控制与决策
Control and Decision

工业信息物理系统安全风险动态表现分析量化评估模型

孙子文，张书国

引用本文:
孙子文, 张书国. 工业信息物理系统安全风险动态表现分析量化评估模型[J]. 控制与决策, 2021, 36(8): 1939−1946.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2019.1479

工业信息物理系统安全风险动态表现分析量化评估模型

孙子文[1,2†], 张书国[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: 针对当前工业信息物理系统的安全风险评估模型极少考虑系统的动态进程对评估准确性的影响,给出一种工业信息物理系统安全风险动态表现分析量化评估模型. 首先,运用贝叶斯网络对攻击在网络层的入侵过程建模,计算网络攻击成功入侵的概率;然后,在攻击成功入侵的前提下,采用卡尔曼状态观测器实时观测被控对象的状态,研究系统的动态表现,定量分析系统的表现损失,从经济损失的角度量化攻击对系统造成的影响,并结合攻击成功入侵的概率,实现对系统安全风险的动态评估. 最后,通过Matlab对攻击下沸水发电厂模型的运行状态进行仿真,结果表明所提模型能有效地评估工业信息物理系统的风险.

关键词: 工业信息物理系统；风险评估；动态分析；贝叶斯网络；网络攻击；攻击影响

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2019.1479 **开放科学(资源服务)标识码(OSID):**

引用格式: 孙子文,张书国. 工业信息物理系统安全风险动态表现分析量化评估模型[J]. 控制与决策, 2021, 36(8): 1939-1946.

Quantitative evaluation model for dynamic performance analysis of security risk in industrial cyber physics systems

SUN Zi-wen[1,2†], *ZHANG Shu-guo*[1]

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

Abstract: In view of the fact that current safety risk assessment models for industrial cyber physical systems (ICPS) rarely consider the impact of dynamic process of the system on the accuracy of the assessment, this paper proposes a quantitative evaluation model for the dynamic performance analysis of security risk in the ICPS. Firstly, the Bayesian network is used to model the intrusion process of the attack in the cyber layer, and the probability of the successful intrusion of the network attack is calculated. Then, under the premise of successful attack, the Kalman state observer is used to observe the state of the controlled object in real time, the dynamic performance of the system is studied, the performance loss of the system is quantitatively analyzed, the impact of the attack on the system from the perspective of economic loss is quantified, and the dynamic assessment of system security risk based on the probability of successful attack is realized. Finally, the running state of the boiling water power plant model under attack is simulated using Matlab. The results show that the model can effectively assess the risk of ICPS.

Keywords: industrial cyber physical systems；risk assessment；dynamic analysis；Bayesian network；network attacks；attack influence

0 引 言

信息物理系统(cyber physical systems, CPS)是一种集网络、通信和自动化控制等技术为一体的网络控制系统[1],被广泛应用于化学、自动化控制、智能电网等工业领域,形成工业信息物理系统(industrial cyber-physical systems, ICPS). 由于ICPS网络层和物理层的紧密耦合以及所处工业环境的复杂多变,使得其极易受到来自物理世界和网络世界的攻击[2],其中网络攻击最为频繁、复杂且更新快,这使得如何精准高效地防御网络攻击成为ICPS的重点研究方向之一. 风险评估量化网络攻击对ICPS造成的风险,得到不同的风险值,为决策者采取精准有效的防御措施提供决策依据[3].

目前,一些关于ICPS的风险评估方法被提出,但大部分集中在IT领域. 文献[4]提出一种基于攻击图的安全风险评估方法,通过分析漏洞的利用模式及影

收稿日期: 2019-10-22；**修回日期:** 2020-04-25.

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).

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

响后果,估计攻击成功概率和攻击影响,建立风险评估模型;文献[5]将BNs、故障树和事件树组成了一个多模型框架,以计算危险事件发生的概率,并评估事件造成的经济损失;文献[6]提出一种基于贝叶斯攻击图的风险评估模型,利用属性攻击图得到静态安全风险,进而运用贝叶斯推理方法更新攻击图中各节点的先验概率,求出攻击成功发生的概率. 上述文献将评估的重心放在了网络层[4-6],而忽略了网络攻击对物理层的影响.

由于ICPS网络层和物理层的紧密耦合,不能单从网络层或者物理层来研究攻击作用. 文献[7]引入弹性控制中攻击发生后系统的可操作时间(time to shut down, TTSD)来量化网络攻击对物理层的影响,并结合攻击成功发生的概率综合评估ICPS的风险值;之后,文献[8]采用随机混合模型对物理层建模,分析物理层在攻击作用下的动态过程,并在TTSD基础上提出攻击达到目的平均所需要的时间(mean time to shut down, MTTSD)来量化ICPS的风险;文献[9]采用一个随机博弈模型求得攻击成功的概率,并结合攻击的MTTSD和系统的可利用性,综合评估网络攻击对ICPS造成的风险. 文献[7-9]都未深入分析系统在网络攻击下的动态过程,且用MTTSD和TTSD来量化攻击对系统的影响过于简单,不能很好地反映出攻击造成的影响.

本文采用贝叶斯网络对攻击在网络层的入侵过程建模,并深入研究系统在网络攻击下的表现变化,给出一种工业信息物理系统安全风险动态表现分析量化评估模型. 首先分析ICPS攻击产生破坏作用的过程,将该过程分为入侵和破坏两个阶段,入侵阶段采用贝叶斯网络对攻击在网络层的入侵过程建模,求出传感器和执行器被成功攻击的概率;破坏阶段采用卡尔曼状态观测器实时监测系统的进程,并对系统在攻击下的动态表现进行定量分析,从表现损失和维护费用两方面量化攻击对系统造成的影响. 然后,综合攻击成功入侵的概率和对系统造成的影响,动态评估ICPS的风险. 最后,采用沸水发电厂模型来验证本文给出的风险评估方法的有效性.

1 工业信息物理系统的安全

1.1 工业信息物理系统的结构

如图1[10]所示,传统的工业信息物理系统主要由网络层和物理层两部分组成. 网络层包括企业网络、隔离区和控制网络:企业网络包含工作站和应用服务器,负责事务管理和用户交互;隔离区是两个防火墙之间的网络区域,主要包含数据库服务器和历史数据,用于存放ICPS中的数据;控制网络由PLCs、控制服务器和人机交互界面组成,负责控制被控对象的正常运行和人机交互,其中控制服务器又包含异常检测器,负责检测网络攻击的发生. 物理层包括执行器、被控对象和传感器. 网络层与物理层通过交流通道进行数据交换,ICPS中有两条主要的交流通道:第1条交流通道是传感器到控制器,用于将传感器的测量值发送给控制器;第2条交流通道是控制器到执行器,用于将控制器发出的控制信号发送给执行器.

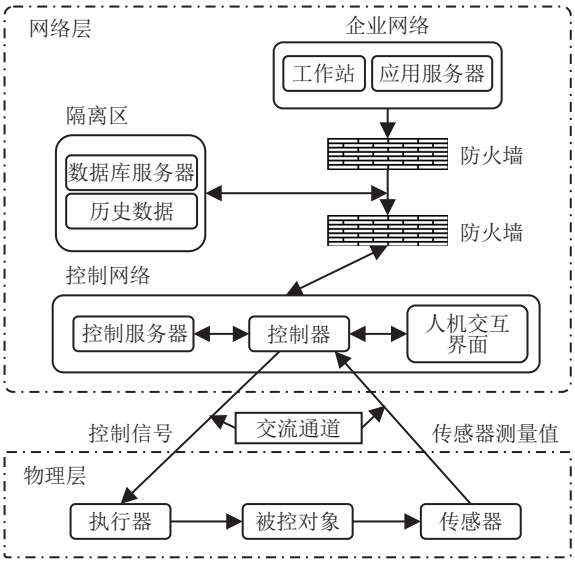

图1 工业信息物理系统结构

1.2 工业信息物理系统中的攻击

ICPS的网络攻击具有与工业网络系统(industrial cyber systems, ICS)网络攻击不同的攻击目的和方法. ICS中网络攻击的目的是破坏信息的完整性、保密性和可利用性,因此只需利用系统的网络漏洞获取操作权限就能达到破坏目的. 不同于ICS中的攻击,ICPS中攻击的目的除了破坏信息的完整性、保密性和可利用性,更严重的是破坏系统的物理进程,从而对生产过程产生影响. 除了掌握网络漏洞和获取操作权限外,还必须了解系统的动态进程、故障条件以及控制算法等,缺乏这些资源,网络攻击则很难破坏系统的物理进程. 因此,将ICPS中攻击产生破坏作用的过程分为入侵和破坏两个阶段:攻击的入侵阶段类似于ICS中攻击的作用,利用系统的漏洞获取所需要的权限,为破坏阶段做准备;在攻击的破坏阶段,攻击者获取系统的动态进程和控制算法等资源,根据发生故障的条件破坏系统的物理进程.

ICPS中的攻击主要是通过掌控物理层的执行器和传感器来破坏系统的物理进程. 当传感器被攻击者掌控时,传感器向控制器发送错误的传感器测量值,控制器就会产生错误的控制信号,影响系统进

程. 当执行器被攻击者掌控时,直接导致执行器执行错误的动作,破坏系统进程.

2　安全风险动态表现分析量化评估模型

建立ICPS安全风险动态表现分析量化评估模型,如图2所示. 基于ICPS中网络攻击的两个阶段,将该评估模型分为网络攻击成功入侵概率的估算和攻击影响的量化两部分. 网络攻击成功入侵概率的估算采用贝叶斯网络对攻击在网络层的入侵过程建模,根据漏洞和权限被利用的概率,求出网络攻击成功入侵的先验概率,进而根据实际攻击事件,运用贝叶斯定理估算网络攻击成功入侵的后验概率;攻击影响的量化通过实时观测被控对象在攻击下的运行状态,动态分析系统的表现,根据表现损失和设备维护花费量化攻击产生的影响. 最后计算网络攻击成功入侵概率与攻击造成影响的乘积,对ICPS的安全风险评估值进行动态更新.

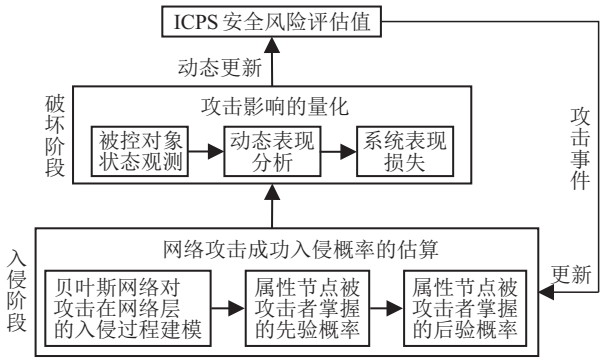

图 2　安全风险动态表现分析量化评估模型

2.1　网络攻击成功入侵概率的估算

2.1.1　网络层的攻击入侵贝叶斯网络模型

贝叶斯网络不仅能够推理出网络状态间的依赖关系,而且可以估算网络攻击发生的概率,所以常被用于安全风险评估模型中. 假设攻击者已知系统的网络结构,并能根据攻击目的设计最佳的攻击路径. 采用贝叶斯网络对攻击在网络层的入侵过程进行建模,形成攻击入侵贝叶斯网络模型. 根据网络攻击对系统资源的利用,将攻击入侵贝叶斯网络模型的属性节点分为3类:权限节点PRI、漏洞节点VUL和目标节点TAR. PRI表示攻击者想要进行攻击操作必须获取的权限,VUL表示可以被攻击者利用的网络系统中的漏洞,TAR表示攻击者拟攻击的传感器或执行器. 属性表示攻击者掌握的系统资源,包括系统漏洞、访问权限以及传感器和执行器. 当目标节点被攻击者掌握时,即表示网络攻击入侵成功.

贝叶斯网络是一个有向无环图,采用五元组 $\mathrm{BN}=\{S,E,R,P_1,P_2\}$ 表示,各元素信息如下:

1) $S=\{S_i^G|i=1,2,\ldots,n,G=\mathrm{PRI},\mathrm{VUL},\mathrm{TAR}\}$ 为属性节点. G 表示属性节点的类型,取值为PRI或VUL或TAR. n 表示贝叶斯网络中属性节点的总数量. S_i^G 取值为0或1,当 $S_i^G=0$ 时,表示该属性节点未被攻击者掌控;当 $S_i^G=1$ 时,表示该属性节点被攻击者掌控.

2) $E=\{e_{ij}|i=1,2,\ldots,n,j=1,2,\ldots,n\}$ 表示贝叶斯网络有向边的集合,e_{ij} 表示从父节点 S_i^G 向子节点 S_j^G 的转移,节点 S_j^G 的父节点集合为 $\mathrm{Fa}(S_j^G)=\{S_i^G|e_{ij}\in E\}$.

3) R 表示属性节点与其父节点之间的关系,父子节点之间的依赖关系可以分为AND或者OR两种. AND表示子节点的所有父节点都必须同时被攻击者掌握才会发生向子节点的转移;OR表示只要子节点的其中一个父节点被攻击者掌握,就会发生向子节点的转移.

4) P_1 表示属性节点被攻击者掌握的先验概率.

5) P_2 表示属性节点被攻击者掌握的后验概率.

图3是攻击入侵贝叶斯网络模型结构示意图. 属性节点 S_1^{VUL}、S_2^{VUL}、S_3^{VUL}、S_4^{VUL} 表示系统中4个可以被攻击者利用的漏洞,其中节点 S_4^{VUL} 的父节点集合为 $\mathrm{Fa}(S_4^{\mathrm{VUL}})=\{S_1^{\mathrm{VUL}},S_2^{\mathrm{VUL}}\}$,它与父节点的关系为AND. S_5^{PRI} 表示攻击者必须获取的权限,它与父节点的关系为OR. S_6^{TAR} 表示攻击者想要攻击的目标节点——传感器或者执行器,它与父节点的关系为AND. 网络攻击的目的是掌握目标节点 S_6^{TAR},攻击者可以选择不同的入侵路线,如路线 e_{14} 至 e_{46} 或 e_{15} 至 e_{56}. 根据利用漏洞和获取权限难度的不同,可以估算出节点 S_6^{TAR} 被攻击者掌握的概率,也即网络攻击成功入侵的概率.

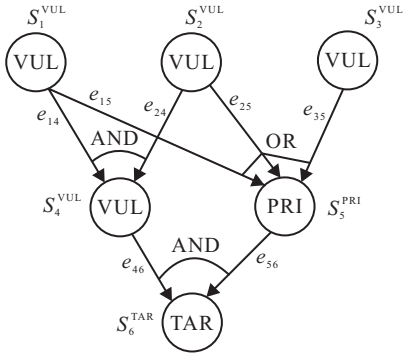

图 3　攻击入侵贝叶斯网络模型结构示意

2.1.2　先验概率的计算

1) 漏洞利用成功的概率. 漏洞利用成功的概率与该漏洞的可利用性(exploitability)有关,采用公共漏洞评分体系(common vulnerability scoring system, CVSS)[11]评估漏洞的可利用性. CVSS由3组

矩阵构成: base、temporal和environmental. base表示漏洞固有的属性, temporal表示漏洞随时间变化的特性, environmental表示漏洞所处的网络环境属性. 漏洞的可利用性只与base有关, CVSS的base有3个子项: 访问向量(access vector, AV)、访问复杂度(access complexity, AC)、认证(authentiation, Au), 它们的评分范围设置为$0 \sim 1$. 根据3个子项的评分, 计算漏洞利用成功的概率[12]为

$$P(S_j^{\mathrm{VUL}}) = 2 \cdot \mathrm{AV} \cdot \mathrm{AC} \cdot \mathrm{Au}, \tag{1}$$

其中S_j^{VUL}表示贝叶斯网络的漏洞节点.

2) 局部条件概率. 局部条件概率反映了某个属性节点被攻击者利用的可能性, 与该属性节点与其父节点的依赖关系有关. 根据父子节点的依赖关系的不同, 局部条件概率的计算分为两种情况:

① 当父子节点的依赖关系为AND时:

$$P(S_j^G|\mathrm{Fa}(S_j^G)) = \begin{cases} 0,\ \exists S_i^G \in \mathrm{Fa}(S_j^G)|S_i^G = 0; \\ \coprod\limits_{S_i^G=1} P(S_i^G),\ \text{otherwise}. \end{cases} \tag{2}$$

② 当父子节点的依赖关系为OR时:

$$P(S_j^G|\mathrm{Fa}(S_j^G)) = \begin{cases} 0,\ \forall S_i^G \in \mathrm{Fa}(S_j^G)|S_i^G = 0; \\ 1 - \coprod\limits_{S_i^G=1} (1 - P(S_i^G)),\ \text{otherwise}. \end{cases} \tag{3}$$

3) 先验概率. 在贝叶斯网络建立后, 根据漏洞利用成功的概率与局部条件概率推算出各节点的先验概率, 每个属性节点的先验概率等于当前节点与其父节点的条件概率之积, 即

$$P_1(S_j^G) = \coprod_{j=1}^{n} P(S_j^G|\mathrm{Fa}(S_j^G)), \tag{4}$$

其中$P_1(S_j^G)$表示节点S_j^G被攻击者掌握的先验概率.

2.1.3 后验概率的计算

根据入侵检测系统, 可以观测到攻击事件的集合$O = \{O_{S_j^G}|j = 1, 2, \ldots, n\}$, 其中$O_{S_j^G}$表示第$j$个属性节点$S_j^G$已经被攻击者掌握. 利用贝叶斯定理来更新节点的先验概率, 有

$$P_2(S_j^G) = P(S_j^G|O) = \frac{P(O|S_j^G) \cdot P_1(S_j^G)}{P_1(O)}. \tag{5}$$

其中: $P(S_j^G|O)$表示在给定攻击事件集合O的前提下, 属性节点S_j^G被攻击者掌握的条件概率; $P(O|S_j^G)$表示在属性节点S_j^G被攻击者掌握的前提下, 攻击事件集合O发生的条件概率; $P_1(S_j^G)$表示属性节点S_j^G被攻击者掌握的先验概率; $P_1(O)$表示攻击事件集合O发生的先验概率.

2.2 攻击影响的量化

2.2.1 被控对象模型

被控对象采用离散的状态空间(linear-time-invariant, LTI)模型:

$$\begin{cases} x(k+1) = Ax(k) + Bu(k) + w(k), \\ y(k) = Cx(k) + Du(k) + v(k). \end{cases} \tag{6}$$

其中: k表示时刻; $x(k)$、$u(k)$和$y(k)$分别表示被控对象的状态变量、控制信号和传感器的测量信号; A、B、C、D分别表示状态转移矩阵、输入矩阵、输出矩阵和直接传递矩阵; $w(k)$和$v(k)$分别表示对象的过程噪声和测量噪声.

2.2.2 状态观测

假设被控对象的状态是可观测的, 采用卡尔曼状态估计器[13]实时监测被控对象在攻击下的运行状态.

根据被控对象模型设计卡尔曼状态估计器:

$$\begin{cases} \hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1), \\ \hat{y}(k) = C\hat{x}(k|k-1). \end{cases} \tag{7}$$

其中: $\hat{x}(k|k-1)$是对k时刻对象状态的估计值, $\hat{x}(k-1|k-1)$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的估计值.

假设$w(k)$和$v(k)$分别是协方差为Q和R的高斯白噪声, 得到如下的卡尔曼状态更新[13]:

$$\hat{x}(k|k) = A\hat{x}(k-1|k-1) + Bu(k-1) + K(k)[y(k) - \hat{y}(k)]. \tag{8}$$

其中: $\hat{x}(k|k)$为k时刻状态最优值; $K(k)$是卡尔曼增益矩阵, 表示为

$$K(k) = P(k|k-1)C^{\mathrm{T}}[CP(k|k-1)C^{\mathrm{T}} + R]^{-1}. \tag{9}$$

这里

$$P(k|k-1) = AP(k-1|k-1)A^{\mathrm{T}} + Q \tag{10}$$

为k时刻对象状态估计值的协方差;

$$P(k|k) = [I - K(k)C]P(k|k-1) \tag{11}$$

为k时刻最优估计值的协方差.

由式(8)得到k时刻对象的最优估计值$\hat{x}(k|k)$后, 根据式(7)推出对象的估计输出$\hat{y}(k)$, 则传感器测量值的残差可表示为

$$r(k) = \tilde{y}(k) - \hat{y}(k), \tag{12}$$

其中$\tilde{y}(k)$为k时刻状态观测器接收到的传感器测量

值.

根据残差$r(k)$的大小,判断系统是否存在攻击,判断规则为

$$\begin{cases}\text{if } \|r(k)\| < \sigma \text{ then } H_0,\\ \text{if } \|r(k)\| > \sigma \text{ then } H_1.\end{cases} \tag{13}$$

其中:σ表示阈值,H_0表示攻击未发生,H_1表示攻击发生. σ的选取与系统要求的虚警率和漏警率有关,选定合适的阈值是保证系统检测性能的前提.

2.2.3　系统动态表现的分析

基于卡尔曼状态观测器,实时监测被控对象的运行状态,进而评估整个系统的动态表现. 用一个关于产出和质量的效益函数来表示系统的动态表现[14]:

$$\mathrm{PE}_i(t) = f(\mathrm{Pro}(t), \mathrm{Qua}(t)) = \alpha \mathrm{Pro}(t) \cdot \mathrm{Qua}(t). \tag{14}$$

其中:$\mathrm{PE}_i(t)$表示系统在t时刻攻击i下的动态表现,α表示系统效益的转化率,$\mathrm{Pro}(t)$表示产出,$\mathrm{Qua}(t)$表示质量. 产出$\mathrm{Pro}(t)$通常用t时刻系统的产量来表示,质量$\mathrm{Qua}(t)$通常指t时刻产出产品的质量. 效益转化率、产出以及质量都会直接对系统的动态表现产生影响.

图4[14]为系统在攻击下的动态表现. t_i^0表示攻击i发生的时刻,t_i^d表示系统表现开始下降的时刻,t_i^m表示系统的表现下降到最低的时刻,t_i^i表示攻击被系统检测到的时刻,t_i^s表示系统表现开始恢复的时刻,t_i^r表示系统完全恢复到正常表现的时刻,有$t_i^d < t_i^i < t_i^s$. T_c表示系统工作周期. PE_0表示系统正常时的表现;PE_i表示系统在攻击i下的最低表现,当$\mathrm{PE}_i = 0$时,系统进程中断.

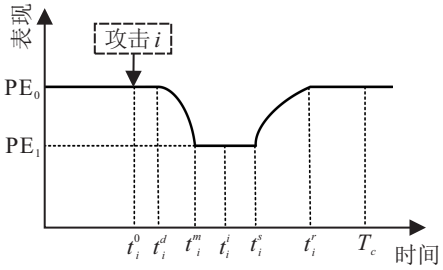

图 4　系统在攻击下的动态表现

描述系统表现动态变化的几个参数如下:

1) 攻击检测时间T_i^i:系统检测到攻击存在所需要的时间.

$$T_i^i = t_i^i - t_i^0. \tag{15}$$

T_i^i的大小间接地反映了系统异常检测的能力,T_i^i越小,观测器检测到异常所需时间越短,系统也就能尽快地采取防御措施.

2) 系统工作时长T_i^w:一个工作周期内,系统能够工作的时间.

$$T_i^w = \begin{cases} T_c, \ \mathrm{PE}_i \neq 0;\\ t_i^m, \ \mathrm{PE}_i = 0.\end{cases} \tag{16}$$

当$\mathrm{PE}_i \neq 0$时,系统进程未中断,故系统的工作时长即为一个工作周期的时长T_c;当$\mathrm{PE}_i = 0$时,攻击i使系统进程中断,系统无产出,故系统的工作时长为系统表现降为0所需要的时间,即t_i^m.

3) 表现损失时间T_i^{pl}:系统的表现低于正常表现的时间.

$$T_i^{\mathrm{pl}} = t_i^r - t_i^d. \tag{17}$$

4) 表现损失PL_i:攻击导致系统的表现下降造成的经济损失.

$$\mathrm{PL}_i = \begin{cases} \mathrm{PE}_0 \cdot T_i^{\mathrm{pl}} - \int_{t_i^d}^{t_i^r} \mathrm{PE}_i(t)\mathrm{d}t, \ \mathrm{PE}_i \neq 0;\\ \mathrm{PE}_0 \cdot T_c - \int_0^{t_i^m} \mathrm{PE}_i(t)\mathrm{d}t, \ \mathrm{PE}_i = 0.\end{cases} \tag{18}$$

当$\mathrm{PE}_i \neq 0$时,表现损失PL_i可用系统无攻击下在T_i^{pl}时间内的表现$\mathrm{PE}_0 \cdot T_i^{\mathrm{pl}}$,与系统有攻击下在$T_i^{\mathrm{pl}}$时间内的表现$\int_{t_i^d}^{t_i^r} \mathrm{PE}_i(t)\mathrm{d}t$的差值来表示;当$\mathrm{PE}_i = 0$时,表现损失$\mathrm{PL}_i$可用系统无攻击下在$T_i^w$时间内的表现$\mathrm{PE}_0 \cdot T_c$,与系统有攻击下在$T_i^w$时间内的表现$\int_0^{t_i^m} \mathrm{PE}_i(t)\mathrm{d}t$的差值来表示.

5) 系统总损失L_i:攻击i造成的总的经济损失,可以用关于表现损失PL_i和设备维护花费R_i^c的方程来表示,即

$$L_i = f(\mathrm{PL}_i, R_i^c) = \beta \mathrm{PL}_i + \lambda R_i^c, \tag{19}$$

其中β、λ分别表示表现损失和设备维护花费对系统造成损失的权重.

2.3　ICPS风险的量化

攻击者的目的是掌握属性节点S_j^{TAR},进而破坏系统的进程. 节点S_j^{TAR}被攻击者i掌握时的安全风险值R_{ij}可表示为

$$R_{ij} = P_2(S_j^{\mathrm{TAR}}) \cdot L_i. \tag{20}$$

其中:$P_2(S_j^{\mathrm{TAR}})$表示属性节点S_j^{TAR}被攻击者掌握的概率,L_i表示攻击i造成的总的经济损失.

3　案例研究

3.1　系统的描述

采用一个沸水发电厂模型(boiling water power plant, BWPP)[15]来验证本文给出风险评估方法的可行性. 如图5所示,在BWPP模型中有3种传感器和阀门:压力传感器、液位传感器和电量传感器;排汽阀,由蒸汽排放控制PLC控制;进水阀,由水位控制PLC控制;进料阀,由进料控制PLC控制.

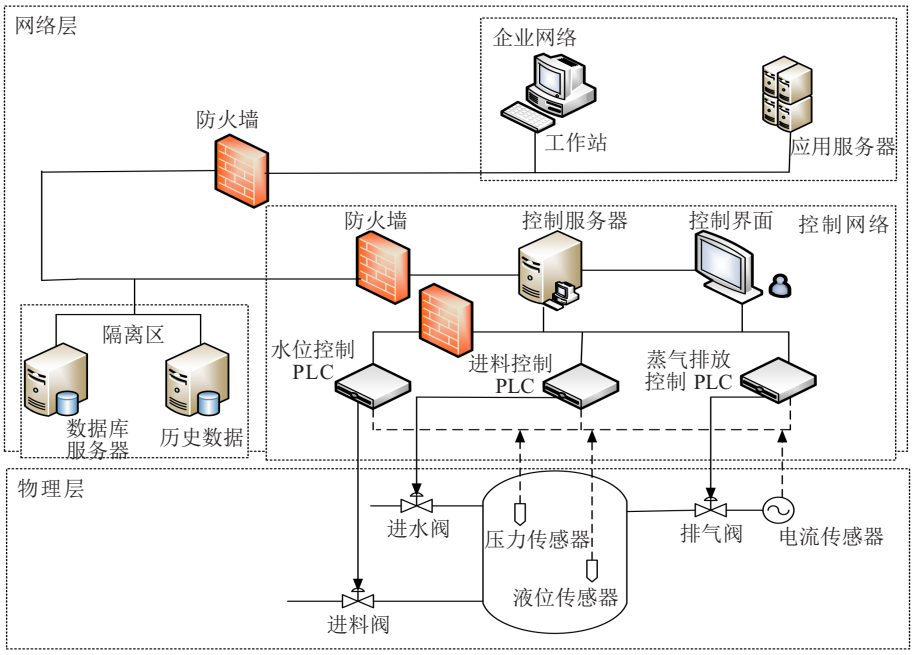

图 5 BWPP工业模型

在BWPP模型中,主要控制系统的3种状态:罐内压力$x_1/(\mathrm{kg/cm^2})$、电量输出x_2/mw、流体密度$x_3/(\mathrm{kg/cm^3})$;u_1、u_2、u_3分别表示进料、排汽和进水阀的位置,且有$0<u_i<1(i=1,2,3)$;y_1、y_2、y_3分别表示压力传感器、电量和液位传感器的读数. 被控对象的数学模型[15]如下:

$$A=\begin{bmatrix}0.999\,8 & 0 & 0\\ 0.006\,9 & 0.990\,0 & 0\\ -0.000\,7 & 0 & 1\end{bmatrix},$$

$$B=\begin{bmatrix}0.09 & -0.034\,9 & -0.015\\ 0.000\,3 & 1.048\,3 & -0.000\,1\\ 0 & -0.139\,8 & 0.165\,9\end{bmatrix},$$

$$C=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0.006\,3 & 0 & 0.004\,7\end{bmatrix}.$$

在BWPP模型中,产出$\mathrm{Pro}(t)$可以用电网传输电量的多少来表示;质量$\mathrm{Qua}(t)$可以用一个关于谐波、频率的方程来表示. 当罐内压力$x_1>250\,\mathrm{kg/cm^2}$时,储料罐就会发生爆炸,此时系统进程中断,系统最低表现$\mathrm{PE}_i=0$,设备维护费用$R_i^c=1$(万元);当罐内压力$x_1<250\,\mathrm{kg/cm^2}$时,储料罐不会爆炸,$\mathrm{PE}_i\neq 0, R_i^c=0$. 由于状态观测器的存在,系统能在较短的时间内检测到攻击的存在,相对于系统表现损坏时间T_i^{pl}和工作时间T_i^w,检测时间T_i^i可以忽略不计,故假设$t_i^0\approx t_i^d\approx t_i^i\approx 0$. 设系统正常运行时的表现为$\mathrm{PE}_0=10\,(\mathrm{RMB/s})$,工作周期$T_c=1\,000\,\mathrm{s}$,取效益转化率$\alpha=1$,攻击$i$下的动态表现为$\mathrm{PE}_i(t)=-\dfrac{\Delta\mathrm{PE}_i}{T_i^w}t+\mathrm{PE}_0$,其中$\Delta\mathrm{PE}_i=(\mathrm{PE}_0-\mathrm{PE}_i)$. 这里取$\beta=\lambda=1$,系统总的经济损失$L_i=f(\mathrm{PL}_i,R_i^c)=\mathrm{PL}_i+R_i^c$.

3.2 评估方法的实施

3.2.1 攻击成功入侵概率的计算

表1总结了BWPP模型中常见的漏洞,漏洞的可利用性通过查询网络公共漏洞数据库得到. 采用贝叶斯网络对攻击在网络层的入侵过程进行建模,如图6所示. 图6中各节点的含义在表2列出,目标节点与其父节点的关系都是AND.

表 1 常见漏洞的信息

漏洞	CVE编号	可利用性/%	描述
VUL_1	CVE-2015-1635	100	允许在Web Server上远程执行任意代码
VUL_2	CVE-2004-0840	100	允许在Mail Server上远程执行任意代码
VUL_3	CVE-2015-7417	68	跨站点Web脚本注入
VUL_4	CVE-2013-3957	100	使得远程攻击者能够通过未指定的向量执行任意SQL命令
VUL_5	CVE-2017-2681	65	影响SIMATIC HMI多面板和HMI移动面板以及S7-300/S7-400设备

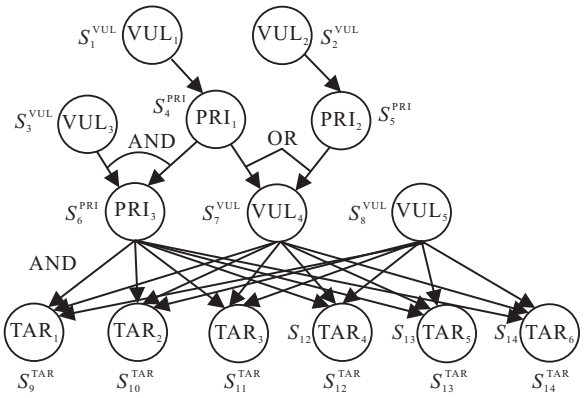

图 6　BWPP中攻击入侵贝叶斯网络模型结构

表 2　贝叶斯网络中各节点的含义

属性节点	概率/%	描述
S_4^{PRI}	60	获得Web Server执行任意代码的权限
S_5^{PRI}	60	获得Mail Server执行任意代码的权限
S_6^{PRI}	86	获得Data Server执行任意代码的权限
S_9^{TAR}	100	进料阀
S_{10}^{TAR}	100	排汽阀
S_{11}^{TAR}	100	进水阀
S_{12}^{TAR}	100	压力传感器
S_{13}^{TAR}	100	电量传感器
S_{14}^{TAR}	100	液体密度传感器

根据父子节点间的关系,利用式(2)和(3)求出局部条件概率. 然后,根据式(4)计算各节点被攻击者掌握的先验概率. 在给定攻击事件集合$O=\{O_{S_6^{\mathrm{PRI}}}\}$的情况下,利用式(5)更新各节点被攻击者掌握的先验概率. 各节点的先验概率和后验概率在表3中列出.

表3　各节点被攻击者掌握的先验概率与后验概率

属性节点	先验概率/%	后验概率/%	属性节点	先验概率/%	后验概率/%
S_1^{VUL}	100	100	S_6^{PRI}	40.8	48.5
S_2^{VUL}	100	100	S_7^{VUL}	84	100
S_3^{VUL}	68	68	S_8^{VUL}	65	65
S_4^{PRI}	60	71.4	S_i^{TAR}	22.3	31.5
S_5^{PRI}	60	71.4	$(i=9,10,\ldots,14)$		

3.2.2　系统动态过程的分析

$\bar{u}_1$、$\bar{u}_2$、$\bar{u}_3$分别表示系统处于稳态时进料、排汽和进水阀所处的位置,$\bar{x}_1$、$\bar{x}_2$、$\bar{x}_3$分别表示系统处于稳态时罐内压力、电量输出以及流体密度所处的状态. 系统处于稳态时,3个阀门的位置分别为$\bar{u}_1=0.34,\bar{u}_2=0.69,\bar{u}_3=0.433$. 罐内压力为$\bar{x}_1=108\,\mathrm{kg/cm^2}$,电量产出为$\bar{x}_2=67.7\,\mathrm{mw}$,罐内液体密度为$\bar{x}_3=427.9\,\mathrm{kg/cm^3}$,如图7所示.

经多次反复实验发现,罐内压力、电量输出以及罐内液位的测量值残差在攻击下具有相似的变化规律. 因此,文中只对罐内压力测量值残差变化进行了讨论. 图8是攻击存在时罐内压力测量值的残差检测结果. 从结果来看,压力测量值的残差值维持在10^{-4}级别,说明在未发生攻击的情况下,利用卡尔曼状态观测器能够很好地跟踪系统的状态,残差值几乎为0.

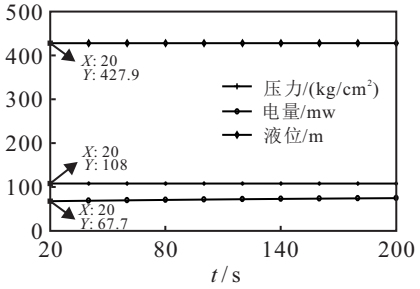

图 7　稳态时系统的状态

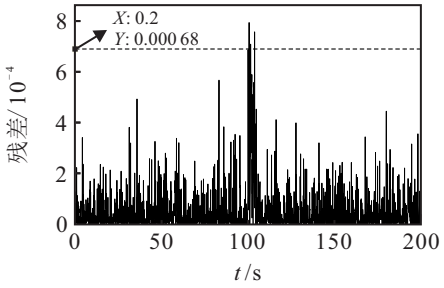

图 8　攻击存在时罐内压力测量值的残差

正常情况下,传感器的稳定输出信号会维持在原测量值的100%±5%之间,考虑到其他因素也可能导致测量值出现大的偏差,引入攻击时间长度,综合判定攻击的有效性. 因此,假定攻击使传感器的测量值偏离原信号5%及以上,且攻击时间长度大于0.2 s的攻击视为有效攻击,反之为无效攻击. 在保证系统漏警率和虚警率不超过2.5%的情况下,通过300次实验得出,压力测量值残差的阈值取6.8×10^{-4}时,观测器有较好的攻击检测效果. 在$t=100\,\mathrm{s}$时,压力测量值的残差大于6.8×10^{-4},系统存在攻击.

图9展示了罐内压力在不同攻击下的动态变化. 攻击作用于u_1时,使进料阀全开,罐内压力迅速上升,在$t=624\,\mathrm{s}$时,系统运行中断;攻击作用于u_2时,使排汽阀全关,这种情况只会降低系统的表现,不会中断系统的运行,$\mathrm{PE}_i(t)=6\,\mathrm{RMB/s}$;攻击作用于$u_3$时,使进水阀全关,这种情况对系统的表现几乎无影响;攻击作用于u_1和u_2时,使进料阀全开,排汽阀全关,在$t=318\,\mathrm{s}$时,系统运行中断;攻击作用于u_1和u_3时,使进料阀全开,进水阀全关,在$t=451\,\mathrm{s}$时,系统运行中断;攻击作用于u_2和u_3时,使排汽阀全关,进水阀全关,这种情况只会降低系统的表现,$\mathrm{PE}_i(t)=4\,\mathrm{RMB/s}$;攻击作用于$u_1$、$u_2$和$u_3$时,使进料阀全开,排汽阀和进水阀全关,在$t=268\,\mathrm{s}$时,系统运行中断.

由图9得到系统在各个执行器被攻击后的工作时长T_i^w和表现损失时间T_i^{pl}后,利用式(18)计算出系统的表现损失PL_i,然后根据式(20)得到ICPS的风险

值,结果如表4所示.

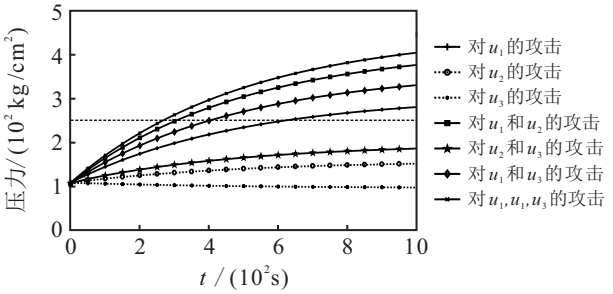

图9 攻击存在时罐内压力的动态变化

表4 ICPS风险估计

控制阀	T_i^w / s	PL_i / 万元	R_{ij}
u_1	624	0.688	52.92
u_2	1 000	0.2	6.3
u_3	1 000	0	0
u_1,u_2	318	0.841	57.96
u_1,u_3	415	0.774 5	55.88
u_2,u_3	1 000	0.3	9.45
u_1,u_2,u_3	268	0.866	58.779

通过对ICPS风险值的分析可知:系统进程没有中断时,系统的表现损失越大,系统的风险也越大;系统进程中断时,系统在一个工作周期内,工作时间T_i^w越长,系统的风险越小. 而系统进程中断时的风险值要远大于系统进程没有中断时的风险值,因此,要提前预防能使系统进程中断的攻击发生. 在得到系统的风险估计值后,可以根据不同的风险值,采用相应的防御措施.

4 结 论

本文采用传统的贝叶斯网络求攻击成功入侵的概率,并深入研究系统在攻击下的动态表现,进而评估整个ICPS 的风险. 采用贝叶斯网络对攻击在网络层的入侵过程进行建模,描绘入侵阶段攻击对系统漏洞和权限的利用,求出攻击成功入侵的概率. 采用卡尔曼状态观测器实时监测系统的动态进程,通过对系统在攻击下表现的分析,量化攻击对系统造成的影响. 根据攻击成功入侵的概率和攻击影响,综合评估ICPS的风险值. 最后研究沸水发电厂模型中的攻击,计算攻击发生的概率,并分析模型在攻击下的动态表现,结果表明本文给出的风险评估方法能有效地评估ICPS的风险.

参考文献(References)

[1] Mahmoud M S, Hamdan M M, Baroudi U A. Modeling and control of cyber-physical systems subject to cyber attacks: A survey of recent advances and challenges[J]. Neurocomputing, 2019, 338: 101-115.

[2] Kure H, Islam S, Razzaque M. An intergrated cyber security risk management approach for a cyber-physical system[J]. Applied Sciences, 2018, 8(6): 898-927.

[3] Cherdantseva Y, Burnap P, Blyth A, et al. A review of cyber security risk assessment methods for SCADA systems[J]. Computer & Security, 2016, 56: 1-27.

[4] Wu W B, Kang R, Li Z. Cyber security risk assessment method of cyber physics system based on attack graph[J]. Journal of Computer Application, 2016, 36(1): 203-206.

[5] Zhang Q, Zhou C J, Xiong N X, et al. Multimodel-based icident prediction and risk assessment in dynamic cybersecurity protection for industrial control systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2015, 46(10): 1-16.

[6] Gao N, Gao L, He Y Y, et al. Dynamic security risk assessment model based on Bayesian attack graph[J]. Journal of Sichuan University: Engineering Science Edition, 2016, 48(1): 111-118.

[7] Hamed O, Mohammad A A. Evaluating the complexity and impacts of attacks on cyber-physical systems[C]. Symposium on Real-Time and Embedded Systems and Technologies (RTEST). Tehran: IEEE, 2015: 1-8.

[8] Huang K X, Zhou C J, Tian Y C. Assessing the physical impact of cyberattaon industrial cyber-physical systems[J]. IEEE Transactions on Industrial Electronics, 2018, 65(10): 8153-8162.

[9] Orojloo H, Azgomi M A. A stochastic game model for evaluating the impacts of security attacks against cyber-physical systems[J]. Journal of Network and Systems Management, 2018, 26(4): 929-965.

[10] Orojloo H, Azgomi M A. A method for evaluating the consequence propagation of security attacks in cyber-physical systems[J]. Future Generation Computer Systems, 2017, 67: 57-71.

[11] Mell P, Scarfone K, Romanosky S. Common vulnerability scoring system[J]. IEEE Security and Privacy Magazine, 2006, 4(6): 85-89.

[12] Poolsappasit N, Dewri R, Ray R. Dynamic security risk management using Bayesian attack graphs[J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9(1): 61-74.

[13] Sun Z W, Zhang Y Q. Research on attack modeling of industrial cyber physics system[J]. Control and Decision, 2019, 34(11): 2323-2329.

[14] Dong W, Kun J. Resilient industrial control system (RICS) concepts, formulation, metrics, and insights[C]. IEEE International Symposium on Resilient Control Systems. Salt Lake City: IEEE, 2010: 15-22.

[15] Tan W, Marquez H J, Chen T W, et al. Analysis and control of a nonlinear boiler-turbine unit[J]. Journal of Process Control, 2005, 15(8): 883-891.

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;

张书国(1995−), 男, 硕士生, 从事工业信息物理系统安全风险评估的研究, E-mail: 6181913048@stu.jiangnan.edu.cn.

(责任编辑: 齐 霁)