol{u}_k + (\Delta \boldsymbol{u}_{k+1} - \Delta \boldsymbol{u}_k) \cdot y_{n+m})] \\ &= 2 \cdot E(y_n \cdot y_{n+m} \cdot \Delta \boldsymbol{u}_k^2) \\ &= 1/2 \cdot \boldsymbol{\Lambda} \end{aligned} \tag{23}$$

定义 $R(h) \triangleq Cov(\Delta \boldsymbol{u}'_n, \Delta \boldsymbol{u}'_{n+h})$ 表示 $\{\Delta \boldsymbol{u}'_n\}$ 在一个平滑段内的自相关函数,可归纳为

$$R(h) = \begin{cases} 2/3 \cdot \boldsymbol{\Lambda} & h = 0 \\ 1/2 \cdot \boldsymbol{\Lambda} & 0 < h < q \\ 0 & h = q \end{cases} \tag{24}$$

由式(13)可以计算出,添加平滑水印后残差的方差变化量为

$$\boldsymbol{\Sigma}'_n = 2 \cdot Cov[\boldsymbol{C} \cdot \sum_{i=0}^{n} \boldsymbol{\Gamma}^{n-i} \cdot \boldsymbol{B} \cdot \Delta \boldsymbol{u}'_{i-d}] \tag{25}$$

定义 $\boldsymbol{\Sigma}' \triangleq \lim\limits_{n \to \infty} \boldsymbol{\Sigma}'_n$,

$$\begin{aligned} \boldsymbol{\Sigma}' &= 2 \cdot \sum_{i=0}^{\infty} \boldsymbol{C} \cdot \boldsymbol{\Gamma}^i \cdot \boldsymbol{B} \cdot (2/3 \cdot \boldsymbol{\Lambda}) \cdot \boldsymbol{B}^{\mathrm{T}} \cdot (\boldsymbol{\Gamma}^i)^{\mathrm{T}} \\ &\quad \cdot \boldsymbol{C}^{\mathrm{T}} + \Delta \boldsymbol{\Sigma} \\ &= 2/3 \cdot \boldsymbol{\Sigma} + \Delta \boldsymbol{\Sigma} \end{aligned} \tag{26}$$

其中,

$$\Delta\boldsymbol{\Sigma} = 2 \cdot \left[\sum_{h=1}^{q-1} \boldsymbol{C} \cdot (\boldsymbol{\Gamma}^h \cdot \Phi(h) + (\boldsymbol{\Gamma}^h \cdot \Phi(h))^{\mathrm{T}}) \cdot \boldsymbol{C}^{\mathrm{T}} \right] \tag{27}$$

$$\Phi(h) = \sum_{k=0}^{\infty} \sum_{i=0}^{q-h-1} \boldsymbol{\Gamma}^{k \cdot q+i} \cdot \boldsymbol{B} \cdot (1/2 \cdot \boldsymbol{\Lambda}) \cdot \boldsymbol{B} \cdot (\boldsymbol{\Gamma}^{k \cdot q+i})^{\mathrm{T}} \tag{28}$$

定义 $\Delta\boldsymbol{\Sigma}' \triangleq \lim\limits_{q \to \infty} \Delta\boldsymbol{\Sigma}$,

$$\Delta\boldsymbol{\Sigma}' = 2 \cdot \left[\sum_{h=1}^{q} \boldsymbol{C} \cdot (\boldsymbol{\Gamma}^h \cdot \boldsymbol{\Theta} + (\boldsymbol{\Gamma}^h \cdot \boldsymbol{\Theta})^{\mathrm{T}}) \cdot \boldsymbol{C}^{\mathrm{T}} \right] \tag{29}$$

$$\boldsymbol{\Theta} = \sum_{i=0}^{\infty} \boldsymbol{\Gamma}^i \cdot \boldsymbol{B} \cdot (1/2 \cdot \boldsymbol{\Lambda}) \cdot \boldsymbol{B} \cdot (\boldsymbol{\Gamma}^i)^{\mathrm{T}} \tag{30}$$

在实际应用中,平滑段长度不可能设置到无穷大,有如下结论:

$$2/3 \cdot \boldsymbol{\Sigma} < \boldsymbol{\Sigma}' \leqslant 2/3 \cdot \boldsymbol{\Sigma} + \Delta\boldsymbol{\Sigma}' \tag{31}$$

$\boldsymbol{\Sigma}'$ 在此范围内随平滑段长度 q 的增大而增大。

定义 β' 表示在平滑水印作用下卡方检测器的渐进检测率。由式(16)及相关分析可知,β' 是关于 $\boldsymbol{\Sigma}'$ 的单调增函数。

由式(22)可知,添加平滑水印后,系统额外性能损失 $\Delta J'$ 是仅关于原始水印信号方差 $\boldsymbol{\Lambda}$ 的单调增函数。由式(25)~(31)及相关分析可知,卡方检测器的渐进检测率 β' 是关于 $\boldsymbol{\Lambda}$ 和平滑段长度 q 的函数。当 $\boldsymbol{\Lambda}$ 一定时,$\boldsymbol{\Sigma}'$ 是仅关于 q 的单调增函数,即卡方检测器的渐进检测率 β' 在一定范围内随着平滑段长度 q 的增大而增大。综上分析,本文提出的平滑水印可以将系统额外的性能损失降低 1/3,而攻击检测率可以通过增大平滑段长度而达到理想水平。

3 仿真实验

考虑平滑水印应用对象的大惯性特性,本文使用单容水箱液位控制系统进行仿真实验。广义被控对象的传递函数为 $G(s) = 10/[(s+1) \cdot (300s+1)]$,其最小时间常数 $T_{\min} = 1$ s。根据采样定理,采样周期 T_S 一般取为 $1/10 \cdot T_{\min} \leqslant T_S \leqslant 1/4 \cdot T_{\min}$,本实验选择采样周期为 0.2 s。单容水箱为单输入单输出系统,最小实现的状态变量为二维向量,其离散状态空间模型的参数如下:

$$\boldsymbol{A} = \begin{bmatrix} 0.8181 & -0.0006 \\ 0.1812 & 0.9999 \end{bmatrix},$$

$$\boldsymbol{B} = \begin{bmatrix} 0.1812 \\ 0.0187 \end{bmatrix}, \ \boldsymbol{C} = \begin{bmatrix} 0 & 0.0333 \end{bmatrix} \tag{32}$$

LQG 控制器性能指标的正定权值矩阵分别为

$$\boldsymbol{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \boldsymbol{U} = 1 \tag{33}$$

系统的过程噪声和观测噪声的协方差矩阵分别为

$$\boldsymbol{Q} = \begin{bmatrix} 0.8 & 0 \\ 0 & 0.8 \end{bmatrix}, \ \boldsymbol{R} = 1 \tag{34}$$

LQG 控制器的最优代价 $J^* = 53.0141$,卡方检测器的误报率控制在约 0.02 的水平。

图 4 显示了向慢过程中添加 IIDGNS 水印信号时,卡方检测器的渐进检测率与性能损失比之间的关系。表 1 是其中的部分数据。分析表 1 数据可知,当控制性能损失与最优性能损失比低于约 10 时,卡方检测器的检测率大致与误报率(约 0.02)持平。若获取约 0.1 的检测率,控制性能损失大约达到最优性能损失的 15 倍。而要获取约 0.5 的检测率,则性能损失将超过最优性能损失的 100 倍,这显然是无法接受的结果。

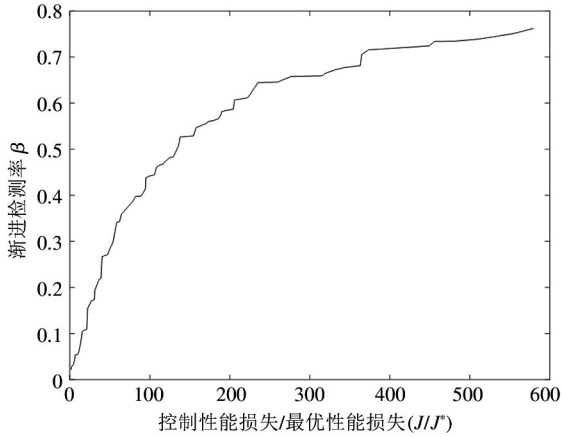

图 4 IIDGNS 水印检测率(β)与损失比(J/J^*)的关系

表 1 IIDGNS 水印检测率与损失比关系的部分数据

β	0.0273	0.0306	0.0538	0.0555	0.1055	0.5077
J/J^*	2.108	3.310	6.749	10.288	16.049	135.983

图 5 为向系统中添加 IIDGNS 水印及相应平滑水印所产生的检测率的对比图。由图 5 可知,从总

体的趋势来看，对 IIDGNS 水印信号进行平滑处理后，可以显著提高卡方检测器的渐进检测率。同时，由于水印信号的方差是驱动卡方检测器的主要因素，当原始水印的方差设置在较低水平时，添加平滑水印与原始水印产生的检测效果并没有明显区别。而当水印信号的方差较大时，平滑水印产生的检测效果明显优于 IIDGNS 水印信号。从图中可以看出，在对 IIDGNS 水印信号进行平滑化处理时，平滑段长度设置得越大，平滑水印产生的检测效果越好。同时也可以看出，当平滑段长度增大到一定水平后，继续增大平滑段长度，对检测率的提升则十分有限。

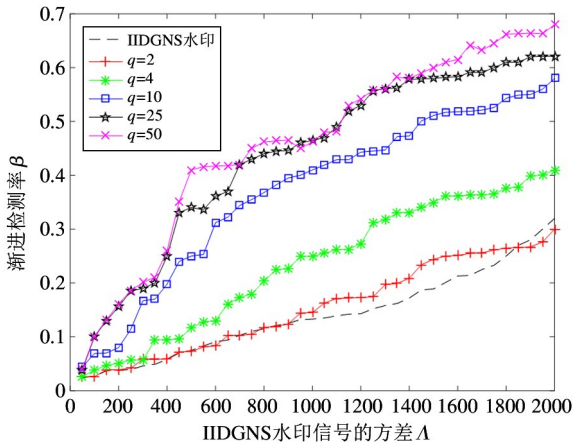

图 5　原始水印及其平滑水印产生的检测率对比
（q 为平滑段长度）

图 6 为添加 IIDGNS 水印及相应平滑水印所导致的性能损失的对比图。从图中可以看出，原始水印进行平滑化处理后，系统额外性能损失显著降低

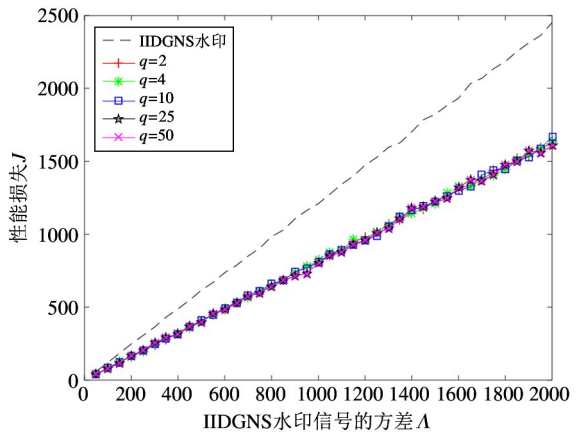

图 6　原始水印及其平滑水印导致的性能损失对比

（比例约为 1/3），且平滑段长度的变化对性能损失并没有影响，与式（22）所述结论一致。

图 7 详细显示了渐进检测率与平滑段长度的关系。从图中可以看出，增大平滑段长度可以有效提高卡方检测器的渐进检测率，但检测率并不会随着平滑段长度的增大而无限升高。可以明显看出，在本次实验中，平滑段长度在小于 20 个步长时，曲线的斜率较大，增大平滑段长度对检测率的提升十分明显。随后，随着平滑段长度继续增大，检测率的提升变缓。显然，设置过大的平滑段长度并没有必要。

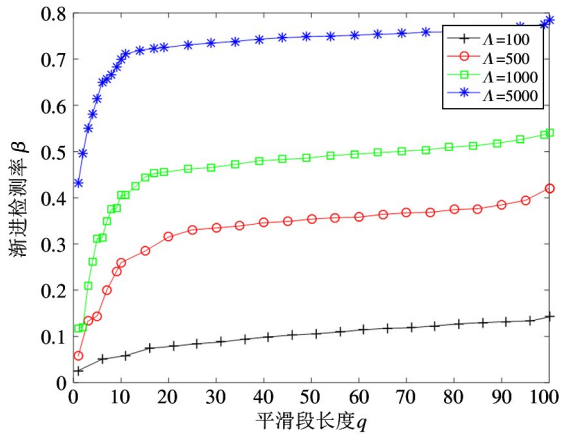

图 7　检测率与平滑段长度的关系

式（20）表明平滑水印的均值为 0，保证了在无穷时域上控制量中无直流偏置。但是，当平滑段长度设置过大时，由于平滑水印在每个平滑段上服从均匀分布，会导致控制量在每个平滑段上产生直流偏置，从而影响控制性能。图8显示了平滑段长度

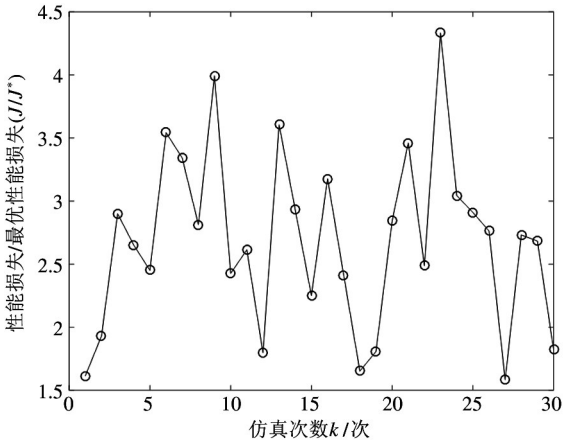

图 8　过大平滑段长度下的性能损失比

设置过大时系统性能损失比的 30 次仿真数据。从图中可以看出,多次仿真的数据波动较大。这是因为平滑段长度过大,所选取的原始水印信号点过少,随机性较大。平滑水印过于依赖所选取的原始水印信号点,可能会在控制量在中产生直流偏置。

综上分析,水印信号的平滑化处理能在降低控制性能损失的同时,明显提升重放攻击检测率,此结论与理论分析相符。同时也表明,在实际应用中,平滑段长度需要设置在合适的水平,才能在不影响控制性能的同时最大程度地提高检测效果。

表 2 是对不同最小时间常数的被控对象进行仿真实验所总结的最优平滑段长度数据。一般来说,传递函数的最小时间常数决定了系统的响应速度,时间常数越小,响应速度越快。由此可归纳出设置平滑段长度的一般规律,即平滑段长度应随着最小时间常数的增大而增大,才能使水印信号适应被控对象的响应速度,从而获取较高的攻击检测率。

表 **2** 最小时间常数与最优平滑段长度的关系

最小时间常数	0.1 s	1 s	5 s	10 s	15 s
最优平滑段长度	约 10	约 20	约 35	约 45	约 50

4 结 论

本文针对一类慢过程被控对象提出了一种基于“平滑水印”的重放攻击检测方法。由于慢过程对高频白噪声几乎无响应,传统的独立同分布高斯噪声水印信号并不能有效破坏重放攻击的隐蔽性。本文从“降频”的角度出发,对独立同分布高斯噪声水印信号进行“平滑化”处理,使水印信号呈现出“低频性”,以适应慢过程的大惯性特性。通过均匀采样和插值的方法平滑高斯水印,并证明了该平滑化处理使水印信号的方差减小、相关性增加,可以在降低控制性能损失的同时,有效提升攻击检测效果。本文选取单容水箱液位控制系统进行仿真实验,验证了该方法的有效性,并通过实验总结出了根据被控对象的最小时间常数设置平滑段长度的一般规律。实验结果也显示出了该方法的两个主要的不足之处。一是在水印信号的方差过小时,平滑化处理对检测效果的提升并不明显;二是当平滑段长度设置过大时,由于平滑水印在各个平滑段上服从均匀分布,可能会给控制量增加直流偏置,从而影响系统性能。后续的研究将从这两点出发,着重于量化被控对象对水印信号的“敏感性”,提出性能更优的水印信号设计方法。

参考文献

[1] 敖伟. 信息物理系统中攻击检测与安全状态估计问题研究[D]. 重庆:重庆大学自动化学院, 2017: 17-20

[2] MEHRA R K, PESCHON J. An innovations approach to fault detection and diagnosis in dynamic systems[J]. *Automatica*, 1971, 7(5): 637-640

[3] MO Y, SINOPOLI B. Secure control against replay attacks[C]//2009 47th Annual Allerton Conference on Communication, Control, and Computing, Allerton, USA, 2009: 911-918

[4] MO Y, WEERAKKODY S, SINOPOLI B. Physical authentication of control systems designing watermarked control inputs to detect counterfeit sensor outputs[J]. *IEEE Control Systems Magazine*, 2015, 35(1): 93-109

[5] FANG C, QI Y, CHENG P, et al. Cost-effective watermark based detector for replay attacks on cyber-physical systems[C]//2017 11th Asian Control Conference, Gold Coast, Australia, 2017: 940-945

[6] FANG C, QI Y, CHENG P, et al. Optimal periodic watermarking schedule for replay attack detection in cyber-physical systems[J]. *Automatica*, 2020, 112: 108698

[7] LIU H, YAN J, MO Y, et al. An on-line design of physical watermarks[C]//2018 IEEE Conference on Decision and Control, Miami, USA, 2018: 440-450

[8] PORTER M, HESPANHOL P, ASWANI A, et al. Detecting generalized replay attacks via time-varying dynamic watermarking[J]. *IEEE Transactions on Automatic Control*, 2020, 66(8): 3502-3517

[9] WEERAKKODY S, MO Y, SINOPOLI B. Detecting integrity attacks on control systems using robust physical watermarking[C]//The 53rd IEEE Conference on Decision and Control, Los Angeles, USA, 2014: 3757-3764

[10] MO Y, CHABUKSWAR R, SINOPOLI B. Detecting integrity attacks on SCADA systems[J]. *IEEE Transactions*

on Control Systems Technology, 2014, 22(4): 1396-1407

[11] SÁNCHEZ H S, ROTONDO D, ESCOBET T, et al. Detection of replay attacks in cyber-physical systems using a frequency-based signature[J]. *Journal of the Franklin Institute*, 2019, 356(5): 2798-2824

[12] WANG D, HUANG J, TANG Y, et al. A watermarking strategy against linear deception attacks on remote state estimation under K-L divergence[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(5): 3273-3281

[13] PANG Z, FAN L, SUN J, et al. Detection of stealthy false data injection attacks against networked control systems via active data modification[J]. *Information Sciences*, 2021, 546: 192-205

[14] TRAPIELLO C, ROTONDO D, SANCHEZ H, et al. Detection of replay attacks in CPSs using observer-based signature compensation[C]//2019 6th International Conference on Control, Decision and Information Technologies, Paris, France, 2019: 1-6

[15] 朱豫才. 过程控制的多变量系统辨识[M]. 长沙:国防科技大学出版社, 2005: 149-160

[16] KUMAR P R, VARAIYA P. Stochastic Systems: Estimation, Identification, and Adaptive Control[M]. Upper Saddle River: Prentice-Hall, 1986: 107-109

A detection technology for the concealment replay attack based on the 'smooth watermark'

SHI Liming*, LIU Bin**, HU Yong***

(*Engineering Research Center of Metallurgical Automation and Testing Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081)

(**Hubei Key Laboratory of Metallurgical Process System Science, Wuhan 430081)

(***Beijing Control Engineering Research Institute, Beijing 100190)

Abstract

Replay attacks are hidden from χ^2 detectors in stable cyber-physical systems. At the cost of losing certain control performance, adding watermark signal to the optimal control quantity can effectively deal with this problem. The watermark signal is usually a set of independent identical distributed Gaussian noise sequences. However, many actual industrial controlled objects are slow processes, whose large inertia will greatly weaken the effect of watermark signal. Aiming at the above problem, this paper proposes an improvement method for watermark signal. The core idea of this method is to use sampling and interpolation algorithm to smooth the watermark signal. The smoothed watermark signal shows 'low frequency' and still responds obviously in the slow process, so that the χ^2 detector acts. In this paper, the liquid level control system of a single tank is used for simulation experiments. The experimental results show that the 'smooth watermarking' can reduce the loss of control performance and effectively improve the attack detection effect under the appropriate smoothing period.

Key words: replay attack, concealment, slow process, large inertia, smooth watermark