文章编号: 1001-0920(2019)11-2323-07 DOI: 10.13195/j.kzyjc.2018.1796

工业信息物理系统的攻击建模研究

孙子文[1,2†], 张炎棋[1]

(1. 江南大学 物联网工程学院，江苏 无锡 214122；2. 物联网技术应用教育部工程研究中心，江苏 无锡 214122)

摘 要: 无线通信网络的脆弱性使得工业信息物理系统易遭受各类网络攻击. 为了更深入地了解不同网络攻击的特征进而建立有效的防御措施,构建一种线性时不变离散系统的工业信息物理系统结构;深入研究信息物理系统攻击者攻击空间及攻击者攻击模型,采用控制理论方法研究攻击空间模型的模型知识、披露资源和破坏资源的数学表达;对拒绝服务攻击、重放攻击、虚假数据注入攻击3种典型网络攻击的基本特性,以及对应攻击下攻击模型的表现形式进行分析. 通过Simulink/Truetime仿真工具对破坏性和隐蔽性性能进行仿真实验. 结果表明,所研究的攻击空间模型及攻击者攻击模型能够有效地描述网络攻击的攻击特性.

关键词: 工业信息物理系统；模型知识；攻击策略；网络攻击；隐蔽性；异常检测器

中图分类号: TP273 **文献标志码:** A

Research on attack modeling of industrial cyber physical systems

SUN Zi-wen[1,2†], *ZHANG Yan-qi*[1]

(1. School of Internet of Things Engineering，Jiangnan University，Wuxi 214122，China；2. Engineering Research Center of Internet of Things Technology Applications of MOE，Wuxi 214122，China)

Abstract: The fragility of wireless communication networks makes industrial cyber physical systems vulnerable to suffer various types of network attacks. In order to research the characteristics of different network attacks for establishing effective defense measures, this paper constructs an industrial cyber physical systems structure based on linear time-invariant discrete systems. The attack space and the attacker attack model of the cyber physics system are deeply studied. The control theory method is used to establish the mathematical expression of the model knowledge, disclosure resource and the broken resource. The basic characteristics of three typical network attacks, such as denial of service attack, replay attack and false data injection attack, and the manifestations of attack models under corresponding attacks are analyzed. Simulation experiments are conducted to evaluate the performance of destructiveness and concealment by using the Simulink/Truetime simulation tool. The simulation results show that the attack space model and the attacker attack model can effectively describe the attack characteristics of network attacks.

Keywords: industrial cyber physical systems；model knowledge；attack strategy；network attacks；stealthy；anomaly detector

0 引 言

近年来,工业自动化控制与计算机、通信等技术的深层次融合,促使信息与物理对象紧密耦合的信息物理系统(cyber-physical-systems, CPS)得到广泛应用. CPS正逐渐被大规模可靠地应用于工业环境,形成了工业信息物理系统(industrial-cyber-physical-systems, ICPS)[1]. ICPS既面临工业环境存在的严重干扰,又面临信息系统和嵌入式设备的无线网络结合带来的新问题,导致ICPS不可避免地存在安全性、实时性和资源限制等挑战. 在众多的安全性挑战之中,网络攻击已成为ICPS的主要安全威胁之一,亟需从系统的角度来研究ICPS的安全性[2],尤其是针对ICPS中的网络攻击的研究.

对于ICPS网络攻击,从是否需要网络密钥将其分为内部攻击、密钥泄露攻击和外部攻击[3];从攻击的影响位置将其分为对传感器网络的攻击、对传感器的网络节点及相关装置的攻击和对传感器网络流量的攻击[4];而目前比较常见的分类将网络攻击分为物理攻击和通信攻击,通信攻击又分为典型的拒绝服务攻击、重放攻击和虚假数据注入攻击[5]. 针对物理

收稿日期: 2018-12-30；**修回日期:** 2019-06-06.

基金项目: 国家自然科学基金项目(61373126)；中央高校基本科研业务费专项资金项目(JUSRP51510)；江苏省自然科学基金项目(BK20131107).

责任编委: 周彬.

†通讯作者. E-mail: sunziwen@jiangnan.edu.cn.

攻击,可以对节点增加认证和访问控制,只有授权的用户才能访问相应节点的数据[6];针对通信攻击,可通过攻击者与防御者之间的最优博弈论来加强DoS攻击下系统的弹性控制[7];可通过对攻击强度的L_1范数惩罚进行正则化来检测和定位恶意数据攻击[8];可以非常规时间间隔扰乱系统来检测重放攻击[9]. 文献[7-9]在攻击防御方面表现出一定的有效性,但缺乏从模型角度来分析各个攻击的性能. 建模是满足ICPS性能分析要求的前提,为此,文献[10]考虑了离散时间线性动态系统下的基于一类拒绝服务攻击模型的最优控制问题;文献[11]描述了配备χ^2故障检测器的一类离散时间线性时不变系统模型,并以此来检测重放攻击;文献[12]提出了基于传感器和远程估计模型的滤波安全算法以滤去发生攻击时恶意传感器观测数据的影响. 但文献[10-12]针对的攻击类型较为单一,没给出统一框架下的模型结构. 文献[13]则考虑了信息物理攻击下的典型攻击类型,从模型知识、披露资源、破坏资源角度阐述攻击下系统模型,但所提出的模型比较抽象,没有围绕攻击向量进行定性定量分析,缺乏整体的信息物理系统结构框架,缺乏对具体的攻击对象进行细化以验证模型的有效性.

本文进一步深入研究信息物理系统攻击者攻击空间模型及攻击者攻击模型,分析一个完整攻击所必需的一系列资源和条件. 首先,在文献[13]的基础上额外构建基于线性时不变工业信息物理系统的典型结构,构建基于反馈控制系统的攻击模型;然后,采用控制理论方法研究攻击者的攻击空间模型的模型知识、披露资源和破坏资源的数学表达,并将异常检测器作为评判攻击隐蔽性的工具,给出不同攻击下攻击向量的不同数学形式并进行定性定量分析;最后,利用所构建的攻击者攻击模型对Dos攻击、重放攻击、虚假数据注入攻击进行数学及仿真分析.

1 ICPS攻击模型的构建

1.1 工业信息物理系统结构模型

工业信息物理系统由被控对象P、反馈控制器F和异常检测器T组成,其结构如图1所示. 其中:$u(k)$是k时刻控制器发出的控制命令,$\tilde{u}(k)$是执行器接收到的控制命令,$y(k)$是k时刻传感器的测量值,$\tilde{y}(k)$是控制器接收到的传感器信号,$x(k)$是被控对象的状态变量,$w(k)$和$v(k)$是干扰信号,$z(k)$是控制器的状态变量,$y_r(k)$是给定的控制信号,A_P、B_P、C_P、D_P为被控对象的系数矩阵,A_F、B_F、C_F、D_F为反馈控制器的系数矩阵,δ_a和δ_s分别为执行器和传感器通道的攻击信号.

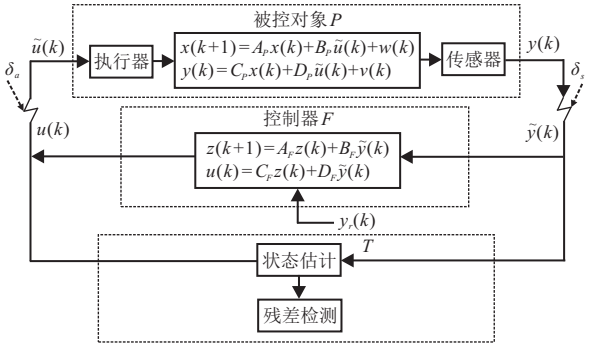


图1 工业信息物理系统结构

1.2 攻击空间与攻击模型

基于攻击者掌握的资源,首次出现了攻击空间的概念,信息物理系统攻击空间模型[13]如图2所示.

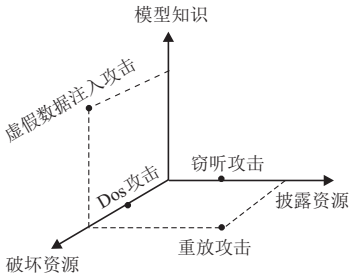


图2 信息物理系统攻击空间模型

攻击空间分为3个维度:先验系统模型知识、披露资源和破坏资源. 先验模型知识可被攻击者用来构造更复杂的攻击,使得攻击更难以检测,并导致更严重的后果. 披露资源用于获得关于系统的敏感信息,其本身不带有破坏性. 攻击者利用掌握的模型知识和披露资源形成攻击者自己的破坏资源,对系统性能进行破坏.

在攻击空间模型基础上对攻击模型[13]进行改进,得到本文的3层通用攻击模型,如图3所示.

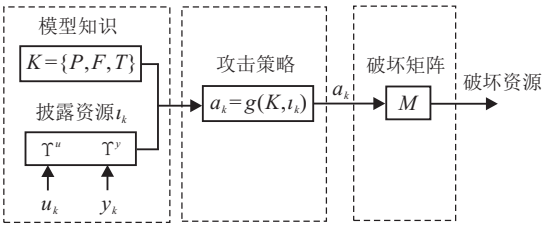


图3 工业信息物理系统攻击者攻击模型

该模型的第1层为攻击者所掌握的先验模型知识K和披露资源ι_k. 攻击者所拥有的先验模型知识由被控对象、反馈控制器和异常检测器构成,即

$$K=\{P,F,T\}. \tag{1}$$

其中:P、F、T分别为被控对象、反馈控制器和异常检测器的模型知识.

披露资源ι_k为攻击者可支配的传感器和执行器数据的集合

$$\iota_k = \iota_{k-1} \bigcup \left\{ \begin{bmatrix} \Upsilon^u & 0 \\ 0 & \Upsilon^y \end{bmatrix} \begin{bmatrix} u_k \\ y_k \end{bmatrix} \right\}, \tag{2}$$

其中Υ^u和Υ^y为二进制关联矩阵,反映执行器或传感器通道数据是否被窃取. 二者的元素取值为1或0:取1时,表示对应的数据通道信息被窃取;取0时,表示对应数据通道的信息安全传输.

第2层为由攻击策略形成的攻击向量,攻击者凭借掌握的先验模型知识K和披露资源ι_k,根据其构成的攻击策略$g(K,\iota_k)$形成攻击向量a_k,即

$$a_k = g(K,\iota_k). \tag{3}$$

第3层为攻击向量生成的破坏资源. 图3中M是破坏矩阵,为攻击者所共有的破坏信息,将攻击向量转换为不同攻击者独特的破坏资源.

一个攻击是否具有破坏性由a_k决定,攻击者根据所掌握的模型知识和披露资源,即拥有的传感器、执行器通道数据信息值以及对系统模型的了解程度,通过各自的攻击策略构成不同的攻击向量,从而发动不同类型的攻击,造成不同的破坏资源.

1.3　系统状态数学模型

1.3.1　无攻击时的系统闭环模型

1) 被控对象和反馈控制器模型.

未受到攻击时,有$\tilde{u}(k)=u(k),\tilde{y}(k)=y(k)$,在不考虑检测器$T$时,系统的模型知识为

$$P\begin{cases} x(k+1)=A_P x(k)+B_P u(k)+w(k), \\ y(k)=C_P x(k)+D_p u(k)+v(k); \end{cases} \tag{4}$$

$$F\begin{cases} z(k+1)=A_F z(k)+B_F y(k), \\ u(k)=C_F z(k)+D_F y(k). \end{cases} \tag{5}$$

将被控对象P的状态量$x(k)$与反馈控制器F的状态量$z(k)$进行融合,则融合闭环系统可表示为

$$\begin{cases} \eta(k+1)=A\eta(k)+Gf(k), \\ y(k)=C\eta(k)+Hf(k), \\ u(k)=E\eta(k)+Nf(k). \end{cases} \tag{6}$$

其中:$\eta(k)=[x^{\mathrm{T}}(k)\quad z^{\mathrm{T}}(k)]^{\mathrm{T}}$为融合状态变量;$f(k)=[w^{\mathrm{T}}(k)\quad v^{\mathrm{T}}(k)]^{\mathrm{T}}$为融合干扰量;模型对应的系数矩阵分别为

$$G=\begin{bmatrix} I & B_P(I-D_F D_P)^{-1}D_F \\ 0 & B_F(I-D_P D_F)^{-1} \end{bmatrix},$$

$$H=[0\quad (I-D_P D_F)^{-1}],$$

$$N=[0\quad (I-D_F D_P)^{-1}D_F],$$

$$A=\begin{bmatrix} A_P+B_P(I-D_F D_P)^{-1}D_F C_P \\ B_F(I-D_P D_F)^{-1}C_P \end{bmatrix} \rightarrow$$

$$\leftarrow \begin{bmatrix} B_P(I-D_F D_P)^{-1}C_F \\ A_F+B_F(I-D_P D_F)^{-1}D_P C_F \end{bmatrix},$$

$$C=[(I-D_P D_F)^{-1}C_P\quad (I-D_P D_F)^{-1}D_P C_F],$$

$$E=[(I-D_F D_P)^{-1}D_F C\quad (I-D_F D_P)^{-1}C_F].$$

通过状态空间方程描述系统,在大多数情况下,被控对象系统的输出方程中并不考虑输入矢量的直接传递,同时为了简化系统,后面的模型都直接假设$D_P=0$,因此有$(I-D_F D_P)^{-1}=I$. 另外,在少数比较复杂的情况下,输入矢量参与输出方程的直接传递,此时要想从输出的测量获得全部状态变量及输出矢量的信息,并能实现对系统的稳定控制,也需假设$D_F D_P$为满秩矩阵进行分析.

2) 异常检测器模型.

这里采用典型的卡尔曼滤波器作为工业信息物理系统的状态估计环节. 在直接传递矩阵$D_P=0$时,已知正常情况下物理对象的模型为

$$\begin{cases} x(k+1)=A_P x(k)+B_P u(k)+w(k), \\ y(k)=C_P x(k)+v(k). \end{cases} \tag{7}$$

根据物理对象设计卡尔曼状态估计器模型

$$\begin{cases} \hat{x}(k|k-1)= \\ A_P\hat{x}(k-1|k-1)+B_P u(k-1), \\ \hat{y}(k)=C_P\hat{x}(k|k-1). \end{cases} \tag{8}$$

其中:$\hat{x}(k|k-1)$是k时刻的状态估计值,$\hat{x}(k-1|k-1)$是$k-1$时刻的状态最优值,$\hat{y}(k)$是k时刻传感器测量数据估计值.

假设物理对象(7)中的噪声$w(k)$和$v(k)$分别是协方差为Q和R的高斯白噪声,可以得到如下的卡尔曼状态更新[14]:

$$\hat{x}t(k|k)=A_P\hat{x}(k-1|k-1)+B_P u(k-1)+ K(k)[y(k)-\hat{y}(k)]. \tag{9}$$

其中:$K(k)$是卡尔曼增益矩阵,满足

$$K(k)= P(k|k-1)C_P^{\mathrm{T}}[C_P P(k|k-1)C_P^{\mathrm{T}}+R]^{-1}; \tag{10}$$

$P(k|k-1)$为k时刻系统状态预测估计的协方差

$$P(k|k-1)=A_P P(k-1|k-1)A_P^{\mathrm{T}}+Q; \tag{11}$$

$P(k|k)$为k时刻最优估计的协方差

$$P(k|k)=[I-K(k)C_p]P(k|k-1). \tag{12}$$

残差$r(k)=y(k)-\hat{y}(k)$,结合式(7)和(8),得

$$r(k)=C_P(x(k)-\hat{x}(k|k-1))+v(k). \tag{13}$$

于是,无攻击时残差$r(k)$的协方差矩阵为

$$V=\mathrm{E}[r(k)\,r^{\mathrm{T}}(k)]=C_PP(k|k-1)C_P^{\mathrm{T}}+R. \quad (14)$$

1.3.2 有攻击时的系统闭环模型

1) 被控对象和反馈控制器模型.

发生攻击时,执行器控制指令变为$\tilde{u}(k)=u(k)+\delta_a$,传感器测量信号变为$\tilde{y}(k)=y(k)+\delta_s$. 在不考虑检测器$T$时,系统模型知识为

$$\begin{cases}x(k+1)=A_Px(k)+B_P(u(k)+\delta_a)+w(k),\\ \tilde{y}(k)=C_Px(k)+v(k)+\delta_s;\end{cases} \quad (15)$$

$$\begin{cases}z(k+1)=A_Fz(k)+B_F(y(k)+\delta_s),\\ \tilde{u}(k)=C_Fz(k)+D_F(y(k)+\delta_s)+\delta_a.\end{cases} \quad (16)$$

同理,融合式(15)和(16),得到攻击后的闭环系统模型

$$\begin{cases}\eta(k+1)=A\eta(k)+Gf(k)+M_1a_k,\\ \tilde{y}(k)=C\eta(k)+Hf(k)+M_2a_k,\\ \tilde{u}(k)=E\eta(k)+Nf(k)+M_3a_k.\end{cases} \quad (17)$$

其中:攻击向量$a_k=[\delta_a^{\mathrm{T}}\ \ \delta_s^{\mathrm{T}}]^{\mathrm{T}}$,其通过破坏矩阵$M$作用于系统进行破坏,破坏矩阵$M$由$M_1$、$M_2$、$M_3$三个子矩阵构成,分别映射闭环系统状态过程、传感器通道和执行器通道的攻击向量,大小为

$$M_1=\begin{bmatrix}B_P & B_PD_F\\ 0 & B_F\end{bmatrix},\ M_2=[0\ \ I],\ M_3=[I\ \ D_F].$$

不难发现,式(17)在(6)的基础上,增加了破坏资源M_1a_k、M_2a_k和M_3a_k,表达了由于攻击造成的影响,其中a_k由执行器和传感器通道上的攻击信号δ_a和δ_s构成. 对于不同类型的攻击,δ_a和δ_s的组成,即所需的模型知识和披露资源,也不尽相同.

2) 异常检测器模型.

类似于无攻击时异常检测器的推导,可得到有攻击时异常检测器的残差$\tilde{r}(k)$为

$$\tilde{r}(k)=y(k)+\delta_s-\hat{y}(k), \quad (18)$$

其协方差矩阵为

$$\tilde{V}=\mathrm{E}[\tilde{r}(k)\,\tilde{r}^{\mathrm{T}}(k)]. \quad (19)$$

发生攻击后残差的协方差矩阵相比攻击前会有一定差异,即定义残差检测器为$\|\tilde{r}(k)\|=\tilde{r}^{\mathrm{T}}(k)\tilde{r}(k)=\mathrm{tr}(\tilde{V})$,通过选定一个合适的阈值$\sigma$,将残差检测器的结果与$\sigma$比较以检测攻击. 假设$H_0$为没发生攻击,$H_1$为发生攻击,检测式如下:

$$\begin{cases}\|r(k)\|<\sigma,\ H_0;\\ \|r(k)\|\geqslant\sigma,\ H_1.\end{cases} \quad (20)$$

即若残差检测结果低于阈值,则接受假设条件H_0,无攻击发生;若检测结果不小于阈值,则接受假设条件H_1,发生了攻击.

2 基于攻击者模型的常见攻击类型分析

2.1 拒绝服务攻击

拒绝服务攻击会阻塞通信信道,造成当前采样时刻信号不能及时送达,反馈回路继续采用上一采样时刻的数据[15]. 描述Dos攻击时,与大多数文献相同,将其以伯努利模型形式进行表示,即

$$\tilde{u}(k)=u(k)-s_k^u(u(k)-u(k-1)), \quad (21)$$

$$\tilde{y}(k)=y(k)-s_k^y(y(k)-y(k-1)), \quad (22)$$

其中s_k^u和s_k^y的元素由0或1构成. 第k个时刻的$s_k^u=0$表示控制指令传输正常,$s_k^u=1$表示执行器通道受到了Dos攻击;同理,$s_k^y=0$表示传感器测量信号传输正常,$s_k^y=1$表示受到了Dos攻击.

将式(21)、(22)代入(15)、(16)中,得到

$$\begin{cases}x(k+1)=A_Px(k)+B_Pu(k)-\\ \qquad B_Ps_k^u(u(k)-u(k-1))+w(k),\\ \tilde{y}(k)=C_Px(k)+v(k)-s_k^y(y(k)-y(k-1));\end{cases} \quad (23)$$

$$\begin{cases}z(k+1)=A_Fz(k)+B_Fy(k)-\\ \qquad B_Fs_k^y(y(k)-y(k-1)),\\ \tilde{u}(k)=C_Fz(k)+D_Fy(k)-s_k^u(u(k)-u(k-1)).\end{cases} \quad (24)$$

可知,若a_k中的$\delta_a=-s_k^u(u(k)-u(k-1))$,$\delta_s=-s_k^y(y(k)-y(k-1))$,则式(17)表示为Dos攻击. Dos攻击采用的攻击策略为$g(\varnothing,\varnothing)$,其中$\varnothing$表示空集,即攻击者在无任何模型知识且无任何传感器以及执行器通道数据下,仅在必要时刻占用通信信道,阻塞信息的及时送达. 这种情况通常被判定为通信延迟.

为简便起见,只考虑传感器通道的攻击,即当$s_k^y=1$时,由式(23)知

$$\tilde{y}(k)=C_Px(k)-(y(k)-y(k-1))+v(k). \quad (25)$$

联立式(7)和(25),得到发生Dos攻击时的传感器测量值为$\tilde{y}(k)=y(k-1)$,发生Dos攻击后的残差$\tilde{r}(k)=\tilde{y}(k)-\hat{y}(k)$. 结合式(8)得

$$\tilde{r}(k)=C_P(x(k-1)-\hat{x}(k|k-1))+v(k-1), \quad (26)$$

则Dos攻击后残差$\tilde{r}(k)$的协方差矩阵为

$$\tilde{V}=C_P(\mathrm{E}((x(k-1)-\hat{x}(k|k-1))\cdot (x(k-1)-\hat{x}(k|k-1))^{\mathrm{T}}))C_P{}^{\mathrm{T}}+R. \quad (27)$$

又由式(14),正常情况下残差$r(k)$的协方差矩阵为

$$V = C_P(\mathrm{E}((x(k)-\hat{x}(k|k-1))\cdot (x(k)-\hat{x}(k|k-1))^{\mathrm{T}}){C_P}^{\mathrm{T}}+R. \tag{28}$$

可见Dos攻击后残差的协方差矩阵会发生变化,所以残差检测器$\|r(k)\|$能够有效地检测出Dos攻击.

2.2 重放攻击

重放攻击的攻击策略是将有效的数据传输被恶意地重复或延迟,通过记录在一定的时间内受损的传感器和执行器的读数并用于事后替换传感器和执行器的实时数据,从而干扰系统的性能.

假设重放攻击发生在$[k_0,k_f]$时间段,重放攻击在第1阶段获取披露资源,即在$k_0 \sim k_r$时间段将传感器和执行器数据通过关联矩阵映射到攻击者,进行数据收集,可表达为

$$\iota_k = \iota_{k-1}\bigcup\left\{\begin{bmatrix}\Upsilon^u & 0\\ 0 & \Upsilon^y\end{bmatrix}\begin{bmatrix}u_k\\ y_k\end{bmatrix}\right\},\ k\in[k_0,k_r]. \tag{29}$$

重放攻击在第2阶段重放披露资源,即在$k_r+1\sim k_f$时间段将之前收集的数据用来替换当前正在传输的数据,达到数据重放的目的,可表达为

$$\iota_k = \iota_{k_r},\ k\in(k_r,k_f]. \tag{30}$$

构造此类攻击不需要知道模型知识,相反却要披露资源,即获取一定的传感器执行器通道信息用于数据的重放构建,所以重放攻击采用了攻击策略$g(\varnothing,\iota_{k_r})$. 与Dos攻击经常被判定为通信延迟不同,重放攻击更像是有意而为之的历史数据回放.

同样,将重放攻击以伯努利模型进行建模

$$\tilde{u}(k) = u(k)-\tau_k^u(u(k)-u(\kappa)), \tag{31}$$

$$\tilde{y}(k) = y(k)-\tau_k^y(y(k)-y(\kappa)). \tag{32}$$

其中: $u(\kappa)$、$y(\kappa)$表示k时刻之前的某时刻的控制和传感器输出信息. τ_k^u和τ_k^y的元素由0或1构成,第k个时刻的$\tau_k^u=0$表示控制指令传输正常,$\tau_k^u=1$表示执行器通道受到了重放攻击;同理,$\tau_k^y=0$表示传感器测量信号传输正常,$\tau_k^y=1$表示传感器通道受到了重放攻击.

与Dos攻击分析类似,即若$a_k=[\delta_a^{\mathrm{T}}\ \ \delta_s^{\mathrm{T}}]^{\mathrm{T}}$,$\delta_a=-\tau^u(u(k)-u(\kappa))$,$\delta_s=-\tau^y(y(k)-y(\kappa))$,则式(17)表示重放攻击. 只考虑传感器通道的攻击,攻击后的传感器测量值为$\tilde{y}(k)=y(\kappa)$,残差为

$$\tilde{r}(k) = C_P(x(\kappa)-\hat{x}(k|k-1))+v(\kappa), \tag{33}$$

发生重放攻击的残差的协方差矩阵为

$$\tilde{V} = C_P(\mathrm{E}((x(\kappa)-\hat{x}(k|k-1))\cdot (x(\kappa)-\hat{x}(k|k-1))^{\mathrm{T}}){C_P}^{\mathrm{T}}+R. \tag{34}$$

由于只考虑了传感器通道的攻击,攻击者并没有捕获执行器通道数据来进行重放,式(8)中对状态值的估计依然为$\hat{x}(k|k-1)=A_P\hat{x}(k-1|k-1)+B_Pu(k-1)$,而非$\hat{x}(\kappa|\kappa-1)=A_P\hat{x}(\kappa-1|\kappa-1)+B_Pu(\kappa-1)$,即不存在$\hat{x}(k|k-1)=\hat{x}(\kappa|\kappa-1)$,所以残差检测器能够有效地检测出重放攻击.

2.3 虚假数据注入攻击

虚假数据注入攻击可以注入恶意数据以降低甚至恶化系统性能,并具有一定的隐蔽性,被认为是ICPS中最危险的网络攻击. 发生虚假数据注入攻击时,控制指令和测量信号可以表达为

$$\tilde{u}(k)=u(k)+\delta_a,\ \tilde{y}(k)=y(k)+\delta_s.$$

为使攻击具有一定的隐蔽性,攻击信号δ_a和δ_s的模型表达根据检测器的不同具有不同的数学形式.

为简单起见,只考虑传感器通道中的攻击. 在本文所用的传统检测方法下,为检测具备一定隐蔽性的攻击以反映该类攻击的特性,虚假数据注入攻击的传感器攻击信号δ_s的数学表达[16]为

$$\delta_s = -y(k)+C_PA_P\hat{x}(k-1)+C_PB_Pu(k-1)+\varsigma(k-1), \tag{35}$$

其中$\varsigma(k-1)$是协方差为$C_PP(k|k-1){C_P}^{\mathrm{T}}+R$的高斯白噪声. 于是,攻击后的传感器测量值为

$$\tilde{y}(k) = C_Px(k)+v(k)-y(k)+C_PA_P\hat{x}(k-1|k-1)+C_PB_Pu(k-1)+\varsigma(k-1). \tag{36}$$

残差$\tilde{r}(k)=\tilde{y}(k)-\hat{y}(k)$,结合式(8)得

$$\tilde{r}(k)=\varsigma(k-1), \tag{37}$$

则攻击后的残差$\tilde{r}(k)$的协方差矩阵为

$$\tilde{V} = \mathrm{E}[\varsigma(k-1)\,\varsigma^{\mathrm{T}}(k-1)] = C_PP(K|K-1){C_P}^{\mathrm{T}}+R. \tag{38}$$

由式(14),正常情况下残差的协方差矩阵也为

$$V=\mathrm{E}[r(k)\,r^{\mathrm{T}}(k)]=C_PP(k|k-1){C_P}^{\mathrm{T}}+R.$$

可以看出,在攻击信号δ_s作用下的残差都是协方差为$C_PP(k|k-1){C_P}^{\mathrm{T}}+R$的高斯白噪声,所以该虚假数据注入攻击能够绕过检测器的检测. 即当$a_k=[\delta_a^{\mathrm{T}}\ \ \delta_s^{\mathrm{T}}]^{\mathrm{T}}$,$\delta_s=-y(k)+C_PA_P\hat{x}(k-1)+C_PB_Pu(k-1)+\varsigma(k-1)$时,模型(17)描述虚假数据注入攻击. 此类攻击需要知道系统的相关模型知识,以便构成复杂的攻击策略绕过检测器的检测. 该类攻击具有一定的欺骗性,持续的虚假数据注入攻击会造成系统不稳定并且令系统难以诊断,往往会造成相比于Dos 攻击

更大的威胁.

3 仿真与结果

3.1 仿真准备

运用 Simulink/TrueTime 进行 ICPS 的攻击仿真,仿真中将网络数据传输率设为80 000 bits/s,丢包率设为0. 执行器节点以事件方式驱动,传感器节点以时间方式驱动.

假设 $A_P = \begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$, $B_P = \begin{bmatrix}1\\ 0.5\end{bmatrix}$, $C_P = [1 \quad 1]$, $D_P = 0$,其满足可控可观,通过控制器的调节使被控对象输出能跟踪设定值. 将被控对象的采样周期设为0.001 s,并将初态 $x(0)$ 设为0.

在不设置攻击节点的情况下,通过调节得到在控制器参数为

$$A_F = \begin{bmatrix}1 & 0.004\,5\\ 0 & 0.548\,8\end{bmatrix},\ B_F = \begin{bmatrix}0\\ 0.004\,5\end{bmatrix},$$

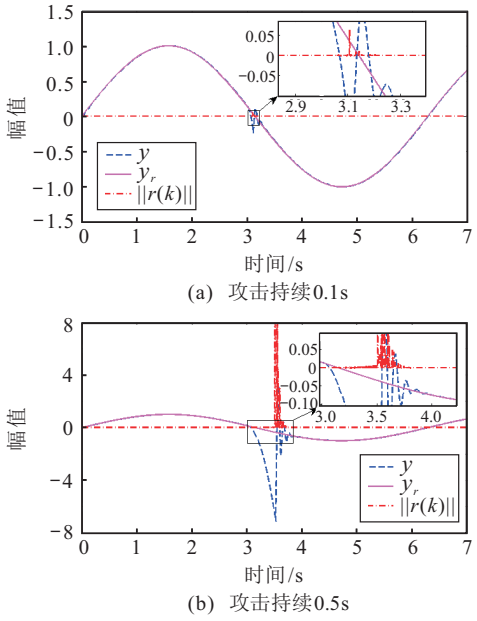
$$C_F = [0.000\,01 \quad -0.079\,999],\ D_F = 0.000\,9$$

时,传感器输出结果跟踪给定值效果良好,残差检测器测得的数值为 10^{-4} 级别. 通过多次实验得出,在检验 Dos 攻击和重放攻击时,残差检测器的阈值 σ 取 8.2×10^{-4} 时具有较好的检测效果.

3.2 攻击结果分析

3.2.1 Dos攻击特性仿真分析

当攻击策略为 $g(\varnothing,\varnothing)$ 且模型(17)中的攻击向量 $a_k = [0 \quad (-s_k^y(y(k)-y(k_a-1)))^{\mathrm{T}}]^{\mathrm{T}}$ 时,表示发生了Dos攻击,仿真结果如图4所示.



(a) 攻击持续0.1s

(b) 攻击持续0.5s

图 4 正弦给定下Dos攻击仿真曲线

从图4可以看出:在第3 s时发生了Dos攻击,使传感器测量信号偏离了给定值且偏离程度超过了5%,为有效攻击;同时,图4(b)的偏离程度大于图4(a)的偏离程度,说明攻击持续时间越长偏离程度越大;图4(a)和图4(b)中的残差检测器的数值远超 8.2×10^{-4},能明显检测出发生了攻击.

3.2.2 重放攻击特性仿真分析

当攻击策略为 $g(\varnothing,\iota_k)$ 且模型(17)中的攻击向量为 $a_k = [0 \quad (-\tau^y(y(k)-y(\kappa)))^{\mathrm{T}}]^{\mathrm{T}}$ 时,表示发生了重放攻击. 在第3 s时给予持续0.1 s的攻击,其仿真结果如图5所示.

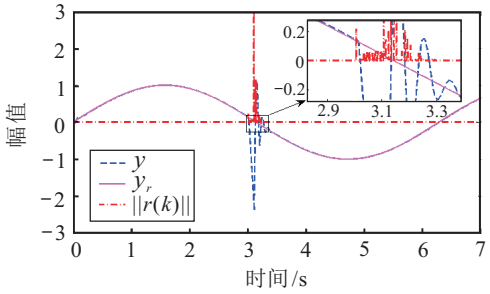
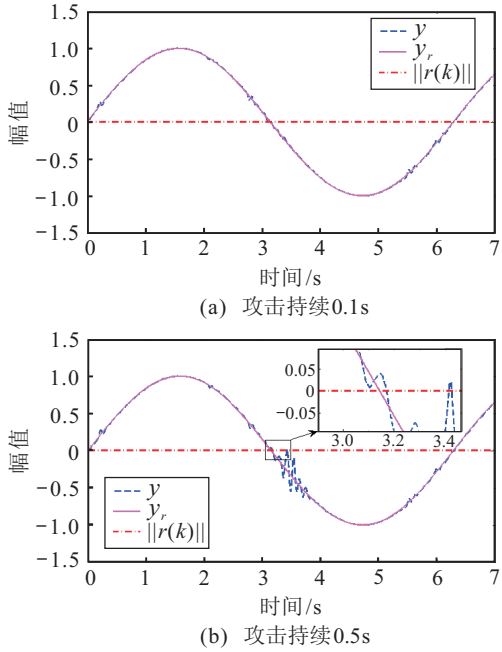


图 5 重放攻击仿真曲线

由图5可以看出,重放攻击相较于Dos攻击,会使真实值的偏离程度更大,并且残差检测器的检测效果也更明显. 因为相比于Dos攻击采用上一时刻的延迟数据,重放攻击用的是更远离于此时刻的过去信息,所以真实数据流失量会更大.

3.2.3 虚假数据注入攻击特性仿真分析

当图3中的攻击策略为 $g(K,\iota_k)$ 且模型(17)中的攻击向量为 $a_k = [0 \quad (-y(k)+C_PA_P\hat{x}(k-1)+C_PB_Pu(k-1)+\varsigma(k-1))^{\mathrm{T}}]^{\mathrm{T}}$ 时,发生了虚假数据注入攻击. 在第3 s分别给予持续0.1 s和0.5 s的攻击信号,仿真结果如图6所示.



(a) 攻击持续0.1s

(b) 攻击持续0.5s

图 6 虚假数据注入攻击仿真曲线

从图6可以看出:在攻击同样持续0.1 s的情况下,整个系统就像没发生攻击一样;而当攻击持续时间加长至0.5 s时才引起测量信号的明显偏离. 这是因为相比于其他类型的攻击,虚假数据注入攻击在原信号的基础上变化非常细微,短时间内的变化不足以引起测量值偏离;而当这种差值慢慢积累,才会在图中显现出来,如图6(b)中长时间的虚假数据注入攻击使测量值偏离了给定值,使系统稳定性受到破坏. 而此时,残差检测器的示数也无明显变化,远低于所给的阈值. 因为虚假数据注入攻击需要知道系统模型知识、传感器的历史数据,以便构成更复杂的攻击绕过防御系统,所以此类攻击更具隐蔽性.

4 结　论

本文根据攻击者所掌握的资源构造通用攻击模型以研究ICPS的网络攻击问题. 利用所提出的攻击模型,研究了Dos攻击、重放攻击、虚假数据注入攻击等攻击策略的不同数学形式,并对相应攻击的性能进行了分析. 分析表明:不同攻击会造成不同的破坏资源,无需披露资源及模型知识的Dos攻击所造成的破坏更像是通信延迟;重放攻击由于采用更远时刻的历史数据作为披露资源进行回放,造成的数据流失量更大;而同时具备模型知识和披露资源的虚假数据注入攻击造成数据破坏的同时又有一定的隐蔽性. 采用Simulink/TrueTime对3种攻击的特性进行了仿真,仿真结果表明本文的攻击者模型能有效地模拟这3种攻击特性. 但本文的研究是假设系统模型(6)的$I - D_F D_P$是可逆的. 对于$I - D_F D_P$不可逆的情况,将作为下一步的研究内容.

参考文献(References)

[1] Wang D, Wang Z, Shen B, et al. Recent advances on filtering and control for cyber-physical systems under security and resource constraints[J]. Journal of the Franklin Institute, 2016, 353(11): 2451-2466.

[2] An L, Yang G H, An L, et al. Improved adaptive resilient control against sensor and actuator attacks[J]. Information Sciences, 2017, 423(C): 145-156.

[3] Cardenas A A, Roosta T, Sastry S. Rethinking security properties, threat models, and the design space in sensor networks: A case study in SCADA systems[J]. Ad Hoc Networks, 2009, 7(8): 1434-1447.

[4] Finogeev A G, Finogeev A A. Information attacks and security in wireless sensor networks of industrial SCADA systems[J]. Journal of Industrial Information Integration, 2017, 5(4): 6-16.

[5] Ding D, Han Q L, Xiang Y, et al. A survey on security control and attack detection for industrial cyber-physical systems[J]. Neurocomputing, 2017, 275(10): 1674-1683.

[6] 季承扬. 信息物理系统安全威胁与防护措施[J]. 科技传播, 2018, 10(4): 111-112.
(Ji C Y. Cyber physical system security threats and protection measures[J]. Science and Technology Communication, 2018, 10(4): 111-112.)

[7] Yuan Y, Yuan H, Guo L, et al. Resilient control of networked control system under DoS attacks: A unified game approach[J]. IEEE Transactions on Industrial Informatics, 2016, 12(5):1786-1794.

[8] Kosut O, Jia L, Thomas R J, et al. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures[C]. IEEE International Conference on Smart Grid Communications. Gaithersburg: IEEE, 2010: 220-225.

[9] Hoehn A, Zhang P. Detection of replay attacks in cyber-physical systems[C]. American Control Conference. Boston: IEEE, 2016: 290-295.

[10] Amin S, Cárdenas A A, Sastry S S. Safe and secure networked control systems under denial-of-service attacks[C]. International Conference on Hybrid Systems: Computation and Control. San Francisco: Springer-Verlag, 2009: 31-45.

[11] Mo Y, Sinopoli B. Secure control against replay attacks[C]. Annual Allerton Conference on Communication, Control, and Computing. Monticello: IEEE, 2009: 911-918.

[12] Chattopadhyay A, Mitra U. Security against false data injection attack in cyber-physical systems[C]. Annual Conference on Information Sciences and Systems. Princeton: Princeton University, 2018: 1-6.

[13] Shames I, Sandberg H, Johansson K H. A secure control framework for resource-limited adversaries[J]. Automatica, 2015, 51(C): 135-148.

[14] 党鑫. 网络攻击环境下的无线网络控制系统设计[D]. 无锡: 江南大学物联网工程学院, 2015.
(Dang X. Design of wireless network control system under network attack environment[D]. Wuxi: School of Internet of Things Engineering, Jiangnan University, 2015.)

[15] Fang Z H, Mo H D, Wang Y. Reliability analysis of cyber-physical systems considering cyber-attacks[C]. 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). Singapore: IEEE, 2017: 364-368.

[16] 杨儒航. 网络化控制系统虚假数据注入攻击的检测方法研究[D]. 北京: 北方工业大学电气与控制工程学院, 2017.
(Yang R H. Research on detection method of false data injection attack in networked control system[D]. Beijing: School of Electrical and Control Engineering, North China University of Technology, 2017.)

作者简介

孙子文(1968−), 女, 教授, 博士, 从事控制理论与控制工程、模式识别、无线传感网络理论与技术等研究, E-mail: sunziwen@jiangnan.edu.cn;

张炎棋(1994−), 男, 硕士生, 从事控制理论与控制工程的研究, E-mail: yanqi.1001@foxmail.com.

(责任编辑：李君玲)