软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2019,30(7):2018−2032 [doi: 10.13328/j.cnki.jos.005756]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

具有多传感器的 CPS 系统的攻击检测*

杨 康[1], 王 瑞[1,2], 关 永[2], 李晓娟[1,2], 施智平[2], Xiaoyu SONG[3]

[1](轻型工业机器人与安全验证北京市重点实验室(首都师范大学 信息工程学院),北京 100048)

[2](电子系统可靠性与数理交叉学科国家国际科技合作示范型基地(首都师范大学),北京 100048)

[3](Portland State University, Portland 97230, OR 97230, USA)

通讯作者: 王瑞, E-mail: rwang04@cnu.edu.cn

摘 要: 信息物理系统(cyber-physical systems,简称 CPS)是基于环境感知实现计算、通信与物理元素紧密结合的下一代智能系统,广泛应用于安全攸关的系统和工业控制等领域.信息技术与物理世界的相互作用使得 CPS 容易受到各种恶意攻击,从而破坏其安全性.主要研究存在瞬态故障的 CPS 中传感器的攻击检测问题.考虑具有多个传感器测量相同物理变量的系统,其中一些传感器可能受到恶意攻击并提供错误的测量.此外,使用抽象传感器模型,每个传感器为控制器提供一个真实值的可能间隔.已有的用于检测传感器被恶意攻击的方法是保守的.当专业攻击者在一段时间内轻微地或不频繁地操纵传感器的输出时,现有方法很难捕获到攻击,如隐身攻击.为了解决这个问题,设计了一种基于融合间隔和历史测量的传感器攻击检测方法.该方法首先为不同的传感器构建不同的故障模型,使用系统动力学方程把历史测量融入到攻击检测方法中,从不同的方面分析传感器的测量.另外,利用历史测量和融合间隔解决了两个传感器的测量相交时是否存在故障的问题.该方法的核心思想是利用传感器之间的成对不一致关系检测和识别攻击.从 EV3 地面车辆上获得真实的测量数据来验证算法的性能.实验结果表明,所提出的方法优于现有方法,对各种攻击类型都有较好的检测和识别性能,特别是对于隐身攻击,检测率和识别率大约提高了 90%以上.

关键词: CPS;安全性;瞬态故障;多传感器融合算法;传感器攻击检测和识别

中图法分类号: TP309

中文引用格式: 杨康,王瑞,关永,李晓娟,施智平,Xiaoyu Song.具有多传感器的 CPS 系统的攻击检测.软件学报,2019,30(7):2018−2032. http://www.jos.org.cn/1000-9825/5756.htm

英文引用格式: Yang K, Wang R, Guan Y, Li XJ, Shi ZP, Song XY. Attack detection of CPS system with multi-sensors. Ruan Jian Xue Bao/Journal of Software, 2019,30(7):2018−2032 (in Chinese). http://www.jos.org.cn/1000-9825/5756.htm

Attack Detection of CPS System with Multi-sensors

YANG Kang[1], WANG Rui[1,2], GUAN Yong[2], LI Xiao-Juan[1,2], SHI Zhi-Ping[2], Xiaoyu SONG[3]

[1](Beijing Key Laboratory of Light Industrial Robots and Safety Verification (College of Information Engineering, Capital Normal University), Beijing 100048, China)

[2](National International Science and Technology Cooperation Demonstration Base of Interdisciplinary of Electronic System Reliability and Mathematics (Capital Normal University), Beijing 100048, China)

[3](Portland State University, Portland 97230, OR 97230, USA)

Abstract: Cyber-physical systems (CPS) are next-generation intelligent systems based on environment-aware computing,

* 基金项目: 国家自然科学基金(61877040, 61702348, 61602325); 国家重点研发计划(2017YFB1301100)

Foundation item: National Natural Science Foundation of China (61877040, 61702348, 61602325); National Key R&D Plan (2017YFB1301100)

本文由“软件形式化验证”专题特约编辑贺飞副教授、张立军研究员推荐.

收稿时间: 2018-07-13; 修改时间: 2018-09-28; 采用时间: 2018-12-13; jos 在线出版时间: 2019-03-28

CNKI 网络优先出版: 2019-03-29 09:14:28, http://kns.cnki.net/kcms/detail/11.2560.TP.20190329.0914.008.html

communication, and physical elements. They are widely used in security-critical systems and industrial control. The interaction of information technology and the physical world makes CPS vulnerable to various malicious attacks, thereby undermining its security. This work mainly studies the attack detection problem of sensors in CPS systems with transient faults. This study considers a system with multiple sensors measuring the same physical variables, and some sensors may be malicious attacked and provide erroneous measurements. In addition, this study uses an abstract sensor model where each sensor provides the controller with an interval of possible values for the true value. Existing methods for detecting sensor malicious attacks are conservative. When a professional attacker manipulates the sensor's output slightly or infrequently over a period of time, existing methods are difficult to capture attacks, such as stealth attacks. In order to solve this problem, this study designs a sensor attack detection algorithm based on fusion intervals and historical measurements. First, the algorithm constructs different fault models for different sensors, integrates historical measurements into the attack detection method using system dynamics equations, and analyzes sensor measurements from different aspects. In addition, combined with historical measurement and fusion interval, the problem of whether there are faults when the two sensors intersect is solved. The core idea of this method is to detect and identify attack by using pairwise inconsistency between sensors. This study obtains real measurement data from EV3 ground vehicles to verify the performance of the algorithm. The experimental results show that the proposed method is superior to the state-of-the-art algorithm, and has better detection and recognition performance for various attack types. Especially for stealth attacks, the detection rate and recognition rate are increased by more than 90%.

Key words: cyber-physical system; security; transient fault; multi-sensor fusion algorithm; sensor attack detection and identification

信息物理系统(cyber-physical system,简称 CPS)在环境感知的基础上实现了计算、网络通信和物理元素的紧密结合[1].CPS 改变了人们与周围物理世界的相互作用和控制方式.与传统的嵌入式系统相比,CPS 具有智能化、控制精确、网络开放性等诸多优点.近年来,随着信息技术与控制技术的发展,CPS 广泛应用于国家关键基础设施和工业控制等领域,包括医疗系统、供水系统、国防和武器系统等[2,3].

随着信息物理系统在安全攸关领域的应用越来越广泛,确保该系统的安全是很有必要的[4−6].因此,CPS 的安全问题也越来越受到人们的重视.攻击者可能有不同的动机去恶意破坏系统,对系统造成不同程度的损坏,从性能上的轻微干扰到完全地控制系统[7].最近的工作表明,攻击者可能通过车辆的车载通信协议或传感器欺骗中的漏洞来劫持现代车辆[8,9].类似地,如 Maroochy Water 漏洞[10]和蠕虫 Stuxnet 能够通过利用监督控制和数据采集(SCADA)系统[11]中的弱点来禁用关键基础设施.

随着传感器技术的迅速发展,现代 CPS 通常配备有测量相同物理变量的多个传感器.例如,可以用 GPS、轮编码器、超声波和 IMU 等传感器来测量汽车的速度.融合这些传感器的数据可以为控制器提供更准确的估计,这可能会对系统的性能和可靠性产生很大影响[12,13].即使这些传感器的精度可能不同,但是融合它们的测量不仅能够产生比任何单个传感器更精确的估计,而且还能够增加系统对外部干扰的鲁棒性[14,15].此外,具有不同精度和可靠性的不同传感器降低了系统对特定传感器的依赖性.Marzullo 开发了一种融合算法,利用多个传感器的测量,产生包含真实值的融合间隔[15].在 Marzullo 提出的融合算法的基础上,Ivanov 等人通过引入传感器传输时间表[16]和使用历史测量[17]提出了一个精确和鲁棒性的传感器融合算法.

目前已经存在许多关于传感器攻击检测问题的研究,大多数检测恶意传感器攻击的方法都主要考虑概率传感器的情况.例如,Kalman 提出的卡尔曼滤波器,把假设的传感器精度和已知的系统动力学模型相结合,产生最佳线性估计[18].Kwon 等人提出把卡尔曼滤波算法用于在检测测量数据是否被篡改的同时执行状态估计[19].另外,Jayasimha 等人提出通过测量数据协方差矩阵中的元素的变化来检测和识别不良数据[20].然而,当系统中存在瞬态故障时,这些方法容易出现残留污染和残留淹没,将导致误报或无法检测到攻击.这些方法的特点是以相同的方式处理瞬态故障和攻击,因此忽略了传感器有时可能由于临时干扰而提供故障测量的事实.瞬态故障是指传感器在短时间内提供错误的测量并在不久之后消失.例如,在隧道中,GPS 经常暂时与卫星失去连接.因此,不应该把瞬态故障看作对系统的安全威胁.为了消除这些局限性,Park 等人通过为每个传感器提供一个瞬态故障模型来区分瞬态故障和攻击,并利用成对传感器关系开发了一种当存在瞬态故障时传感器攻击检测和识别的方法[21].该方法的局限性在于它们仅考虑了两个传感器不相交的情况,忽略了它们相交时也可能提供有故障测量的事实,致使有些攻击不能被检测出来.例如,隐身攻击.即,攻击者有足够的能力,为了保持不被检测到,以最大化融合间隔,并使得任意两个传感器的间隔尽可能地相交.

本文主要解决 CPS 的安全问题,这通常涉及系统的最坏情况,因此,本文采用抽象传感器模型.为了解决现有的攻击检测方法存在的一些局限性,本文提出了一种基于融合间隔和历史测量的传感器攻击检测和识别方法,并考虑了两个传感器间隔相交时的故障检测问题.该方法的核心思想是从不同角度考虑传感器的测量,并为不同的传感器建立不同的故障模型,结合融合间隔和历史测量,利用两个传感器之间的不一致关系来检测攻击.此外,本文还提出了一种基于构建 ROC(receiver operating characteristic)曲线来选择瞬态故障模型参数的方法.最后,本文从 EV3 机器人平台上获得真实的实验数据来评估算法的性能.通过获取不同窗口和攻击场景下超声波和两个大型电机的传感器数据,训练合适的瞬态故障模型,验证了本文提出的传感器攻击检测和识别方法的检测和识别性能,并分析了误报情况.实验结果表明,本文提出的方法对各种攻击类型都非常有效.特别是对于隐形攻击,本文提出的算法的优势是显而易见的,与现有的检测算法相比,检测的准确率大约提高了 90%以上.

本文第 1 节介绍本工作的一些基础知识,包括系统模型、传感器模型、瞬态故障模型以及攻击模型.第 2 节介绍两种多传感器融合算法.第 3 节详细阐述本文提出的基于融合间隔和历史测量的传感器攻击检测和识别方法,并提出一种基于构建 ROC 曲线选择瞬态故障模型参数的方法.为了增加算法的可读性,第 3 节给出一种算法的实例.第 4 节通过 EV3 平台验证算法的性能,并介绍在实际场景中瞬态故障模型的构建.第 5 节总结本文工作并提出未来的研究方向.

1 预备知识

为了更好地理解本文提出的基于融合间隔和历史测量的传感器攻击检测和识别方法,本节主要介绍本工作中的一些基础知识.首先介绍系统模型,并简单分析了在具有非线性系统模型的 CPS 中,本文提出的检测算法的可扩展性.之后介绍了本文使用的抽象传感器的模型,详细描述了传感器间隔的构建.然后介绍用来区分瞬态故障和攻击的瞬态故障模型.最后介绍本文用到的 3 种攻击类型:偏差攻击、随机攻击和隐身攻击.

1.1 系统模型

本文考虑具有测量相同物理变量(例如,速度)的 N 个传感器组成的 CPS,利用多个传感器产生的冗余信息为控制器提供更准确的估计值.本文假设 CPS 在 T(系统运行的总时间)周期内以周期性的方式查询所有传感器,不断地获取所有传感器的测量,然后执行传感器融合算法和攻击检测算法.由于大多数 CPS 具有已知的动力学,本文假设 CPS 由一个离散线性时不变系统组成,其形式如下:

$$x_{t+1} = Ax_t + Bu_t + \omega_t,$$
$$y_t = Cx_t + v_t,$$

其中, $x_t = (x_{t1}, x_{t2}, x_{t3}, ..., x_{tn}) \in R^n$, 表示在 t 时刻的状态; $u_t = (x_{t1}, x_{t2}, x_{t3}, ..., xt_p) \in R^p$, 表示控制器在 t 时刻的输入信号; $y_t \in R^m$ 为系统输出,即传感器的测量值; $\omega_t \in R^q$ 和 $v_t \in R^l$,分别表示系统干扰和测量噪声; A、B、C 属于系统信息, $A \in R^{n\times m}$ 和 $B \in R^{n\times p}$,分别为状态矩阵和控制输入矩阵, $C \in R^{m\times n}$ 为输出矩阵.

本文主要研究具有上述线性系统模型的 CPS 中传感器的攻击检测和识别问题.然而,本文提出的传感器攻击检测方法可以很容易地扩展到具有非线性系统模型的情况,如变速和转弯的情况.扩展思路主要有两个:(1) 根据实际的输入调整瞬态故障模型.例如,可以使用查找表,该查找表包含系统在不同操作模式下训练的多个瞬态故障模型的参数值.即,每一个速度段给出一组瞬态故障模型参数,针对不同的系统操作模式动态地选择合适的参数.(2) 针对转弯的情况.对于转弯的瞬间,那个时刻点 t 不使用历史测量,对于其他的时刻点,本文提出的算法都是适用的.

如图 1 所示,本文专注于一个广泛使用的 CPS 架构,它的基本工作流程是:CPS 通过传感器实时地感知物理对象的变化,然后通过网络把传感器的测量信息发送给控制器.控制器中的计算单元根据多个传感器的测量信息正确地认识环境,并根据一定的算法做出相应决策,再通过网络将决策反馈给执行器,最后由执行器执行该决策,进而改变物理对象.在整个过程中,可能有些地方会受到恶意攻击.例如,在通过网络把传感器的测量发送给控制器这个过程中,攻击者可能会通过物理攻击或网络攻击篡改传感器的值,导致控制器收到的测量值是错误

的,这可能导致整个 CPS 不能正常运行.要实现对 CPS 的精确控制,必须保证各种传感器输出真实可靠.因此,实现传感器故障的快速准确检测,对于提高 CPS 的安全可靠性具有十分重要的意义.本文结合 CPS 安全可靠性这一迫切需求,利用多传感器冗余信息,开展传感器攻击检测和识别方法及应用的研究.本文在控制器这里添加一种传感器攻击检测方法,当控制器收到传感器的测量值时,首先经过攻击检测算法,判断收到的值是否是正确的,然后再进行下一步工作.

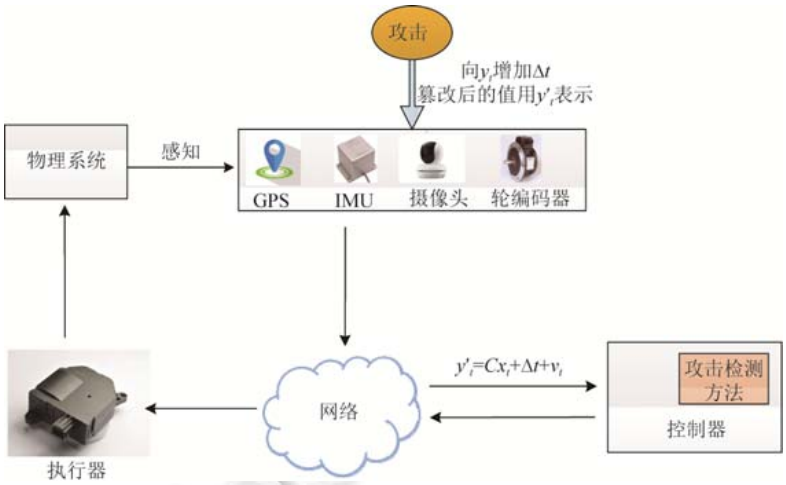

Fig.1 System architecture

图 1 系统框架

1.2 传感器模型

目前文献中使用的传感器模型大致分为概率传感器模型和抽象传感器模型两类.在前者中,每个传感器向控制器提供具有已知概率分布的被噪声破坏的单个测量,如高斯分布.噪声分布对更可能成为真正有价值的点赋予更多的权重.因此,概率传感器模型非常适合分析预期的系统性能,但需要了解噪声分布的知识.其中,在错误的噪声假设条件下设计的检测器会降低攻击检测的准确性[22].在后者中,每个传感器向控制器提供一个包含所有可能为真实值的点的间隔,这种类型的传感器模型假定不知道区间上的噪声分布,但即使在最坏的情况下,也可以构造这些区间来包含未知的真实值.由于抽象测量是真实值的最坏情况界限,因此,它们非常适用于拜占庭故障和传感器攻击下系统运行的最坏情况分析.

本文采用抽象传感器模型.抽象间隔可以基于制造商关于传感器的精度和规范以及诸如采样抖动和同步误差的物理限制来构造.在本文中,每个传感器围绕其自身测量构造一个区间$[S_i(t)-\delta_i,S_i(t)+\delta_i]$,用 $S_i^F(t)$ 表示.其中,$S_i(t)$是传感器 S_i 在 t 时刻的测量值,δ_i 是传感器的误差界限(即,误差精度).传感器 S_i 的间隔大小为 $2\delta_i$.间隔大小反映了传感器的精度,小的间隔意味着获得的测量具有较高的可信度.另外,传感器 S_i 在时间 t 的真实值由 $\tau_i(t)$ 表示(通常是未知的).有时传感器可能会提供错误的测量,如果 t 时刻 S_i 的测量值偏离真实值,并且其偏离值大于 δ_i,则 $S_i(t)$被称为故障测量.用谓词 $F(S_i,t)$表示传感器 S_i 在 t 时刻提供了错误的测量.故障测量的定义形式化为如下定义.

定义 1(故障测量).

$$F(S_i,t) \equiv |S_i(t)-\tau_i(t)| > \delta_i.$$

1.3 瞬态故障模型

传感器 S_i 的瞬态故障模型是一个三元组(δ_i,f_i,w_i),其中,δ_i 是误差界限;(f_i,w_i)是瞬态阈值,其指定在任何大小为 w_i 的窗口中,S_i 可以提供至多 f_i 个错误测量.瞬态阈值用于定义瞬态故障和非瞬态故障之间的边界.如果 S_i 不违反其瞬态阈值,则称为瞬态故障,由谓词 $TF(S_i,t)$表示,否则称为非瞬态故障.

定义 2(瞬态故障).

$$TF(S_i,t)\equiv\sum_{t'=t-\omega_i+1}^{t}F(S_i,t)\leqslant f_i,\ t\geqslant w_i,$$

其中,如果 S_i 在 t 时刻提供了故障的测量,则 $F'(S_i,t)=1$.否则,$F'(S_i,t)=0$;如果($t\geqslant w_i$),则从第 1 轮检测开始累加,即 $t'=1$.

本文把永久性故障和恶意攻击都形式化为非瞬态故障,以相同的方式对待它们.当然,也可能存在符合瞬态故障模型的攻击,这将留作以后的工作,本文仅考虑表现为非瞬态故障的攻击.因此,如果传感器 S_i 在整个系统操作期间至少出现 1 次非瞬态故障,则称 S_i 受到了攻击,用谓词 $A(S_i,t)$表示.

定义 3(攻击).

$$A(S_i,t)\equiv\exists t\leqslant T,\neg TF(S_i,t),$$

其中,T 是系统运行的总时间.

1.4 攻击模型

目前,攻击者一般有物理攻击和网络攻击两种攻击手段.物理攻击是指通过损坏或替换传感器来完成,或通过其他物理手段引入偏差来实现[23].网络攻击是指通过利用其软件中的漏洞或用一个全新的版本替换其代码来危害传感器.因此,攻击者可以通过网络访问传感器,甚至不需要物理接近.在现代传感器中也可以利用在其他嵌入式系统中可能发生的任何软件缺陷.

如第 1.1 节中所述,本文假设系统中的传感器通过共享总线进行通信.因此,通过获得对传感器的控制,攻击者能够获得当前轮次的所有传感器测量以及之前所有传感器发送到总线上的历史测量数据.本文假设攻击者可以代表损坏的传感器发送任何测量,并且最坏情况下攻击者具有无限的计算能力和全系统知识,包括传感器的规范、采用的传感器融合算法和攻击检测算法等.由于每个传感器的输出是一个间隔,那么,攻击者的目标是通过篡改某些传感器的值,使其间隔最大化,从而降低了所提供的测量的置信度,使系统处于不安全状态.由于 CPS 的性质和体系结构,攻击者可以利用物理层和网络层的弱点.根据攻击者不同的攻击目的,本文考虑 3 种攻击类型.

(1) 偏差攻击(bias attack).攻击者为被攻击的传感器添加一个偏差值Δ,其中,$\Delta=2\delta_{\max}$,$\delta_{\max}$ 是指所有传感器中最大的误差精度.$W=(w_1,w_2,w_3,\ldots,w_T)$,即窗口的大小即为系统的总运行时间 T.对 $t=1,2,\ldots,T$,有

$$\tilde{S}_i(t)=\begin{cases}S_i(t)+2\delta_{\max}, & t=k\\ S_i(t), & t\neq k\end{cases},$$

其中,$\tilde{S}_i(t)$ 是被篡改后的测量;k 是被攻击的时刻,即在 k 时刻被攻击的传感器提供错误的测量.

(2) 随机攻击(random attack).该攻击类型与偏差攻击非常相似,其主要区别在于,该攻击是为 t 时刻被攻击的传感器添加一个随机的偏差,偏差的范围是$(0,2\delta_{\max})$.当添加的偏差值很小时,系统很难发现此攻击.显然,随机攻击比偏差攻击更难被检测到.

(3) 隐身攻击(stealth attack).攻击者用一个新的测量代替当前时刻的测量值,最大化融合间隔的大小并尽可能地使任意两个传感器的间隔相交.这类攻击是在假设攻击者已知系统所使用的融合算法、所有测量数据以及传感器的规范等知识的前提下所设计的攻击.其攻击的目的是尽可能地保持不被检测到.该攻击类型是 3 种攻击类型中危害性和隐蔽性最强的.

2 多传感器融合算法

2.1 经典的融合算法

本节介绍一种基于抽象传感器模型的融合算法,该算法由 Marzullo 在 1990 年提出[15].该算法的输入是 N 个传感器的间隔和一个数字 f,其中,f 表示系统中被破坏的传感器数目的上限.算法的输出是一个间隔,在本文中称为融合间隔,其跨越包含至少 $N–f$ 个间隔中的最小到最大点.由于存在至多 f 个损坏的传感器,至少存在 $N–f$

个正确的传感器,因此真实值可以在任何 $N–f$ 个间隔组中.最后,算法输出包含所有这样的组的最小间隔.

定义 4(融合间隔). l 表示包含至少 $N–f$ 个传感器间隔的最小值,h 表示包含至少 $N–f$ 个传感器间隔的最大值. $F_N^f(S,t)$ 表示融合间隔$[l..h]$.

融合间隔的计算步骤如下:首先,我们对所有传感器的下界和上界分别以升序和降序进行排序,结果分别存放在 *str_start* 和 *str_end* 数组中.然后,我们按照数组下标(从 0 开始)从小到大的顺序开始扫描*str_start*,直到扫描到融合间隔的下界 l 与所有传感器的交点共 $N–f$ 个为止,这个数值就是融合间隔的下界 l.按同样的方法扫描 *str_end*,得到融合间隔的上界 h.

图 2 所示为该融合算法的一个示例.从图中可以看出,融合间隔的准确性取决于 f 的值.因为 f 是未知的,通常我们将 f 设得保守地高,本文假设 $f=\lceil N/2\rceil-1$,这在文献[17]中已被证明.

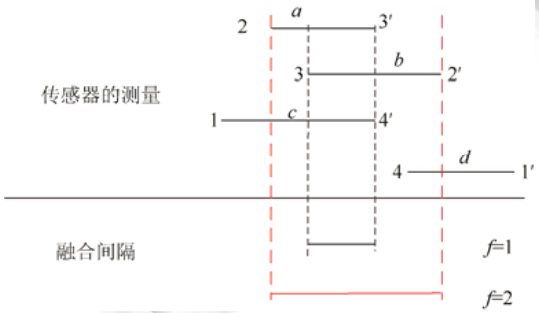

Fig.2 The fusion interval with N=4 sensors and f=1, f=2

图 2 具有 4 个传感器且 f=1 和 f=2 的融合间隔

2.2 基于历史测量的改进的融合算法

在上述经典的传感器融合算法的基础上,本文提出了一种改进的传感器融合算法,其核心思想是利用历史测量来减小融合间隔的大小,从而提高融合间隔的精确度.

如第 1.1 节所述,CPS 具有已知的系统动力学模型,因此,可以通过系统动力学方程来使用历史测量以改进融合算法.本文将历史测量加入到融合算法中的基本思路是:把 $t–1$ 时刻的所有传感器(共 N 个)的测量,通过动力学方程映射到 t 时刻,这时,t 时刻就有 $2N$ 个测量值,此时发生故障的传感器的数量最大是 $2f$,然后使用经典的融合算法计算融合间隔.在添加过去的测量之后,改进的融合算法仍然需要满足两个标准:(1) 在包含真实值的前提下,融合间隔尽可能地小.(2) 融合间隔不大于未使用历史测量时获得的融合间隔.

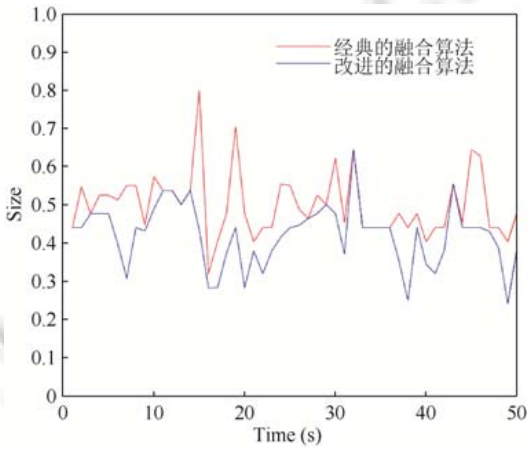

Fig.3 The fusion interval size of two fusion algorithms in the random attack scenario

图 3 随机攻击场景下两种融合算法的融合区间大小

为了比较上述两种融合算法,本文考虑了超声波传感器受到随机攻击的情况,分别使用两种融合算法来计算融合间隔.实验结果如图 3 所示.从图中可以看出,使用历史测量的融合算法得到的融合间隔永远不会大于不使用历史测量的融合算法得到的融合间隔.该现象说明,历史测量的使用确实可以缩小融合间隔的大小,提高融合算法的精确度.

由于真实值通常是不知道的,因此很难确定哪个传感器提供了错误的测量值.Marzullo 在 1990 年提出基于

抽象传感器的融合算法的同时,也提出了一种基于融合间隔的传感器攻击检测方法.该方法的主要思想是让每个传感器与融合间隔进行比较,任何不与融合间隔相交的传感器都是故障的.

命题 1. 给定 $S_i^F(t), F_N^f(S,t)$ 和时间 t,

$$S_i^F(t) \cap F_N^f(S,t) = \varnothing \Rightarrow F(S_i,t).$$

证明:在故障传感器数量有限的情况下,融合算法会产生一个包含真值的融合区间.由于真实值不可能存在于两个不相交的区间中,因此,如果传感器 S_i 的间隔不与融合区间相交,则意味着 S_i 一定是故障的. □

3 基于融合间隔和历史测量的传感器攻击检测方法

本节描述了本文提出的用于传感器攻击检测和识别问题的改进的成对不一致(boosted pairwise inconsistency,简称 BPI)算法.该方法在基于 PI 方法[21]的基础上考虑了融合间隔和历史测量,并解决了当两个传感器的间隔相交时传感器的攻击检测和识别问题.

3.1 攻击检测

本文提出的传感器攻击检测和识别方法基于传感器之间的成对不一致关系.本节首先介绍两个关键概念,即传感器之间的成对不一致类型:弱不一致性和强不一致性.该部分建立在假设真实值通常不知道的前提下,因此不知道哪些传感器提供了正确的测量.然而,已知的是传感器测量之间的相互关系,本文正是利用这种相互关联的信息来检测和识别攻击.两个传感器 S_i 和 S_j 之间的第 1 种关系是弱不一致性,由谓词 $WI(S_i,S_j,t)$ 表示.如果当且仅当其中一个传感器提供了故障的测量,那么两个传感器在给定的回合中是弱不一致性的.

定义 5(弱不一致性).

$$WI(S_i,S_j,t) \equiv F(S_i,t) \vee F(S_j,t).$$

由于真实值是未知的,这个条件通常很难验证.因此,该定义不能用于判断两个传感器是否是弱不一致性.然而,因为真实值不能位于两者中,因此,如果两个传感器的间隔彼此不相交,则它们中的一个必须已经提供了错误的测量.

由上述弱不一致性的定义可知,瞬态故障和攻击可能在单个回合中都表现出弱不一致性.因此,本文引入强不一致性来消除两者之间的歧义.如果两个传感器中至少有一个是非瞬态故障(传感器不符合其瞬态故障模型),那么这两个传感器之间存在强不一致性.形式化地,有如下定义.

定义 6(强不一致性).

$$SI(S_i,S_j,t) \equiv \neg TF(S_i,t) \vee \neg TF(S_j,t).$$

与弱不一致性的定义类似,因为不知道真实值,很难判断 S_i 是否是非瞬态故障.特别地,如果两个传感器在给定的窗口中频繁地发生弱不一致,那么它们变得强不一致.

引理 1. 给定传感器 S_i、S_j 和时间 t,

$$\sum_{t'=t-\min(w_i,w_j)+1}^{t'=t} WI(S_i,S_j,t') > f_i + f_j \Rightarrow SI(S_i,S_j,t), t \geqslant \min(w_i,w_j).$$

注意,如果 $t<\min(w_i,w_j)$,那么就从 $t'=1$ 开始累加弱不一致信息.

证明:根据定义 5,如果两个传感器 S_i 和 S_j 在 t'时是弱不一致性的,这意味着在 t'时至少有一个传感器是有故障的,那么在 $\min(w_i,w_j)$或者更小($t<\min(w_i,w_j)$)的窗口中,如果 S_i 和 S_j 的弱不一致信息的累加和大于(f_i+f_j),则意味着,至少有一个传感器是非瞬态故障,即, $(f_i' > f_i) \vee (f_j' > f_j)$, 其中, f_i' 表示 S_i 提供故障测量的总次数.

3.1.1 弱不一致检测

弱不一致检测方法是本文提出的传感器攻击检测和识别方法的核心部分,是提高攻击检测和识别性能的关键步骤.由于 2015 年 Park 等人提出的基于 PI 的攻击检测方法仅考虑了两个传感器不相交的情况,忽略了两个传感器相交时也可能存在故障的情况,并且该方法无法检测和识别到隐身攻击.针对这些情况,本文提出了一种新的弱不一致检测方法来解决这些问题.本文提出的弱不一致检测方法分为两步.

(1) 两个测量不相交.

首先比较当前时刻(t)任意两个传感器的测量,如果两个传感器的测量不相交,则至少有一个传感器提供了故障测量,即它们在 t 时刻是弱不一致关系.

$$S_i^F(t)\cap S_j^F(t)=\varnothing \Rightarrow WI(S_i,S_j,t).$$

然后融合过去和当前的测量,将所有传感器的测量从 t–1 时刻映射到 t 时刻,在 t 时刻共 $2N$ 个测量.最后在 t 时刻比较任意两个测量值,注意,不包括同一个传感器不同时刻的两个测量.如果这两个测量不相交,那么这两个传感器之间存在弱不一致关系.

$$S_i^{F'}(t)\cap S_j^{F'}(t)=\varnothing \Rightarrow WI(S_i,S_j,t),\ i\neq j.$$

(2) 两个测量相交.

当两个测量相交时,主要是从不同的角度利用融合间隔和历史测量来判断故障.

第 1 步:首先计算 t 时刻的融合间隔,让每个传感器和融合间隔进行比较,如果两个传感器 S_i 和 S_j 均不与融合间隔相交,并且这两个传感器的测量相交,则 S_i 和 S_j 是弱不一致关系.

$$\begin{cases} S_i^F(t)\cap F_{N+1}^f(S,t)=\varnothing \\ S_j^F(t)\cap F_{N+1}^f(S,t)=\varnothing \\ S_i^F(t)\cap S_j^F(t)\neq\varnothing, i\neq j \end{cases} \Rightarrow WI(S_i,S_j,t).$$

第 2 步:结合历史测量进行判断.首先计算 t–1 时刻的融合间隔,将每个传感器在 t–1 时刻的测量与该融合间隔进行比较;然后把 t–1 时刻的测量映射到 t 时刻,再进行两两比较.当满足下列条件时,说明传感器 S_i 和 S_j 是弱不一致关系.

$$\begin{cases} S_i^F(t-1)\cap F_{N+1}^f(S,t-1)=\varnothing \\ S_j^F(t)\cap F_{N+1}^f(S,t)=\varnothing \\ S_i^{F'}(t)\cap S_j^F(t)\neq\varnothing, i\neq j \end{cases} \Rightarrow WI(S_i,S_j,t).$$

算法 1 给出了具体实现.第 1 行~第 8 行实现了在 t 时刻任何两个测量之间的比较,并将不一致性信息存储在弱不一致数组中.第 9 行将所有传感器的测量值从时间 t–1 时刻映射到 t 时刻.第 10 行调用融合算法计算融合间隔.第 11 行实现了当前时刻和历史的传感器测量之间的比较.第 12 行完成了单个传感器和融合间隔的比较,并将这些传感器的信息存储在不一致数组中.使用弱不一致检测方法最终会得到一个弱不一致数组,里面存放了在每一个时刻存在弱不一致关系的传感器的信息.这些信息将会在传感器的攻击检测和识别中用到.

Algorithm 1. Weak inconsistent detection algorithm.

Input: N measurements of the abstract sensor.

1: $w\leftarrow 0$;
2: **for** $i=0\rightarrow N-2$ **do**
3: **for** $j=i+1\rightarrow N-1$ **do**
4: **if** $S_i^F(t)\cap S_j^F(t)==\varnothing$ **then**
5: $weaks[w{+}{+}]\leftarrow(i{+}1,j{+}1,t)$
6: **end if**
7: **end for**
8: **end for**
9: $S_i^{F'}(t)\leftarrow S_i^F(t-1)$ (i=1,2,…,N);
10: fusion();
11: compareTwoSensorHistory ($S_i^{F'}(t),S_j^F(t)$);
12: SensorAndFusionCompare(sensors, $F_N^f(S,t)$);
13: return weaks;

3.1.2 强不一致检测

本小节通过采用上述的不一致概念来展示本文提出的传感器攻击检测方法.由定义 6 可知,如果两个传感

器之间存在强不一致关系,则说明两个传感器中至少有一个是非瞬态故障.因此,由定义 3 攻击的定义可知,如果存在时间 $t \leqslant T$,使得传感器 S_i 和 S_j 是强不一致的,则说明系统中存在被攻击的传感器.其中,T 是系统运行的总时间.

引理 2(攻击检测).

$$\sum_{t'=1}^{t} SI'(S_i,S_j,t) \geqslant 1 \Rightarrow AD(S_i,S_j,t), t \geqslant 1,$$

其中,如果 $SI(S_i,S_j,t)$存在,则 $SI'(S_i,S_j,t)=1$;否则,$SI'(S_i,S_j,t)=0$.

证明: $\sum_{t'=1}^{t} SI'(S_i,S_j,t) \geqslant 1$ 说明传感器 S_i 和 S_j 在 t 时刻是强不一致关系,由强不一致的定义可知,S_i 和 S_j 至少有一个是非瞬态故障.根据定义 3 可知, $\neg TF(S_i,t) \vee \neg TF(S_j,t) \Rightarrow A(S_i,t) \vee A(S_j,t)$, 因此,可以断定系统中存在攻击.证毕. □

3.2 攻击识别

上述的攻击检测方法仅考虑检测系统中是否存在传感器被攻击,并没有考虑哪个传感器被攻击的问题.在本小节中,为了确定哪个传感器受到攻击,本文假设系统中至多有 $s(s<N-1)$个传感器受到攻击.本文提出的攻击识别方法是:累积强不一致信息,在强不一致对中,如果传感器 S_i 出现的次数超过 s,则称 S_i 被攻击了.

引理 3. 给定传感器 S_i 和时间 t,用 $degree(S_i,t)$表示在 t 时刻与 S_i 存在强不一致关系的传感器的数量.

$$degree(S_i,t) > s \Rightarrow A(S_i,t).$$

证明:假设有 $n(n>s)$个传感器和 S_i 相连,与 S_i 相连的传感器用 S_j 表示.由于 S_i 和 S_j 是强不一致关系,如果 S_i 没有被攻击,那么 S_j 必须被攻击.此时,共有 n 个传感器受到攻击,这与最多有 s 个传感器被攻击的假设相矛盾.证毕. □

3.3 瞬态故障模型参数选择

本文提出的 BPI 方法需要精确的瞬态故障模型.现在的制造商一般会提供传感器的瞬态故障规范.尽管制造商有时候会提供算法所需要的全部参数,但对应不同的应用场景和需求,这些参数可能并不是最佳的(例如,在被高建筑物包围的环境中使用 GPS).此外,针对不同的传感器攻击算法,有些参数制造商可能无法提供,比如本文提出的 BPI 算法的参数 f_i.那么,就有必要基于经验数据来开发瞬态故障模型.因此,本节提出一种通过构建 ROC 曲线来选择瞬态故障模型参数的新方法,用于抽象传感器的攻击检测.

受试者工作特征曲线为 ROC 曲线(receiver operating characteristic curve).通常,该曲线的横轴是假阳性概率(false positive rate),纵轴是真阳性概率(true positive rate),其目的主要是用来选择最佳的界限值.将多个实验得到的 ROC 曲线绘制到同一个坐标中,就能很直观地看出,ROC 曲线的最左上角的点表示错误率最少且准确率最高,意味着该点对应的阈值是最好的.

本文提出的选择瞬态故障模型参数方法的思想是:从实际的实验平台上获取大量的实验数据(传感器的值),把这些数据应用到攻击检测算法中(BPI 方法),从而得到不同场景下传感器的攻击识别率和误报率.然后把各个场景得到的 ROC 曲线绘制到同一个坐标系中.找到识别率最高、误报率最小的点,即图中最左上角的点,该点对应的参数就是最佳的参数.然而,由于每个参数对应的值可能有无数多个,不可能把每一个参数的每个值都进行实验来得到其对应的识别率和误报率.因此,在进行实验之前,需要筛选出这些参数可能的取值范围,然后再取这些范围内的点进行实验,最终确定参数的取值.在第 1.3 节中提出的瞬态故障模型有 3 个参数.参数 δ_i 使用制造商提供的值,另外两个参数 f_i 和 w_i 使用本文提出的基于构建 ROC 曲线选择参数的方法进行确定.其基本步骤如下.

(1) 初步筛选参数,选择若干个值作为 f_i 的候选值.根据受到攻击的传感器数量,本文考虑 3 种攻击情况:只有一个传感器受到攻击,另外两个传感器没有故障;一个传感器受到攻击,另外两个传感器仅有一个传感器存在瞬时故障;一个传感器受到攻击,另外两个传感器均存在瞬时故障.通过对无攻击无故障的传感器数据添加上面 3

种攻击和故障,会得到一些训练数据.此外,分别用 f_{i-1} 和 f_{i+1} 代替 3 种类型的攻击中的 f_i 来形成一系列新的训练集.通过实验得到不同 f_i 对应的识别率和误报率.最后,本文选择几个误报率较小但识别率还能接受的 f_i 作为候选值.这样做虽然不能保证攻击检测率最大,但可以保证误报率最小.

(2) 根据第 1.4 节中设计的攻击模型,为原始数据(无攻击、无故障的传感器数据)添加 3 种类型的攻击.针对(1)中选取的每个候选值进行实验,获得对应窗口的误报率和识别率后,构建 ROC 曲线,然后确定最终的参数值.本文选取参数的具体细节将在第 4.1 节中加以详细介绍.

3.4 传感器攻击检测实例

本小节主要是想通过一个简单的例子来进一步解释上述传感器攻击检测和识别算法.为了便于解释,本文使用一个不一致图 $G(V,E)$来描述该算法.在不一致图中,每个传感器对应一个顶点,两个顶点之间的关系代表它们之间的不一致关系.如果该关系是弱不一致关系,则该图表示弱不一致图,用 *WI_Graph* 表示;如果该关系是强不一致关系,则该图表示强不一致图,用 *SI_Graph* 表示.该不一致图的定义如下.

定义 7(不一致图).

$$V=\{S_1,S_2,\ldots,S_N\},$$

$$E=\{(i,j)|WI(S_i,S_j,\mathrm{t})\ \text{or}\ SI(S_i,S_j,t)\}.$$

本文设计了如图 4 所示的实例,图中有 4 个传感器,窗口 W=5,垂直虚线表示真实值(未知).假设系统中传感器 S_1 和 S_4 受到攻击(最多有 2 个传感器被攻击).注意,不要求被攻击的传感器在每一轮次中均受损,但需要满足在给定窗口中是非瞬态故障,每个传感器的瞬态故障模型见表 1.

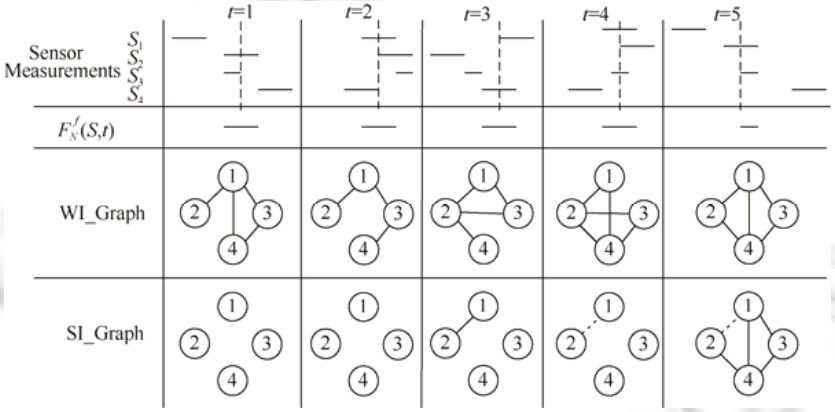

Fig.4 BPI algorithm example

图 4 BPI 算法实例

Table 1 The parameters of the transient fault model in the example of Fig.4

表 1 图 4 实例中对应的瞬态故障模型的参数

传感器	δ_i	f_i	w_i
S_1	1	1	5
S_2	1	1	3
S_3	0.5	2	5
S_4	1	1	5

根据第 3.1.1 节介绍的弱不一致检测算法,得到如 *WI_Graph* 所示的弱不一致关系图.从图中可以看出,系统在每一轮次中都检测到了弱不一致对.在 t=1 时,传感器 S_1 和 S_2 之间存在弱不一致关系,这意味着这两个传感器至少有一个提供了错误的测量.到 t=3 为止,S_1 和 S_2 之间共出现了 3 次弱不一致关系.根据第 3.1.2 节介绍的强不一致检测方法可知,S_1 和 S_2 之间存在强不一致关系,这意味着系统中存在攻击,其他类似.该实例中所有传感器的强不一致关系如图 4 中的 *SI_Graph* 所示.*SI_Graph* 中的虚线代表在之前的检测中已经出现过该强不一致对.虽然,在 t=3 时,检测到系统中存在攻击,但此时无法判断哪个传感器受到了攻击.直到 t=5 时,发现传感器 S_1 和 S_4

的度均为 3,大于 2,根据第 3.2 节中给出的攻击识别算法可以识别出在 t=5 时 S_1 和 S_4 均受到了攻击.

4 实验评估

本节从 EV3 机器人平台上获取实际的实验数据来评估 BPI 算法的性能.首先介绍实验的基本设置,之后介绍瞬态故障模型的参数选择问题,然后评估几种算法的攻击检测和识别性能,最后分析误报的原因.

4.1 实验设置

本文的实验平台选择 LEGO EV3 地面车辆,如图 5 所示.EV3 是 2013 年 LEGO 公司开发的第三代 MINDSTORMS 机器人.它可以安装多个传感器,包括超声波、电机(内嵌角度传感器)、陀螺仪、颜色传感器等.根据需求,本文使用 2 个大型电机和 1 个超声波传感器来测量 EV3 的速度.这 3 个传感器均可提供 10Hz 的测量.

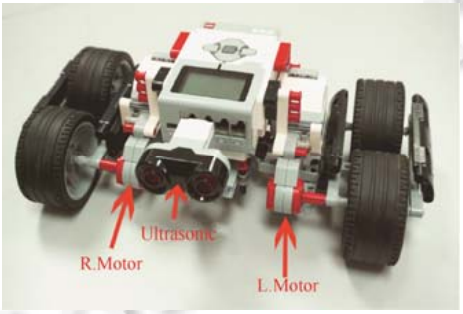

Fig.5 EV3 robot

图 5 EV3 机器人

为了获得鲁棒的瞬态故障模型参数,本文根据第 3.3 节提出的参数选择方法选择最终的实验参数.具体是,根据制造商提供的测量误差设置传感器的瞬态故障模型的参数 δ_i.对某个确定的窗口 w_i,通过真实的训练数据来确定 f_i 的值.本文首先以 0.7m/s 的恒定速度驱动 EV3 机器人直线运动,每个传感器收集 400 个测量数据.利用第 3.3 节提出的参数选择方法进行实验,为了模拟真实的攻击情形,窗口 w_i 中每个传感器提供的故障数是随机的.根据实验结果,本文建立了如图 6 所示的 ROC 曲线来确定 f_i 的值.ROC 曲线的 x 轴表示误报率(误报的数量/识别的总数量),y 轴表示识别率((识别的总数–误报的总数)/测试总数).为了避免混淆,图 6 中不包括 w_{400} 的情况.从图中可以看出,左上角的数据点具有更高的识别率、更低的误报率,该点对应的 f_i 是最佳阈值.表 2 总结了可变窗口大小的故障模型参数,其中,w_i 表示窗口大小为 i 的检测器.

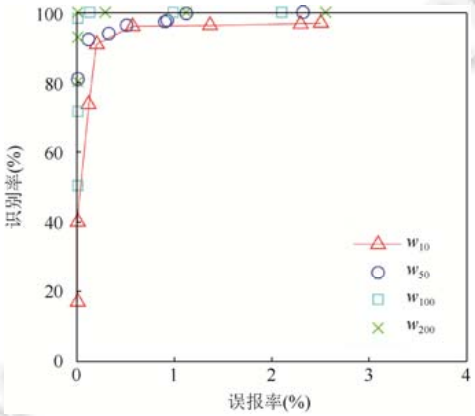

Fig.6 ROC curve: Identification rate and false alarm rate under three types of attacks

图 6 ROC 曲线:3 种类型攻击下的识别率和误报率

Table 2 Fault models for the sensors on EV3

表 2 EV3 上瞬态故障模型的参数

Detector	Ultrasonic		L.Motor		R.Motor	
	δ	f	δ	f	δ	f
W_{10}	0.2	1	0.037	1	0.037	2
W_{100}	0.2	20	0.037	16	0.037	18
W_{200}	0.2	31	0.037	20	0.037	24
W_{300}	0.2	40	0.037	35	0.037	30
W_{400}	0.2	64	0.037	45	0.037	50

4.2 检测性能分析

为了评估本文在第 3 节中介绍的 BPI 算法,本节首先在平坦的地面上以恒定的速度驱动 EV3 来收集每个传感器的数据.本文收集了 20 次来自 EV3 设备上的未被攻击的传感器数据,其中,所有传感器以 10Hz 平均采样 40s,因此,每个传感器有 400 个测量值.

本文假设在固定窗口中,攻击者对 3 个抽象传感器中的 1 个进行攻击,具体哪一个传感器被攻击是未知的.同时,对于被攻击的传感器,本文添加第 1.4 节中设计的 3 种攻击类型.注意,被攻击的传感器不要求在每一轮中都提供故障的测量,每个窗口中传感器提供故障测量的数量是随机的,但需要保证在给定窗口中是非瞬态故障.此外,在所有的检测数据中,其他传感器可能存在瞬态故障,也可能不存在瞬态故障.表 3 给出了不同窗口和攻击场景下 3 种攻击检测和识别方法的检测率.其中,KF 方法是一种基于卡尔曼滤波器的传感器攻击检测方法.从表中可以看出,本文提出的 BPI 算法能够比已有的算法检测到更多的攻击.对于 3 种攻击,BPI 方法在任何窗口中都比 PI 和 KF 方法更加鲁棒,并且随着窗口大小的增加,BPI 方法逐渐达到稳态检测率.对于偏差攻击,KF 方法与 BPI 的检测性能相近,BPI 检测器的平均检测率比 PI 检测器大约高 25%,比 KF 方法高约 2%.然而,对于随机攻击,KF 方法的检测性能明显低于 BPI 方法,小窗口中 BPI 检测器比 PI 检测器平均约高 53%,在大窗口中约高 35%,平均比 KF 方法高 14%.特别地,对于隐身攻击来说,BPI 方法的优势是显而易见的.现有方法的检测率几乎是 0,而 BPI 方法能够检测到攻击,其检测率平均能达到 90%以上.PI 方法不能检测到隐形攻击,是因为隐身攻击的特点就是尽可能地最大化融合间隔并使当前时刻任意两个传感器之间的间隔尽可能地相交.而 PI 方法恰好是基于两个传感器之间的间隔不相交来判断故障的,所以,基于 PI 的方法无法检测到这种攻击.基于 KF 的方法在大的窗口中偶尔可以检测到攻击,但检测率仅有 0.01%左右.并且,其误报率很高,这将在下一小节中详细加以介绍.

Table 3 Detection rate

表 3 检测率

(a) 偏差估计

检测率(%)	W=10	W=100	W=200	W=300	W=400
PI	54.6	63	79.82	83.5	88.97
KF	99.7	99.48	94.16	96.21	98.1
BPI	99.68	99.71	99.86	99.95	99.99

(b) 随机攻击

检测率(%)	W=10	W=100	W=200	W=300	W=400
PI	31.53	38.97	45.89	55.79	71.73
KF	86.22	69.67	65.54	87.01	94.58
BPI	86.59	92	97.46	99.18	99.89

(c) 隐身攻击

检测率(%)	W=10	W=100	W=200	W=300	W=400
PI	0	0	0	0	0
KF	0	0	0.01	0.02	0.021 4
BPI	89.36	93.23	93.69	96.45	98.59

4.3 识别性能分析

本文提出的 BPI 算法的识别性能与检测性能几乎相同,但是识别比检测需要花费更长的时间.表 4 显示了

不同窗口和攻击情形下 3 种传感器攻击检测方法的识别率.这些结果表明,对于偏差攻击和随机攻击,除了 KF 方法外,其他两种方法的识别率通常随窗口大小而有所改善.从表 4(a)可以看出,对于偏差攻击,BPI 方法的识别率大约平均比 PI 方法高 30%左右,仅比基于卡尔曼的方法高 3.3%左右.对于随机攻击来说,BPI 方法大约平均比 PI 方法高 49%左右,比基于 KF 的方法约高 14%左右.从表 4(b)可以看出,基于 KF 的方法对于随机攻击的识别性能明显降低.这是由于,随机攻击为被攻击的传感器在被攻击的时刻随机添加一个 0~0.7 的一个偏差值,当这个偏差值较小时,由于测量值和估计值的偏差比较小,就会导致 KF 方法无法检测到故障.特别地,从表 4(c)中可以看出,对于隐身攻击,基于 PI 的方法的识别率为 0,基于 KF 的方法的识别率在大的窗口中偶尔能够检测到 1 次攻击,仅 0.01%左右,并且其误报非常高.然而,本文提出的 BPI 方法平均可达到 93%.

Table 4 Identification rate

表 4 识别率

(a) 偏差估计

识别率(%)	W=10	W=100	W=200	W=300	W=400
PI	53	59.75	76.09	79.1	80.67
KF	99.7	99.48	94.16	96.21	98.1
BPI	99.48	99.64	99.68	99.78	99.99

(b) 随机攻击

识别率(%)	W=10	W=100	W=200	W=300	W=400
PI	29.25	38.92	41.65	50.9	68.37
KF	86.22	69.67	65.54	87.01	94.58
BPI	86.54	92.1	96.8	98.41	99.17

(c) 隐身攻击

识别率(%)	W=10	W=100	W=200	W=300	W=400
PI	0	0	0	0	0
KF	0	0	0.01	0.02	0.021 4
BPI	86.33	90.87	95.54	96	97.11

4.4 误报分析

表 5 总结了不同窗口大小中 3 种方法的误报率.实验结果表明,3 种方法都存在误报,这可能是由于这些窗口中存在瞬态故障的原因.此外,我们注意到,BPI 方法的误报率略高于 PI 方法.主要原因是瞬态故障的存在不能保证每轮被损坏的传感器数量不超过 f,即无法保证融合间隔一定包含真实值.此外,由于使用传感器历史测量,在某些攻击情形中可能由于从根据历史测量预测的当前测量不够准确,导致误报增加.然而,从表中可以看出,KF 方法的误报率非常高.其出现误报的原因主要是由于瞬态故障的存在,导致 KF 方法产生残留污染和残余淹没,从而造成误报.此外,需要注意的是,对于偏差攻击和随机攻击,KF 方法的误报率与 BPI 相近,但对于隐身攻击,其误报率能达到 50%以上.

Table 5 False rate

表 5 误报率

误报率(%)	W=10	W=100	W=200	W=300	W=400
PI	0.75	0.14	0	0	0
KF	0.41	7.39	94.18	51.71	45.87
BPI	1.05	0.26	1.31	0.82	1.05

5 总 结

本文研究了在存在瞬态故障时 CPS 的安全问题.首先在 Marzullo 提出的经典融合算法的基础上,通过融入历史测量提出了一种改进的融合算法,该算法可以得到更精确的融合间隔,具有更强的鲁棒性.此外,本文结合历史测量和融合间隔提出了一种新颖的传感器攻击检测和识别方法,用于具有测量相同物理变量的多个传感器的 CPS.并且,提出了一种基于构建 ROC 曲线的方法来选择瞬态故障模型.最后在 EV3 平台上获得实际的实验数据,验证了算法的性能,并与现有的基于卡尔曼滤波器的方法和基于 PI 的方法进行了比较.实验结果表明,

该算法在各种攻击场景下均优于现有的算法.基于本文的评估,未来的工作包括尝试使用形式化验证方法来验证所提出算法的正确性,并将该算法部署在实际系统上进行工业实践.

References:

[1] Miao F, Zhu Q, Pajic M, Pappas GJ. Coding schemes for securing cyber-physical systems against stealthy data injection attacks. IEEE Trans. on Control of Network Systems, 2017,4(1):106–117.

[2] Kim KD, Kumar PR. Cyber-physical systems: A perspective at the centennial. Proc. of the IEEE, 2012,100:1287–1308.

[3] Kong LL. Analysis of deception models and detection algorithms on CPS control layer [MS. Thesis]. Shanghai: East China University of Science and Technology, 2015 (in Chinese with English abstract).

[4] Jiang Y, Song H, Wang R, Gu M, Sun J, Sha L. Data-centered runtime verification of wireless medical cyber-physical system. IEEE Trans. on Industrial Informatics, 2017,13(4):1900–1909.

[5] Jiang Y, Zhang H, Song X, Jiao X, Hung WNN, Gu M, Sun J. Bayesian-network-based reliability analysis of plc systems. IEEE Trans. on Industrial Electronics, 2013,60(11):5325–5336.

[6] Yang K, Wang R, Jiang Y, Luo C, Guan Y, Li X, Shi Z. Enhanced resilient sensor attack detection using fusion interval and measurement history. In: Proc. of the 2018 Int'l Conf. on Hardware/Software Codesign and System Synthesis (CODES+ISSS). 2018. 1–3. [doi: 10.1109/CODESISSS.2018.8525941]

[7] Cardenas AA, Amin S, Sastry S. Secure control: Towards survivable cyber-physical systems. In: Proc. of the Int'l Conf. on Distributed Computing Systems Workshops. IEEE, 2008. 495–500.

[8] Checkoway S, Mccoy D, Anderson D, Kantor B, Shacham H, Savage S, Koscher K, Czeskis A, Roesner F, Kohno T. Comprehensive experimental analyses of automotive attack surfaces. In: Proc. of the Usenix Conf. on Security. 2012. 6.

[9] Koscher K, Czeskis A, Roesner F, *et al*. Experimental security analysis of a modern automobile. IEEE Journal of Selected Topics in Quantum Electronics, 2010,41(3):447–462.

[10] Slay J, Miller M. Lessons learned from the maroochy water breach. In: Proc. of the Int'l Conf. on Critical Infrastructure Protection. 2007. 73–82. [doi: 10.1007/978- 0-387-75462-8_6]

[11] Farwell JP, Rohozinski R. Stuxnet and the future of cyber war. Survival, 2011,53(1):23–40.

[12] Xiao L, Boyd S, Lall S. A scheme for robust distributed sensor fusion based on average consensus. In: Proc. of the Int'l Symp. on Information Processing in Sensor Networks. IEEE, 2005. 9.

[13] Olfati-Saber R, Shamma JS. Consensus filters for sensor networks and distributed sensor fusion. In: Proc. of the IEEE Conf. and the European Control Conf. on Decision and Control, CDC-ECC 2005. 2006. 698–6703.

[14] Yang K, Wang R, Jiang Y, Song H, Luo C, Guan Y, Li X, Shi Z. Sensor attack detection using history based pairwise inconsistency. Future Generation Computer Systems, 2018,86:392–402.

[15] Marzullo K. Tolerating failures of continuous-valued sensors. ACM Trans. on Computer Systems, 1990,8(4):284–304.

[16] Ivanov R, Pajic M, Lee I. Attack-resilient sensor fusion for safety-critical cyber-physical systems. ACM Trans. on Embedded Computing Systems, 2016,15(1):1–24.

[17] Ivanov R, Pajic M, Lee I. Resilient multidimensional sensor fusion using measurement history. In: Proc. of the Int'l Conf. on High Confidence Networked Systems. 2014. 1–10.

[18] Kalman RE. A new approach to linear filtering and prediction problems. Journal of Basic Engineering Transactions, 1960, 82(Series D):35–45.

[19] Kwon C, Hwang I. Security analysis for cyber-physical systems against stealthy deception attacks. In: Proc. of the American Control Conf. IEEE, 2013. 3344–3349.

[20] Jayasimha DN. Fault tolerance in a multisensory environment. In: Proc. of the 13th Symp. on Reliable Distributed Systems, SRDS'94. 1994. 2–11.

[21] Park J, Ivanov R, Weimer J, *et al*. Sensor attack detection in the presence of transient faults. In: Proc. of the 6th ACM/IEEE Int'l Conf. on Cyber-physical Systems. ACM, 2015. 1–10.

[22] Willsky AS. A survey of design methods for failure detection in dynamic systems. Automatica, 1975,12(6):601–611.

[23] Shoukry Y, Martin P, Tabuada P, Srivastava M. Non-invasive spoofing attacks for anti-lock braking systems. In: Proc. of the Int'l Conf. on Cryptographic Hardware and Embedded Systems. Springer-Verlag, 2013. 55–72.

附中文参考文献:

[3] 孔令霖.CPS 控制层欺骗攻击模型与检测算法的研究[硕士学位论文].上海:华东理工大学,2015.

杨康(1992－),女,山东菏泽人,硕士,主要研究领域为 CPS 的安全性,形式化验证.

李晓娟(1968－),女,博士,教授,CCF 专业会员,主要研究领域为系统形式建模与验证,机器人系统软件安全,计算机网络协议分析.

王瑞(1981－),女,博士,副教授,CCF 专业会员,主要研究领域为形式化方法.

施智平(1974－),男,博士,教授,CCF 高级会员,主要研究领域为形式化,人工智能.

关永(1966－),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为形式化验证,系统可靠性,嵌入式系统.

Xiaoyu Song(1963－),男,博士,教授,博士生导师,主要研究领域为形式化方法.