面向传感器攻击的概率时间窗感知融合算法研究

陈彦峰 邓庆绪 张天宇 孙 磊
(东北大学计算机科学与工程学院 沈阳 110169)

摘 要 信息物理系统需要部署传感器对真实物理状态进行测量,通过网络传输到控制器实现各类功能.传感器攻击使得控制器接收的物理状态测量值存在较大误差,不能准确反映真实物理状态,从而促使控制器生成错误的决策和控制指令.除传感器攻击外,物理状态估测的准确性还受到两个因素影响:传感器测量过程中存在随机噪声,测量值在网络传输过程中存在随机延迟.针对以上因素,本文设计了能够抵抗传感器攻击的概率时间窗感知融合算法,对随机延迟进行补偿并计算在时间窗内不同传感器的正确概率累积值,进而作为权重调整卡尔曼滤波更新值,减小受攻击的传感器测量值带来的负面影响.自动驾驶车队仿真结果表明,在噪声和延迟的概率信息准确的条件下,与传统卡尔曼滤波、欧拉卡尔曼滤波、间隔融合等算法相比,本文所提出的延迟补偿-概率时间窗卡尔曼滤波算法具有最小的测量误差,物理状态估测累积误差较传统卡尔曼滤波降低67%;即使攻击者完全掌握本文融合算法对传感器进行自适应攻击,本文算法能够将融合值误差控制在可接受范围,系统仍能正常工作.

关键词 信息物理系统;随机延迟;随机噪声;概率论;时间窗;传感器攻击;感知融合
中图法分类号 TP393 **DOI号** 10.11897/SP.J.1016.2023.01227

Research on Probability-Time-Window Sensor Fusion Algorithm for Sensor Attack

CHEN Yan-Feng DENG Qing-Xu ZHANG Tian-Yu SUN Lei
(*School of Computer Science and Engineering, Northeastern University, Shenyang* 110169)

Abstract Cyber-physical systems deploy sensors to measure the real physical state, and transmit their measurements to the controller through the network to achieve multiple functions. The sensor attack causes a large error in the physical state measurement value received by the controller, which cannot accurately reflect the real physical state, thus prompting the controller to make wrong decisions and control instructions. This directly affects the normal work of the system and threatens the related industrial production, personal and property safety. The security of cyber-physical systems is facing the increasingly serious problem caused by sensor attacks, and the research on related anti-attack algorithms is of great significance. This paper mainly focuses on sensor attacks, and studies how to improve the accuracy of physical state estimation through sensor fusion algorithm. Besides sensor attacks, there are two factors that affect the accuracy of physical state estimation: random noise in the sensing, and random delays in the network transmission. In view of the above factors, this paper designs a probability time window sensor fusion algorithm which can resist sensor attacks. Firstly, on the basis of using Kalman filter to deal with random noise, the corresponding state estimation value and probability are calculated by compensating the random communication delay. Then, the probability cumulative value of each sensor in the time window is used to evaluate the confidence of the normal operation of the sensor. Finally, the normal working confidence of each sensor is taken as the weight, and the physical state

收稿日期:2022-01-24;在线发布日期:2023-01-12.本课题得到国家自然科学基金地区联合基金重点项目(U1908212)、国家自然科学基金面上项目(62072085)、国家自然科学基金(62102073)和兴辽人才计划项目(XLYC1902017)资助.**陈彦峰**,博士研究生,主要研究方向为嵌入式系统.E-mail:yfchen0@163.com.**邓庆绪**(通信作者),博士,教授,中国计算机学会(CCF)杰出会员,主要研究领域为实时嵌入式系统、物联网安全.E-mail:dengqx@mail.neu.edu.cn.**张天宇**,博士,主要研究方向为嵌入式系统.**孙 磊**,博士研究生,主要研究方向为嵌入式系统.

estimation value is updated by Kalman filter to reduce the negative impact brought by the measured value of the attacked sensor. The autonomous vehicle platoon MATLAB simulation results show that under the condition that the probability information of noise and delay is accurate, the algorithm proposed in this paper has the lowest sensing error, compared with ordinary Kalman filter, Euler-Distance Kalman filter and Interval Fusion. And the cumulative error of physical state estimation is 67% lower than that of ordinary Kalman filter. In addition, we evaluate the performance of the proposed algorithm under inaccurate probability information of noise and delay with different degrees of error, including the range and the parameters of distribution functions. The results show that all relative algorithms need relatively accurate information to generate acceptable estimation of the physical state to make the autonomous vehicle platoon work normally. The proposed algorithm can tolerate the maximum error of the probability information among the above methods. When the relative error between the communication delay time range obtained by the controller and the real range is ±60%, the proposed algorithm can still achieve higher fusion accuracy than the traditional Kalman filter algorithm. If the attacker obtains the full-knowledge of the proposed method and launches adaptive sensor attacks, it will pre-check whether the designed attack vector can be detected by the proposed algorithm. To keep stealthy, the fused errors are limited to acceptable range under which the attacked sensor will still be regarded as normal sensor. In this condition, the maximum negative affect the attacker can produce is the fused value skipping between extreme points of the acceptable range, under which the system can still work normally.

Keywords cyber-physical system; random delay; random noise; probability theory; time window; sensor attack; sensor fusion

1 引 言

信息物理系统综合计算、网络、控制,实现信息世界和物理世界之间的感知、控制和信息服务,涉及到军事、工业、生活等领域.为实现信息世界和物理世界之间的交互,大量的传感器设备被部署到系统中来对相关物理状态进行测量.传感器的测量值是信息物理系统感知物理世界状态的最原始数据,通过网络传输到控制器,经过特定的控制、服务算法实现各类功能.对真实物理状态进行准确测量是信息物理系统做出正确决策和控制的基础[1-2].一旦信息物理系统控制器接收到的传感器数据不准确,不能真实反应实际物理状态,控制器就会做出错误决策并发出错误的控制指令.恶意传感器攻击直接造成控制器接收的测量数据被篡改,轻则影响系统性能,重则直接威胁用户的生命财产安全[3-5].例如,目前已有相关研究指出了电力设施[6]、军事无人机[7]、自动驾驶汽车[8]等典型信息物理系统受到传感器攻击造成的灾难性后果.对物理状态进行高准确性测量是保证信息物理系统正常工作的基础,而传感器攻击直接影响测量的准确性.研究面向传感器攻击的感知融合算法,实现对物理状态高准确性的测量,对保护信息物理系统安全具有重要意义.

传感器攻击的最终目的是使控制器接收到错误的传感数据,进而生成错误的决策和指令.传感数据既包括原始的传感器测量值,也包括时间戳、通信密钥等附加数据.传感器攻击可能发生在传感测量阶段,通过物理方式直接干扰传感器的测量,强制传感器生成错误的测量值;也可能针对通信网络,在数据传输过程中进行监听、拦截、篡改.通信加密算法可以解决针对网络的攻击,但是难以应对针对传感器的物理方式攻击.以自动驾驶车队这一典型的信息物理系统为例,该系统由多辆自动驾驶汽车组成,每辆汽车配有 LiDAR(Light Detection and Ranging)、摄像头、雷达等传感器设备.自动驾驶车队保持恒定车距行驶在道路上,汽车之间通过无线网络共享各自的速度、加速度、车距等状态信息.攻击者可以通过物理方式发射虚假反射波,直接控制某辆或多辆自动驾驶汽车的传感器设备输出错误的测量数据,也可以直接对车队内车辆间的通信进行拦截篡改,进而使车辆接收错误的状态信息,迫使车队内

前后车辆做出瞬间加速、紧急刹车等危险动作，造成重大生命财产安全损失[9-11].

感知融合算法通过在系统内部署多个传感器对相同物理状态进行直接或间接的冗余测量，采用合适的攻击检测处理算法，利用多传感器数据间的约束关系，检测、处理可能被攻击的传感器数据，融合经过处理后的正确数据对真实物理状态进行高准确性的估测，从而有效抵御传感器攻击[12-13].从控制器的角度，利用接收到的测量数据进行感知融合后生成的状态估测值和真实物理状态之间的误差是否在可接受范围内，是衡量融合估测值是否准确的标准.然而传感器对真实物理状态的测量得到的数据本身包含一定的自身误差，导致攻击检测处理算法很难辨别多传感数据间的差异是攻击造成的还是传感器的自身误差，给感知融合带来不利影响.

传感数据的自身误差主要来源于两个环节：(1)传感器在测量过程中存在一定的随机噪声；(2)数据从传感器产生到通过网络传输到控制器进行处理之间存在一定的随机延迟.在多数实际系统中，高斯噪声是随机噪声最为广泛的分布形式，可以通过适当的滤波算法进行处理；而随机延迟受到通信协议、网络状态、外部干扰等因素的影响，不同系统间的延迟存在较大差异.虽然对于确定的实际系统，可以认为随机延迟也服从一定的概率分布特征，但与随机噪声不同，并不是多数随机延迟都服从零均值高斯分布[14].只有合理处理随机噪声和随机延迟，在进行攻击检测处理算法前，先对随机噪声和随机延迟产生的误差进行适当补偿，才能实现更高准确性的物理状态感知融合估测.因此，面向传感器攻击的感知融合算法需要处理 3 个因素：(1)随机噪声；(2)随机延迟；(3)传感器攻击.

针对随机噪声，目前已有众多成熟的研究成果，通过各种数字滤波滤除服从特定分布特征的随机噪声.以最广泛存在的零均值高斯噪声为例，基于卡尔曼滤波的算法可以有效滤除服从此类分布的随机噪声.与随机噪声不同，随机延迟不符合零均值高斯分布.把随机延迟视为随机噪声，用传统的卡尔曼滤波算法进行处理，融合值的精度会有所降低.为保证信息物理系统的控制精度，必须加入合适的随机延迟处理算法.延迟的处理算法可以分为两大类：其中一类是通过同步时钟时间戳或者其他方式直接对延迟进行在线估测；另外一类是通过建模、实验等方式，获得某一确定系统随机延迟的概率分布特征，利用该信息进行延迟补偿.前类算法可以直接准确地处理延迟，但是考虑到攻击者可以监听、篡改时间戳，或直接进行时钟同步攻击[15]，因此在面向传感器攻击的感知融合中具有有限的有效性.通过概率信息处理延迟，利用确定系统固有的通信特征，对所有可能范围内的延迟进行一定范围的补偿，能够应对多种方式的传感器攻击.但是此类方法的问题是：一方面，延迟噪声的概率分布信息可能存在一定误差；另一方面，系统可能受未知扰动的影响，数据在某些时刻存在较大的噪点.这两方面因素会影响攻击检测处理的准确性.

攻击检测处理算法是面向攻击感知融合算法的核心.目前对攻击的检测多基于多传感数据的一致性，利用冗余传感数据间的约束条件，通过投票法等方式对传感数据进行范围、特征层面的检测，进而识别出可能受到攻击的传感数据.被攻击传感数据的处理可以分为两大类：一类是直接丢弃被攻击数据[12-13]；另外一类是在丢弃源数据基础上，利用对应传感器的历史数据基于物理模型进行预测补偿[16].前者具有简单稳定的特点，在传感数据冗余度足够的条件下，可以很好地满足感知融合的要求；后者相对复杂，且依赖于物理模型，实现较为复杂，但是在传感数据冗余度较低时，能够实现更好的效果.

目前，缺少能够同时处理随机噪声和随机延迟的面向传感器攻击的感知融合算法研究.基于卡尔曼滤波的算法可以有效处理零均值高斯噪声，但是无法处理所有的随机延迟；基于时间戳延迟估测的算法无法应对时钟同步攻击或可篡改时间戳的攻击；基于延迟概率信息的算法在应对概率信息误差、未知扰动方面适应性较差.针对上述问题，本文提出了一种基于概率时间窗的抗攻击感知融合算法，结合了传感器攻击检测和卡尔曼滤波算法，设计了传感器在时间窗内正常工作概率评估方法，并将其作为该传感器数据的时间窗权重来计算感知融合估测值.首先，利用随机延迟的概率分布信息，对各传感器的测量值进行可能范围内的延迟补偿，获得各传感器在不同延迟补偿下对应的测量值并计算出相应的概率信息；然后，用历史融合值和各传感器测量值的概率信息评估该传感器正常工作的概率，将累积概率作为传感器数据的权重；最后，结合各传感器的时间窗权重对卡尔曼滤波的状态估测值进行更新，获得真实物理状态的融合估测值.

本文的主要贡献如下：

(1)在卡尔曼滤波处理随机噪声基础上，设计了利用随机延迟的概率分布信息进行延迟补偿的机

制,利用一个时间窗内卡尔曼滤波估测对应概率的累积值作为评估传感器正常工作的可能性,在卡尔曼滤波的状态预测环节,将该累积值作为更新状态估测值的权重,减小受攻击数据带来的负面影响;

(2)在仿真软件中建立自动驾驶车队控制模型,对本文所提出的感知融合算法和其他相关抗攻击融合算法进行了对比仿真,分析验证了噪声和延迟的概率分布信息误差对各种算法融合精度的影响,对比了不同算法的容错性,验证了自适应攻击下该方法的有效性.

本文第 2 节主要介绍已有基于卡尔曼滤波和基于间隔的面向传感器攻击的感知融合算法;第 3 节说明了所研究问题的背景和数学模型;第 4 节分析传统卡尔曼滤波和间隔算法的适用性;第 5 节详细介绍了基于概率时间窗的抗攻击融合算法;第 6 节对所提出的算法和其他相关算法进行了对比仿真验证;第 7 节对文章进行了总结和展望.

2 相关工作

面向传感器攻击的感知融合算法需要处理 3 个因素:随机噪声、随机延迟和传感器攻击.针对最广泛存在的高斯噪声,卡尔曼滤波是最有效的解决算法之一.然而,卡尔曼滤波本身不具备处理非高斯分布的随机延迟的能力,必须与随机延迟处理算法相结合,才能实现能够抵抗传感器攻击的高准确性感知融合的目标.在处理随机噪声并进行随机延迟补偿后,再进行相应的攻击检测处理,进而实现抗攻击的感知融合.

随机延迟的处理可以分为两大类:一类是直接根据同步时钟下的时间戳等方式估测随机延迟;另外一类是根据系统的随机延迟概率分布特征,进行所有可能范围的延迟补偿.Zhao 等人将传统卡尔曼滤波器扩展到 2 维空间,建立了 2 维正则化最小二乘估测模型,利用历史方差对随机延迟产生的误差进行最小化处理[17].Wang 等人在多采样频率和通信延迟条件下,利用被控系统的数学模型,对传感器测量值进行了不确定补偿[18].Ravi 等人研究了不同采样频率和通信延迟下的状态估测算法,通过一段时间内滚动预测的方式实现了对状态量的平滑估测[19].Dorigoni 等人对一个时间窗内的卡尔曼滤波矩阵进行观测,在状态估测中引入系统不确定性处理随机延迟的影响[20].然而,上述随机延迟的相关卡尔曼滤波算法均未考虑任意传感器攻击的影响.

面向传感器攻击,目前有学者基于卡尔曼滤波算法进行了相关研究.其中传感器攻击检测的原理都是基于各传感器测量数值间的一致性进行的,通过历史数据、累积方差等指标识别传感器攻击.Bai 等人分析了卡尔曼滤波框架下的攻击行为,评估了攻击可以造成的最坏影响,给出了最优攻击策略[21].Chang 等人在智能电网背景下,利用系统数学模型设计了一个预测误差评估值,在此基础上提出了结合卡尔曼滤波和安全估计的感知融合算法[22].然而,常规基于卡尔曼滤波的攻击检测算法其基于方差检测器,假设攻击数据服从高斯分布.在实际系统中的攻击可能是任意的,导致常规基于方差检测的算法难以应对任意攻击.Manandhar 等人提出了欧拉距离检测算法,对传感器数据进行一致性计算,以应对复杂的虚拟数据注入攻击[23].然而上述算法未考虑随机通信延迟的影响,且未对其和随机延迟算法相结合的可能性做出深入讨论或给出相应方式.

目前,同时考虑噪声、延迟和攻击的感知融合算法研究很少.Marzullo 在 1990 年提出的基于数据间隔的融合算法中,综合考虑了测量误差和通信延迟,依据多数投票原则设计了攻击检测和融合算法,根据误差和延迟可能的最大范围,计算出真实物理状态可能的范围间隔.在多个传感器形成的范围间隔之间,设计间隔的重合部分作为最终融合估测值[12].Ivanov 在 Marzullo 的算法基础上进行了改进,在 2015 年提出了基于时间窗的攻击检测算法,改善了攻击检测过程中未知干扰的影响[13].在 2016 年对 Marzullo 的间隔算法中的攻击处理环节做了改进,在检测到攻击传感数据后,利用历史数据对当前被攻击数据进行估测替换,然后进行数据融合,而不是直接丢弃被攻击数据[16].

据我们所知,Marzullo 的间隔算法及其优化算法是目前少数综合考虑噪声、延迟和攻击的算法之一,是与本文所研究问题最直接相关的研究.然而该算法的主要问题是,针对噪声和延迟的分析基于限定值计算所有可能情况时,没有考虑噪声和延迟的随机性.同时,噪声和延迟的限定值可能存在一定误差,在传感攻击检测处理环节可能造成误判断.虽然 Ivanov 提出的基于时间窗的优化算法在一定程度上改善了此类干扰的影响,但是该方法需要设定合理的判断阈值.阈值直接影响最终的判断结果,阈值设定没有可靠标准.

本文综合考虑随机噪声和随机延迟,利用卡尔曼滤波对服从高斯分布的随机噪声进行处理,利用随机延迟的概率分布信息进行延迟补偿,根据多传

感器之间的数据关系和对应传感数据概率对传感数据正常工作概率进行评估. 结合各传感数据正常工作概率在时间窗内的累积值，进行定量的融合估测，无需设置判定阈值.

3　背景和模型

本节将介绍信息物理系统的组成框架，说明本文所研究多感知融合算法的背景，介绍其数学模型，为后文分析问题提出算法做基础.

3.1　数学模型

信息物理系统的框架如图 1 所示，真实世界的物理系统状态 x_k 由 N 个传感器进行测量. 考虑到系统的通用性，N_l 个系统内部传感器配置在与控制器相同的位置，$N-N_l$ 个系统外部传感器以分布式的方式配置在与控制器不同的位置. 利用多个传感器对同一状态量进行测量并引入系统外部传感器数据，可以有效提高测量精度和稳定性[1]. 例如在自动驾驶领域，在多辆智能车组成车队行进时，会共享各车的速度、加速度等信息，从而达到更稳定的车队控制效果[24-25]. 对于每辆车来说，内部传感器即为本车上的传感器，其测量值通过有线或无线通信方式传递到控制器；外部传感器为其他车辆上的传感器，通过无线通信的方式传输到本车控制器.

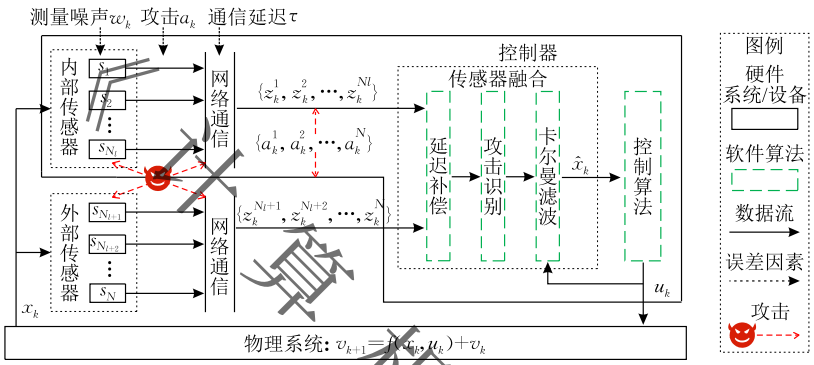

图 1　典型信息物理系统框架

(1) 物理系统

给定一个物理系统，其离散状态方程为

$$x_{k+1}=Ax_k+B\boldsymbol{u}_k+v_k \tag{1}$$

式中 x_k 是系统在第 k 时刻的状态量，$\boldsymbol{u}_k$ 是对应时刻的输入控制向量，v_k 是符合高斯分布的过程白噪声，$v_k\sim\mathcal{N}(0,Q_k)$，$A$ 是第 $k+1$ 时刻状态量 x_{k+1} 同第 k 时刻状态量 x_k 之间的关系，B 是第 $k+1$ 时刻状态量 x_{k+1} 同控制输入向量间 $\boldsymbol{u}_k$ 的关系.

(2) 随机噪声

N 个传感器对上述系统的测量模型为

$$z_k^i=\boldsymbol{H}^i(x_k)+w_k^i \tag{2}$$

式中 z_k^i 是第 i 个传感器 ($i\in[1,N]$) 对第 k 时刻状态量的测量值，$\boldsymbol{H}^i$ 是第 i 个传感器的状态测量矩阵，w_k^i 是其测量噪声，服从高斯分布 $w_k^i\sim\mathcal{N}(0,R_k^i)$，$(w_k^i\in[-e_u^i,+e_u^i])$，其中 e_u^i 为随机噪声绝对值最大值. 根据高斯分布特征，取 $e_u^i=3\sqrt{R_k^i}$，此时噪声在 $[-e_u^i,+e_u^i]$ 范围内对应概率为 99.7%.

(3) 随机延迟

如果传感器测量值在生成的一瞬间就被控制器接收到，那么传统的卡尔曼滤波或者其他滤波算法就可以用来对当前状态量进行有效的正确估计. 然而，在信息物理系统通信网络中通常含有网络延迟 $\tau_k^i, i\in[1,N]$，且一般具有随机性. 导致控制器在第 k 时刻接收到的测量值实际是 $k-\tau_k^i$ 时刻由传感器生成的：

$$z_k^i=\boldsymbol{H}^i(x_{k-\tau_k^i})+w_{k-\tau_k^i}^i \tag{3}$$

式中 τ_k^i 是系统的通信随机延迟，一般可以通过离线测试或在线估测等算法获得随机延迟的概率分布信息，记为 $\tau_k^i\sim pt^i(\tau)$，$\tau_k^i\in[t_l^i,t_u^i]$. t_l^i,t_u^i 是确定系统随机延迟的最小/最大限定值，$0<t_l^i<t_u^i$.

(4) 系统正常工作条件

任何信息物理系统正常工作均需保证一定程度的传感数据准确性. 在不考虑攻击的前提下，多个传感数据融合后准确性高于其中最高准确性的传感数据[12,21,25]. 假设满足物理系统准确性要求的传感器最大协方差为 R_0，则 N 个传感器组成的传感器系统满足控制要求的充分条件为至少有一个传感器协方差小于 R_0：$\exists i\in[1,N], R_k^i\leq R_0$.

3.2　攻击模型

除随机延迟的影响外，系统可能受到攻击者的恶意攻击，攻击可以直接通过物理方式改变传感器的测量值[6]，也可以在通信过程中进行拦截篡改、延

迟、干扰[1].虽然攻击的方式有多种,但是最终的攻击效果都可以用一个攻击向量$\{\boldsymbol{a}_k^1,\cdots,\boldsymbol{a}_k^N\}$来表示,其为控制器接收数据和传感器生成数据间的差.在考虑攻击行为时,控制器实际接收的值为

$$z_k^i=\boldsymbol{H}^i(x_{k-\tau_k^i})+w_{k-\tau_k^i}^i+\boldsymbol{a}_k^i \tag{4}$$

为了实现更好的通用性,对攻击做如下假设:

A1. 针对某一特定传感器,可以在任意时刻对其进行任意的攻击,即不对攻击向量作任何形式的假设限制,$\boldsymbol{a}_k^i\in(-\infty,+\infty)$,且攻击既可能是连续的,也可能是间歇触发的.某传感器在某时刻受到攻击并不意味着在后续时刻一直受到攻击.这样的假设更具备通用性,对应的感知融合难度也更高.

A2. N 个传感器中,只有少于一半的传感器可能受到攻击,$countif(\boldsymbol{a}_k^i\neq\boldsymbol{0})<N/2$.做出该假设的原因是目前多数的传感器攻击检测算法基于数据间的约束关系,如果超过一半的传感数据被任意篡改,则攻击者可以完全控制最终的融合估测值,此时无法保证系统正常工作.Marzullo[12]、Park[13]、Yang[25]等学者分别在其研究中进行了理论论证且得出了相同的结论.当系统中半数以上传感器完全受攻击者控制时,物理状态的反馈值将被攻击者任意修改,且能保证不被攻击检测算法识别,可以直接导致系统崩溃[26],已超出了攻击检测、信息融合所能解决的范围,不在本文的考虑范围之内.

攻击目标是最大程度地干扰感知融合估测值的准确性,进而使控制器做出错误决策,并输出错误的控制指令.假设多传感器的信息融合算法为$\mathcal{F}(z_k^1,z_k^2,\cdots,z_k^N)=\hat{z}_k$,$\hat{z}_k$为融合值,融合误差为$e_k=\hat{z}_k-\boldsymbol{H}(x_k)$,$x_k$为真实物理系统状态;攻击检测算法为$\mathcal{D}(z_k^1,z_k^2,\cdots,z_k^N)=\{\delta_k^1,\delta_k^2,\cdots,\delta_k^N\}$,$\delta_k^i=0/1$,1表示第$i$个传感器被判定为受攻击,0表示被判定为正常工作.根据攻击者是否掌握整个系统的感知融合算法,将攻击类型分为图2所示的两种:

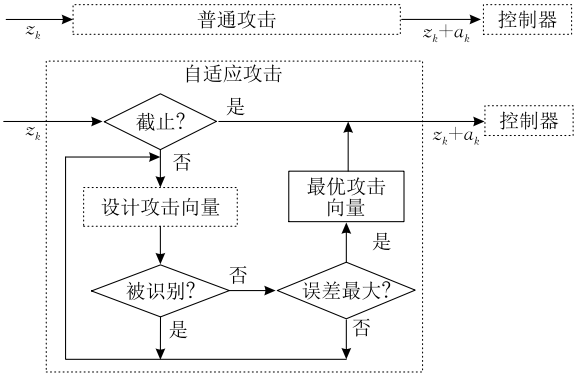

图2 攻击分类

普通攻击.攻击者未掌握信息融合算法$\mathcal{F}(\cdot)$和系统攻击检测处理算法$\mathcal{D}(\cdot)$,无法获取系统其他传感器数据$\{z_k^1,z_k^2,\cdots,z_k^N\}$,只能操纵被攻击传感器输出任意值,即$\boldsymbol{a}_k^i\in(-\infty,+\infty)$.

自适应攻击.攻击者掌握信息融合算法$\mathcal{F}(\cdot)$和系统攻击检测算法$\mathcal{D}(\cdot)$以及其他传感器数据$\{z_k^1,z_k^2,\cdots,z_k^N\}$,可以操纵被攻击传感器输出任意值,且能提前判断是否会被攻击检测算法识别,属于闭环攻击.会在保证不被识别的条件下,输出能够最大程度干扰感知融合估测值的篡改传感数据,如算法1所示.

算法1. 自适应攻击.

输入:融合算法$\mathcal{F}(\cdot)$,攻击检测算法$\mathcal{D}(\cdot)$,所有传感数据$\{z_k^1,z_k^2,\cdots,z_k^N\}$

输出:最优攻击向量$\{\boldsymbol{a}_k^1,\cdots,\boldsymbol{a}_k^N\}$

//初始化攻击向量,计算真实融合值

1. 生成$\{\boldsymbol{a}_k^1,\cdots,\boldsymbol{a}_k^N\}$,$\mathcal{F}(z_k^1,z_k^2,\cdots,z_k^N)=\hat{z}_k$,$e_{\max}\leftarrow0$
2. WHILE(未到截止时间)
3. 生成$\{\boldsymbol{a}_k^1,\cdots,\boldsymbol{a}_k^N\}$,$e\leftarrow\mathcal{F}(z_k^1+\boldsymbol{a}_k^1,\cdots,z_k^N+\boldsymbol{a}_k^N)-\hat{z}_k$

//判断攻击是否被识别

4. IF $\sum\mathcal{D}(z_k^1+\boldsymbol{a}_k^1,\cdots,z_k^N+\boldsymbol{a}_k^N)==0$

//判断造成的干扰是否最大

5. IF $e>e_{\max}$
6. $e_{\max}\leftarrow e$, $a_{opt}\leftarrow\{\boldsymbol{a}_k^1,\cdots,\boldsymbol{a}_k^N\}$
7. END WHILE
8. RETURN a_{opt}

3.3 问题描述

根据上述模型分析,本文的主要研究问题是:已知控制器在k时刻接收到存在延迟的数据$\{z_k^1,\cdots,z_k^N\,|\,z_k^i=\boldsymbol{H}^i(x_{k-\tau_k^i})+w_{k-\tau_k^i}^i+\boldsymbol{a}_k^i\}$,以及测量噪声和延迟的分布特征$w_k^i\sim\mathcal{N}(0,R_k^i)$,$\tau_k^i\sim pt^i(\tau)$.控制器需要采取感知融合算法,对当前物理状态$x_k$进行高准确性的估测.为实现该目标,需要解决以下3个具体问题:

P1. 如何消除高斯噪声$w_k^i\sim\mathcal{N}(0,R_k^i)$;

P2. 如何处理随机延迟$\tau_k^i\sim pt^i(\tau)$;

P3. 如何检测和处理攻击向量$\{\boldsymbol{a}_k^1,\cdots,\boldsymbol{a}_k^N\}$.

另外,随机噪声和随机延迟的处理都依赖于其概率分布信息,考虑到实际执行过程中控制器获取的概率信息可能存在一定误差,如何设计相应机制提高算法对信息误差的容错性也是需要考虑的问题.

需要说明的是,传感器故障不在本文研究范围内,即本文假设所有传感器自身均能正常工作.传感器不同故障状态下的行为有所不同,如断电/断通信等故障可以通过检测数据更新状态进行故障识别,

与本文研究的攻击有明显区别；测量误差超过标定范围则与本文研究的攻击有相似之处，可以通过相似算法进行识别. 本文不考虑传感器故障，但是所设计的算法可以与传感器故障检测算法相结合，能够处理更复杂情况.

4　传统算法及适用性分析

根据本文所研究的主要问题，需要分别对随机噪声、随机延迟和攻击进行处理. 目前已有相关研究解决了其中的某项问题，主要分为基于卡尔曼滤波和基于间隔的解决方法两大类，本节将给出相关传统算法并分析论证其适用性.

4.1　基于卡尔曼滤波的算法

卡尔曼滤波是处理高斯噪声最有效的方法之一，利用概率论推导出最大概率的补偿和估测，在估测时考虑了历史数据趋势，是应用最广泛的噪声处理算法. 在传统卡尔曼滤波基础上，衍生出针对延迟和攻击的相关算法.

(1) 随机噪声的处理

当系统中没有通信延迟时，可以直接通过分布式扩展卡尔曼滤波算法，利用历史数据对当前状态进行估测[9]，对第 i 个传感器预测方程：

$$\begin{aligned}&\hat{x}^i_{k|k-1}=A\hat{x}^i_{k-1|k-1}+Bu_{k-1},\\&P^i_{k|k-1}=AP^i_{k-1|k-1}A'+Q_k\end{aligned}\tag{5}$$

测量和校正方程：

$$\begin{aligned}&K^i_k=P^i_{k|k-1}\boldsymbol{H}^{i'}(\boldsymbol{H}^iP^i_{k|k-1}\boldsymbol{H}^{i'}+R^i_k)^{-1},\\&\hat{x}^i_{k|k}=\hat{x}^i_{k|k-1}+K^i_k(z^i_k-\boldsymbol{H}^i\hat{x}^i_{k|k-1}),\\&P^i_{k|k}=(I-K^i_k\boldsymbol{H}^i)P^i_{k|k-1}\end{aligned}\tag{6}$$

式中 $\hat{x}^i_{k|k}$ 是传感器 S_i 在第 k 时刻输出的状态估测值，$\hat{x}^i_{k|k-1}$ 是传感器 S_i 在第 k 时刻根据 $k-1$ 时刻估测值和系统状态方程做出的预测值，$P^i_{k|k}=E[(\hat{x}^i_{k|k}-x_k)(\hat{x}^i_{k|k}-x_k)']$ 是 $\hat{x}^i_{k|k}$ 相对于真实状态 x_k 的协方差，$P^i_{k|k-1}=E[(\hat{x}^i_{k|k-1}-x_k)(\hat{x}^i_{k|k-1}-x_k)']$ 是 $\hat{x}^i_{k|k-1}$ 相对于真实状态 x_k 的协方差. N 个传感器共同组成的分布式卡尔曼滤波测量模型为[9]

$$z_k=\boldsymbol{H}(x_k)+W_k\tag{7}$$

式中：

$$\begin{aligned}&z_k=(z^{1'}_k,\cdots,z^{N'}_k)',\\&\boldsymbol{H}=(\boldsymbol{H}^{1'},\cdots,\boldsymbol{H}^{N'})',\\&W_k=(w^{1'}_k,\cdots,w^{N'}_k)',\\&Cov(W_k)=R_k=diag(R^1_k,\cdots,R^N_k)\end{aligned}\tag{8}$$

控制器可以获得所有传感器的数据，设计状态量的分布式卡尔曼滤波融合估测算法[27]，预测方程：

$$\begin{aligned}&\hat{x}_{k|k-1}=A\hat{x}_{k-1|k-1}+Bu_{k-1},\\&P_{k|k-1}=AP_{k-1|k-1}A'+Q_k\end{aligned}\tag{9}$$

测量和校正方程：

$$\begin{aligned}&K_k=P_{k|k-1}\boldsymbol{H}'(\boldsymbol{H}P_{k|k-1}\boldsymbol{H}'+R_k)^{-1},\\&\hat{x}_{k|k}=\hat{x}_{k|k-1}+K_k(z_k-\boldsymbol{H}\hat{x}_{k|k-1}),\\&P^{-1}_{k|k}=P^{-1}_{k|k-1}+\sum_i^N P^{i-1}_{k|k}-P^{i-1}_{k|k-1}\end{aligned}\tag{10}$$

基于卡尔曼滤波的融合算法本质是根据历史测量数据计算各传感器当前测量值的协方差，协方差用以表征对应传感器数据的正确性，协方差越大表示正确性越小，对最终状态估测值的影响权重越小，且多感知融合值精度不低于所有传感器的最高精度[27-29]，即

$$\forall i\in[1,N],\ P_{k|k}\le P^i_{k|k}\tag{11}$$

(2) 不进行随机延迟补偿的影响

由于存在随机延迟，导致控制器当前所接收的数据实际上是延迟时间 τ^i_k 之前的测量值：$z^i_k=\boldsymbol{H}^i(x_{k-\tau^i_k})+w^i_{k-\tau^i_k}$，与当前时刻应收到的数值间存在误差：

$$\begin{aligned}\Delta z^i_{k|\tau^i_k}&=\boldsymbol{H}^i(x_k)-\boldsymbol{H}^i(x_{k-\tau^i_k})\\&=\frac{\boldsymbol{H}^i(x_k)-\boldsymbol{H}^i(x_{k-\tau^i_k})}{\tau^i_k}\tau^i_k\approx\frac{\mathrm{d}z^i_k}{\mathrm{d}t}\tau^i_k=\dot{z}^i_k\tau^i_k.\end{aligned}$$

测量噪声和通信延迟互为独立事件，因此延迟造成的测量噪声差值的协方差为

$$\begin{aligned}&Cov(\dot{z}^i_k\tau^i_k+w^i_{k-\tau^i_k},\dot{z}^i_k\tau^i_k-w^i_{k-\tau^i_k})=\\&Cov(\dot{z}^i_k\tau^i_k)+2Cov(\dot{z}^i_k\tau^i_k,w^i_{k-\tau^i_k})+Cov(w^i_{k-\tau^i_k})=\\&Cov(\dot{z}^i_k\tau^i_k)+Cov(w^i_{k-\tau^i_k})=R^i_k+|\dot{z}^i_k|Cov(\tau^i_k)\le\\&R^i_k+|\dot{z}^i_k|^2_{\max}|t^i_u|^2.\end{aligned}$$

其中延迟后的随机噪声仍服从高斯分布：$w^i_{k-\tau^i_k}\sim\mathcal{N}(0,R^i_k)$，仍可以利用卡尔曼滤波器滤除. 而随机通信延迟带来的测量误差为 $\dot{z}^i_k\tau^i_k$，服从通信延迟的概率分布函数 $pt^i(\tau)$，误差大小和当前物理系统状态及状态变化率相关，是一个动态误差. 在不进行延迟补偿的前提下，第 i 个传感器测量数值的协方差应取上限值 $R^i_k+|\dot{z}^i_k|^2_{\max}|t^i_u|^2$，其中 $|\dot{z}^i_k|_{\max}$ 和 t^i_u 分别为传感器数值变化率和延迟时间的上限值. $|\dot{z}^i_k|_{\max}$ 由物理系统的特性确定，通过对物理系统的最大变化速率 $|\dot{x}_k|_{\max}$ 计算获得 $|\dot{z}^i_k|_{\max}=\boldsymbol{H}|\dot{x}_k|_{\max}$；$t^i_u$ 和网络通信协议及实时网络状态相关，可以通过分析协议或者直接通过离线网络延迟测量实验获得. 网络随机延迟的存在增大了传感器测量过程中的协方差，即减小了对应传感器测量的准确性. 系统变化速度越快、

网络通信延迟时间越长，测量精度降低幅度就越大.

(3) 攻击的检测和处理

目前有相关研究基于传统卡尔曼滤波进行开展，考虑了一定程度的攻击. 其核心思想是利用传感数据间的一致性，选取欧拉距离[22]等指标作为衡量传感器是否受攻击的判定标准. 目前相关算法的主要问题是未考虑随机延迟的影响，没有对延迟进行准确的处理，且缺少对概率分布误差的分析和处理. 在未进行随机延迟补偿时，只能将延迟带来的误差归为传感器噪声，这就导致当延迟较大时，对应的卡尔曼滤波算法中的协方差必须设置为很大，否则无法有效滤除噪声. 但是这会使融合值灵敏度降低，不利于检测和处理攻击.

4.2 基于间隔的算法

基于间隔的方法由 Marzullo 最早提出，综合考虑了噪声、延迟和攻击. 实际上该方法可以理解为在随机噪声和随机延迟精确模型基础上做的最差情况分析. 该方法只分析随机噪声和随机延迟在限定值条件下，传感数据可能的范围. 然后依据多数投票原则对传感数据受到的攻击进行检测.

(1) 噪声和延迟的处理

根据已知的噪声和延迟的范围($w_k^i \in [-e_u^i, +e_u^i], \tau_k^i \in [t_l^i, t_u^i]$)，利用限定值计算出补偿最大噪声和最大延迟后，真实传感数据可能的范围为 $z_k^i + [-e_u^i - |\dot{z}_k^i|_{\max} |t_u^i|, +e_u^i + |\dot{z}_k^i|_{\max} |t_u^i|]$，构造出第 i 个传感器的间隔，如果传感器未受到攻击，则该间隔必定包含真实物理状态[12].

(2) 攻击的检测和处理

所有未受攻击传感器的间隔均至少包含真实物理状态，因此所有未受攻击传感器的间隔必定有重叠部分，真实状态在此重合间隔内. 考虑到假设条件 A2，大部分的传感器未受到攻击，因此如果某一传感器的间隔与大多数传感器的重合间隔没有重叠，则判断该传感器受到攻击. 在进行感知融合时，直接丢弃该数据. 以图 3 为例，假设对 5 个传感器数据进行融合，$S_1 \sim S_5$(从上到下)的初始间隔分别为[14,19]、[11,17]、[13,18]、[12,15]及[10,13]，每个传感器的间隔都表示真实测量值的可能范围. 在少于半数传感器受攻击的条件下，真实被测值一定落在所有或多数传感器重叠范围内. 在获取了 5 个传感器的间隔后，计算不同范围内传感器间隔重叠的个数，表征了真实值落在该范围内的可能性大小. 在[10,11]、[11,12]、[12,13]、[13,14]、[14,15]、[15,17]、[17,18]和[18,19]这 8 段范围内，分别有 1、2、3、3、4、3、2、1 个重叠间隔. 采用多数投票法进行融合，则间隔[14,15]有 4 个传感器有共同重叠($S_1 \sim S_4$)，具有最高的投票权重，选取该间隔作为最终融合间隔，S_5 数据被丢弃.

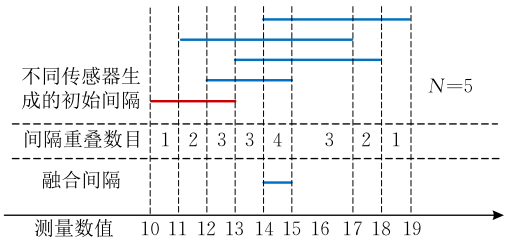

图 3 基于间隔的感知融合算法

(3) 时间窗-间隔算法

考虑到基于间隔的算法在某些时刻可能受到未知干扰的影响，同时噪声和延迟的限定值也可能存在一定误差. Ivanov 将攻击检测由单一时刻改进为某一时间窗内的综合评估，当时间窗内某传感器与多数传感器重合间隔没有重叠次的数超过限定阈值时，判定该传感器受到攻击[16]. 然而文章也指出，此方法的效果很依赖限定阈值的设定，这是该方法的主要缺点.

4.3 任意攻击的影响及适用性分析

(1) 攻击的影响分析

由假设 A1，本文中不对攻击方式做任何假设，即攻击向量 $\boldsymbol{a}_k^i$ 可以是任意值. 随机通信延迟和任意攻击使传感器上的测量噪声不再服从假设的高斯分布，使得传感器 S_i 的协方差变为 $R_k^i + |\dot{z}_k^i|_{\max} |\tau_k^i|_{\max}^2 + Cov(\boldsymbol{a}_k^i)$. 如果仍然用传统的卡尔曼滤波器处理，会带来更大的状态估测误差，且该误差是任意大的. 不进行延迟补偿和攻击检测，系统仍能正常工作的条件将变得更为苛刻.

(2) 无需延迟补偿和攻击检测的条件

当系统存在延迟时间 τ_k^i 且只有少于一半的传感器可能受到任意攻击 $\boldsymbol{a}_k^i$，N 个传感器组成的基于卡尔曼滤波的传感器系统在不进行攻击识别时满足控制要求的充分条件为至少有 $N/2+1$ 个传感器在最大延迟时间误差条件下协方差小于 R_0：$countif(R_k^i + |\dot{z}_k^i|_{\max}^2 |t_u^i|^2 \leq R_0) > N/2+1$. 此时，未受攻击的传感器中至少有一个满足存在通信延迟时的精度控制要求. 存在攻击时所有传感器的最终融合值精度高于所有未受攻击传感器的最高精度，此时即便不对攻击行为进行识别和处理，直接用传统卡尔曼滤波算法也可以达到精度控制要求. 这一条件要求选用精度过高的传感器，会在很大程度上增加系统的

硬件成本，甚至当系统延迟高于某一限定值时，当前市场上最高精度传感器也不能满足该条件. 当不满足上述条件时，必须对延迟进行补偿，并检测、处理受攻击传感数据，才能保证系统的正常运行.

(3) 适用性分析

目前，基于卡尔曼滤波和基于间隔的相关改进算法均能在一定程度上分别处理高斯噪声、延迟和攻击. 基于间隔的算法本质上简化了概率信息、只取限定值做最差分析，能够在很大程度上减小计算量，当系统噪声和延迟较小时，此方法可以获得较好的综合效果，但是此方法没有利用随机延迟的概率信息，因此当系统存在较大的噪声或延迟时，此方法的精度较差. 尤其对自适应攻击来说，过大的间隔允许自适应攻击在不被发现的前提下产生较大范围的抖动，不利于系统的稳定控制. 相比于基于间隔的算法，卡尔曼滤波算法有效利用了高斯噪声的分布特征，能够实现更高准确性的状态估测. 目前，基于卡尔曼滤波算法的主要问题是，缺少对利用随机噪声的概率信息进行攻击检测和处理的算法.

5　基于概率时间窗的抗攻击融合

根据上述分析，随机延迟和任意攻击对基于卡尔曼滤波的融合算法具有很大的影响，本节提出了基于概率时间窗的卡尔曼滤波融合算法. 首先，针对问题 P1，利用卡尔曼滤波处理基本的高斯噪声；其次，利用概率信息对各传感器测量值进行随机延迟补偿以解决问题 P2；再次，针对问题 P3，利用所获得的概率信息在一个时间窗内的累积值作为评估传感器正常工作的可能性；最后，结合各传感器时间窗权重校正卡尔曼滤波器的协方差迭代矩阵，处理可能受到攻击的传感数据.

5.1　基于概率的延迟补偿

对于实际的连续物理系统，传感器对其状态量的测量值也是连续的. 针对问题 P2，在不考虑攻击向量 $\boldsymbol{a}_k^i$ 的前提下，补偿通信延迟 τ_k^i 期间的测量值变化量：

$$\tilde{z}_{k|\tau_k^i}^i = z_k^i + \Delta z_{k|\tau_k^i}^i = \boldsymbol{H}^i(x_{k-\tau_k^i}) + w_{k-\tau_k^i}^i + \dot{z}_k^i \tau_k^i \quad (12)$$

实际获得的延迟数据中含有测量噪声，仍可以通过卡尔曼滤波进行处理，因此只需要对随机延迟造成的误差项 $\Delta z_{k|\tau_k^i}^i$ 进行补偿. 其中，$\dot{z}_k^i$ 为传感器测量值的微分，是一个动态确定变量，可以利用多种数值分析算法获得(最简单的，用当前测量值和上一周期测量值之差除以采样周期)；τ_k^i 是随机通信延迟，且服从一定的概率分布 $\tau_k^i \sim pt^i(\tau)$. 将进行延迟补偿后的测量值 $\tilde{z}_k^i$ 代入式(5)、(6)，我们可以计算出针对测量值进行随机延迟补偿后的状态估测值 $\tilde{x}_{k|k}^i \sim \mathcal{P}_k^i(x)$，

传感器预测方程：

$$\begin{aligned} \tilde{x}_{k|k-1}^i &= A\tilde{x}_{k-1|k-1}^i + Bu_{k-1}, \\ \tilde{P}_{k|k-1}^i &= A\tilde{P}_{k-1|k-1}^i A' + Q_k \end{aligned} \quad (13)$$

测量和校正方程：

$$\begin{aligned} \tilde{K}_k^i &= \tilde{P}_{k|k-1}^i \boldsymbol{H}^{i'} (\boldsymbol{H}^i \tilde{P}_{k|k-1}^i \boldsymbol{H}^{i'} + R_k^i)^{-1}, \\ \tilde{x}_{k|k}^i &= \tilde{x}_{k|k-1}^i + \tilde{K}_k^i (\tilde{z}_{k|\tau_k^i}^i - \boldsymbol{H}^i \tilde{x}_{k|k-1}^i), \\ \tilde{P}_{k|k}^i &= (I - \tilde{K}_k^i \boldsymbol{H}^i) \tilde{P}_{k|k-1}^i \end{aligned} \quad (14)$$

其概率密度函数中包含了以下信息：

(1) 延迟时间 τ^i 范围比较大的传感器，相应的补偿值概率密度 $\mathcal{P}_k^i(x)$ 更分散，不确定性更高；

(2) 该补偿值是在取攻击向量 $\boldsymbol{a}_k^i$ 为 $\boldsymbol{0}$ 的前提下计算出来的，当第 i 个传感器在 k 时刻受到攻击时，对应的实际概率分布函数会发生偏移，概率密度函数的差异可以表征攻击向量的大小，这一特性可以用来进行对传感器攻击的检测.

对应的 N 个传感器共同组成的考虑随机延迟测量模型为

$$\tilde{z}_{k|\tau_k} = \boldsymbol{H}(x_k) + W_k \quad (15)$$

式中 $\tilde{z}_{k|\tau_k} = (\tilde{z}_{k|\tau_k^1}^{1'}, \cdots, \tilde{z}_{k|\tau_k^N}^{N'})'$，不同测量序列对融合值进行预测和更新，利用式(13)、(14)获得对应的融合更新值 $\tilde{x}_{k|k}$ 对应概率为 $\mathcal{P}_k(x)$.

不考虑攻击时，最终输出值为对应概率最大的估测值 $\tilde{x}_{k|k}$，$\mathcal{P}_k(\tilde{x}_{k|k}) = \max(\mathcal{P}_k(x))$. 当系统中第 i 个传感器未受攻击时($\boldsymbol{a}_k^i = 0$)，对应的状态量估测值期望概率 $\mathcal{P}_k^i(\tilde{x}_{k|k}^i) > 0$. 当系统中第 i 个传感器受攻击时($\boldsymbol{a}_k^i \neq 0$)，对应的状态量估测值期望概率发生偏移，如果此时 $\mathcal{P}_k^i(\tilde{x}_{k|k}^i) > 0$，则说明尽管受到了攻击，但是状态估测值仍在可接受范围内波动；否则若 $\mathcal{P}_k^i(\tilde{x}_{k|k}^i) = 0$，说明该传感器受攻击的影响较大，已经完全超出了可能的范围.

5.2　概率信息误差的影响

然而，仅用单个周期的 $\mathcal{P}_k^i(\tilde{x}_{k|k}^i) = 0$ 作为传感器受到攻击的判据可能存在问题. 一方面，上述所有传感器工作的概率分析都依据于已知系统内随机通信延迟的概率分布信息，即 $pt^i(\tau)$. 但是无论 j 离线测量方式还是在线评估方式，在获取 $pt^i(\tau)$ 时均难以避免存在误差. 延迟时间 τ^i 的最大值和最小值确定了对应状态估测值 $\tilde{x}_{k|k}^i$ 的范围，如果延迟时间范围存在误差，那么势必会导致 $\tilde{x}_{k|k}^i$ 存在误差. 延迟时间

范围比真实延迟大，则会存在传感器已经受到攻击但是$\mathcal{P}_k^i(\tilde{x}_{k|k}^i)>0$的情况，会降低攻击识别的准确性；延迟时间范围比真实延迟小，则会存在传感器未受到攻击但是$\mathcal{P}_k^i(\tilde{x}_{k|k}^i)=0$的情况，此时该传感器会被误认为受到攻击，对应的传感器提升融合值精度的能力被削弱，最终融合值的精度受到影响. 另一方面，在实际系统中，卡尔曼滤波算法所有的参数都只能靠对系统的认知和专业经验在一定范围内确定，比如各传感器测量噪声的协方差R_k^i、处理过程协方差Q_k. 各传感器微分值的计算也同样存在一定误差. 这些因素积累到一起，都会对卡尔曼滤波估测和随机延迟补偿时造成小幅度的误差，且该误差难以估计和处理. 这类误差同样可能造成对应的概率信息$\mathcal{P}_k^i(\tilde{x}_{k|k}^i)$不准确，影响传感器攻击识别.

5.3 基于概率时间窗的攻击检测

单一时刻的攻击检测可能受到概率信息误差的较大影响，对某一时间窗内的概率进行综合评估，可以有效改善类似干扰造成的不良影响. 设定一个长度为L个周期的时间窗，对第i个传感器在最近L个采样周期内的期望概率进行累积，获得概率时间窗的值：

$$PW_{k|L}^i=\sum_{k-L}^{k}\mathcal{P}_{k-1}^i(\tilde{x}_{k-1|k-1}) \tag{16}$$

用以衡量在过去L个周期内，将各传感器概率时间窗数值相对所有传感器概率时间窗最大值的比值作为传感器的时间窗权重：

$$\alpha_{k|L}^i=\frac{PW_{k|L}^i}{\max(PW_{k|L}^1,\cdots,PW_{k|L}^N)} \tag{17}$$

区别于只用一个周期的概率作为评估传感器是否正常工作的标准，对一个时间窗的概率进行累积可以减小上述分析中各类偶然因素的影响. 传感器的时间窗权重可以与某一阈值进行比较，判断传感器是否受到攻击，并直接将认为受到攻击的数据剔除，然后再进行信息融合；也可以作为权重确定某一传感器数据对最终融合数据的影响. 前者的处理算法更简单，带来的资源开销低，更适用于实时性要求高、资源受限的场景，但是需要设置合理的判断阈值，且剔除时可能产生较大的振荡，这需要对系统有准确的认知且更依赖于开发者的经验；后者的处理算法过渡更平滑，对各种系统的通用性更好，容易扩展到各类应用场景.

5.4 基于时间窗权重的融合

上文已经通过概率时间窗的方式对传感器过去一段时间内正常工作的概率进行了评估，计算出了各传感器的时间窗权重. 根据感知融合的基本思想，在对多个传感器数据进行融合时，应保证高时间窗权重传感器的数值对最终融合值的影响更大. 将式(13)、(14)内加入各传感器时间窗权重，调整为

$$\begin{aligned}\tilde{x}_{k|k-1}&=A\tilde{x}_{k-1|k-1}+Bu_{k-1},\\ \tilde{P}_{k|k-1}&=A\tilde{P}_{k-1|k-1}A'+Q_k\end{aligned} \tag{18}$$

测量和校正方程：

$$\begin{aligned}\tilde{K}_k&=\tilde{P}_{k|k-1}\boldsymbol{H}'(\boldsymbol{H}\tilde{P}_{k|k-1}^i\boldsymbol{H}'+R_k)^{-1},\\ \tilde{x}_{k|k}&=\tilde{x}_{k|k-1}+\alpha_{k|L}\tilde{K}_k(\tilde{z}_{k|\tau_k^i}^i-\boldsymbol{H}\tilde{x}_{k|k-1}),\\ \tilde{P}_{k|k}&=(I-\tilde{K}_k\boldsymbol{H})\tilde{P}_{k|k-1}\end{aligned} \tag{19}$$

式中$\alpha_{k|L}=[\alpha_{k|L}^1,\cdots,\alpha_{k|L}^N]$为各传感器构成的时间窗权重.

现已将传感器时间窗权重融入到卡尔曼滤波算法中，在过去L个周期的时间窗内，如果对应某传感器时间窗权重较低，则在进行融合值计算时，其对应测量值对融合值的影响也较低. 而且此算法并不是简单地将可能受到攻击的数据在满足某一条件的瞬间剔除，而是在整个测量周期内平滑地进行过渡，减小了融合过程中数值的扰动. 当某一传感器逐渐恢复正常工作后，相应的概率时间窗数值也会逐渐恢复到正常水平，对应的数据会重新参与融合值的计算，有效对应 A1 中对传感器攻击间歇发生的假设.

如果自适应攻击者完全掌握了本文所设计的融合算法，且认为攻击者能够获得所有传感器数据，但是只能对其中少于一半的传感器数据进行攻击. 其攻击目标为在不被发现的前提下，使最终融合值误差最大. 则在攻击过程中，攻击者应始终保证其所攻击传感器估测状态值对应概率大于 0. 根据第 4 节中的分析，最终融合值的协方差小于所有传感器最小协方差，则自适应攻击的最优目标对象为精度较高、延迟较小的传感器. 在自适应攻击下，最终融合值范围由未受攻击传感器共同确定，融合值被自适应攻击者控制在该范围内变化，但是不会超出该范围，否则攻击会被识别.

6 对比仿真验证及结果分析

本节在仿真软件 MATLAB/Simulink 中建立自动驾驶车队系统模型，将对传统卡尔曼滤波、随机延迟补偿卡尔曼滤波、欧拉距离卡尔曼滤波、时间窗-间隔的融合和本文所提出的概率时间窗卡尔曼滤波感知融合，共 5 种感知融合算法的性能进行验证分析. 并通过仿真验证，分析噪声和延迟概率分布信息对概率时间窗融合精度的影响，讨论自适应攻

击下概率时间窗融合算法的性能.

6.1　仿真设置

自动驾驶车队的框架如图 4 所示，整个车队由头车带领行驶，后面若干台尾车跟随头车行驶并保持与前车车距恒定. 每台车辆配置 4 个传感器，其中 2 个编码器测量车辆主动轮转速，1 台雷达测量本车车头与前车车尾间距离，1 台摄像头用以实现路径跟随，且通过特定的视觉处理算法估测与前车距离. 特别地，为了更好地分析感知融合效果，本文不考虑车队的转向等复杂情况，即头车以一定速度沿直线前进. 此时，2 个编码器测量的主动轮转速和车辆行驶速度成正比. 为更好地分析不同算法，本文主要对利用不同传感器测量值估测头车速度进行重点分析.

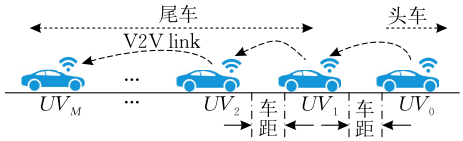

图 4　车队系统示意图

(1) 传感器配置及参数

为了简化分析，在仿真中建立的验证模型只包含 2 台自动驾驶车 UV_0 和 UV_1，构成最小车队. 该系统框图如图 5 所示，尾车融合的数据来自 6 个传感器，本地传感数据 2 个编码器测量的转速数据 (z_k^1, z_k^2)，雷达和摄像头测量的车距数据 (z_k^3, z_k^4)；外部传感器数据来自头车的 2 个编码器 (z_k^5, z_k^6)，精度与尾车编码器相同.

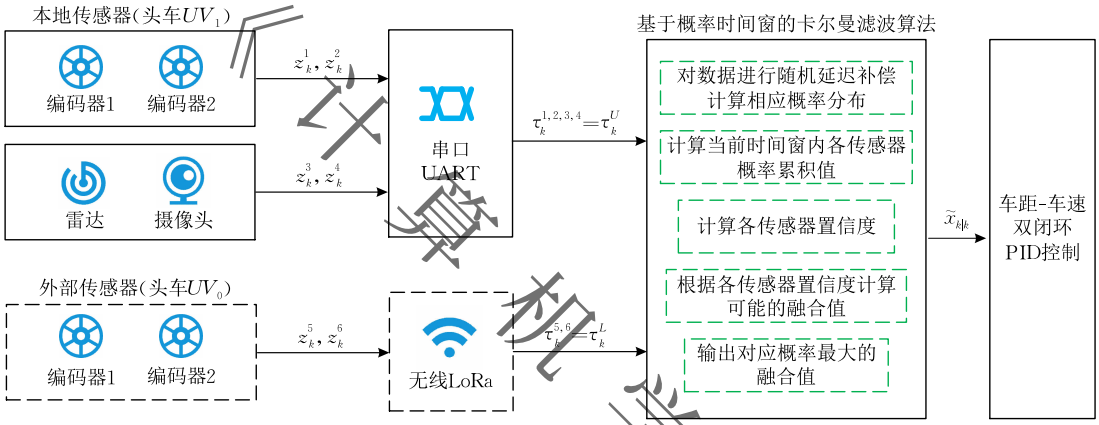

图 5　多感知融合-车队控制系统示意图

依据上述条件，本文所仿真的车队控制系统状态主要有三个量，分别为头车速度 V、尾车速度 V'、车距 d，分别由 $z_k^{5,6}$、$z_k^{1,2}$、$z_k^{3,4}$ 处理获得. 本文重点对比头车速度 V 的估测值，该值可以直接通过前车编码器测量获得：$z_k^{5,6}=V+w_k^{5,6}$；也可以通过尾车速度 V' 和车距 d 间接测量：$\mathrm{d}d/\mathrm{d}t=V-V'$，$z_k^{1,2}=V'+w_k^{1,2}$，$z_k^{3,4}=d+w_k^{3,4}$. 因此对头车速度 V，共有 4 组传感器进行测量：$\{z_k^5\}$，$\{z_k^6\}$，$\{z_k^1+z_k^3\}$，$\{z_k^2+z_k^4\}$，对应的进行相应卡尔曼滤波后的状态估测值用上角标区分，记为 V^1，V^2，V^3，V^4.

(2) 算法对比设置

为对比本文所提出算法同传统算法的优势，设置以下 5 种算法，各算法对应的估测值用下角标$_{0,1,2,3,4}$区分，其中第 4 种算法随机延迟补偿-概率时间窗卡尔曼滤波(KF-PW)是本文所提出的算法，第 0～3 种算法是作为对照组的基础算法.

0 传统卡尔曼滤波算法(Kalman Filter，KF). 不考虑通信延迟和攻击，直接将原始数据输入到卡尔曼滤波器；

1 欧拉距离卡尔曼滤波算法(KF-E). 在传统卡尔曼滤波器基础上，不考虑通信延迟，利用欧拉距离对攻击进行检测和处理[22]；

2 基于时间窗-间隔的融合算法(IT-W). 只考虑噪声和延迟的限定值，确定各传感器补偿最大噪声和延迟后的范围，根据投票原则检测攻击，时间窗内被识别为攻击的次数超过设定阈值时，判定传感器受到攻击[16]，在进行融合时弃用判定为受攻击的传感数据；

3 随机延迟补偿卡尔曼滤波算法(KF-P). 在传统卡尔曼滤波器 KF 基础上，对原始数据进行随机延迟补偿，将对应概率最大的数据作为滤波器输入数据；

4 随机延迟补偿-概率时间窗卡尔曼滤波(KF-PW). 在延迟补偿卡尔曼滤波器 KF-P 基础上，对历史数据在时间窗内的概率进行累积并作为时间窗权重，参与卡尔曼滤波中协方差矩阵的更新.

(3) 时间窗设置

时间窗的大小会影响感知融合的效果，时间窗

太小则抗干扰能力较差，检测结果存在较大偶然性；时间窗过大，导致系统灵敏度降低，不能及时检测出攻击. 为测试不同时间窗大小对时间窗权重的影响，针对图 6 中所示的真实延迟时间下 4 个传感器进行 10 万次蒙特卡洛模拟，统计时间窗权重的概率分布情况，结果箱线图如图 7 所示. 横坐标为时间窗大小，纵坐标表示 10 万次蒙特卡洛模拟的统计结果，有数据点的是模拟中存在的样本，其中第一四分位数和第三四分位数的范围用实线框标出. 时间窗权重的分布范围越小，对应的实线框位置越高，说明算法的偶然性越低，准确性越高. 时间窗大小从 1 增加到 100 个采样周期，随着时间窗的增大，对应的时间窗权重分布越稳定. 当时间窗在 1 到 50 个采样周期阶段增大时，时间窗权重的概率分布范围迅速减小，在时间窗取 30 个采样周期时，时间窗权重的中位值已超过 0.9；时间窗超过 80 个采样周期后，时间窗增大给消弱偶然性带来的改善能力越来越小，同时增大了算法的响应时间，降低了算法灵敏度. 结合上述模拟结果分析，30～80 个采样周期是比较合理的时间窗大小，在该范围内所提出的算法具有较高的准确性和灵敏度，后续仿真中取时间窗大小为 50 个采样周期.

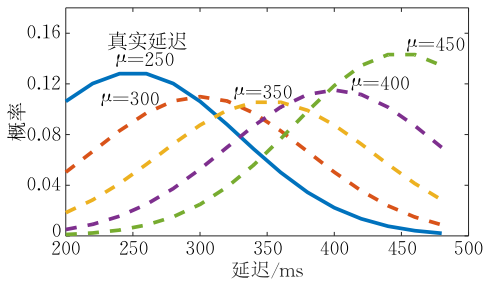

图 6 延迟概率分布

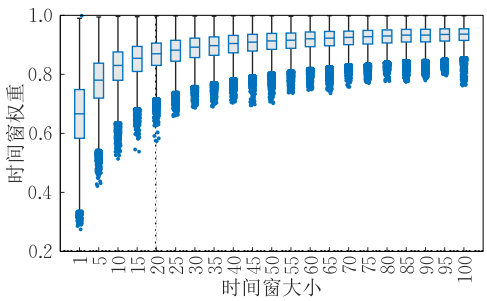

图 7 时间窗大小影响蒙特卡洛分析结果

(4) 工况和评估标准

本文所有仿真均在相同工况下进行，头车车速在 0～15 s 期间为恒速 1.0 m/s，在 15～20 s 期间以 0.4 m/s^2 的加速度运行，在 20～30 s 期间保持匀速 3.0 m/s 运行. 整个运行时间段内，目标车距保持 0.3 m 不变. 在存在通信延迟和任意攻击的条件下，对比上述 5 种算法在状态估测累积误差、车距控制精度、攻击检测三个方面的效果，其中状态估测累积误差用真实状态和估测状态的差值取平方后积分来表示，记为 $\int e^2$.

(5) 随机噪声设置

为验证各类算法受随机噪声概率信息误差的影响，分别设置控制器获取真实和存在误差的随机噪声信息.

真实随机噪声. 所有传感器的测量噪声服从高斯分布，其中 4 个编码器对车速测量的噪声协方差相同，为 $w_k^{1,2,5,6}\sim\mathcal{N}(0,0.05)$，雷达对距离的测量噪声为 $w_k^3\sim\mathcal{N}(0,0.005)$，摄像头对距离的测量噪声为 $w_k^4\sim\mathcal{N}(0,0.1)$.

存在误差的随机噪声. 在保证随机噪声仍满足零均值的前提下，设置控制器获取的高斯分布中表征传感器准确性的协方差存在误差，为实际协方差的 γ 倍，即 4 个编码器服从分布 $w_k^{1,2,5,6}\sim\mathcal{N}(0,\gamma*0.05)$，雷达对距离的测量噪声为 $w_k^3\sim\mathcal{N}(0,\gamma*0.005)$，摄像头对距离的测量噪声为 $w_k^4\sim\mathcal{N}(0,\gamma*0.1)$.

(6) 随机延迟设置

为验证各类算法受随机延迟概率信息误差的影响，分别设置控制器获取真实和存在误差的随机延迟概率信息.

真实通信延迟. 本地传感器通过串口 UART 通信协议与控制器进行信息交互，通信延迟均服从正态分布 $\tau_k^U\sim\mathcal{N}(0.001,0.0001)$. 外部传感器数据在头车和尾车之间通过 LoRa 无线通信模块进行交互，通信延迟为 τ_k^L，其概率密度函数如图 6 中蓝色实线表示，用函数 $\mathcal{N}_{pdf}([200,500],250,80)$ 拟合而成，第 1 个参数表示随机延迟范围在 200～500 ms 之间，第 2 个参数表示最大概率对应延迟时间为 250 ms，第 3 个参数表示正态分布协方差为 80.

存在误差的通信延迟. 实际系统工作过程中，网络通信延迟随时可能发生变化，因此控制器获得的网络延迟概率分布信息可能与实际系统中的延迟存在一定的误差. 为了验证本文所提出的算法在通信延迟概率信息存在误差时的表现，设置多组不同误差的延迟概率分布信息.

通信延迟的概率信息误差主要表现在 2 个方面：范围和对应的概率值. 因此设置不同延迟范围作为一组对照实验，控制器获得的概率分布函数为 $\mathcal{N}_{pdf}(\lambda*[200,500],250,80)$，其中 λ 为延迟时间缩放系数. 设置不同的概率分布信息作为另外一组对照实验，控制器获得的概率分布函数为 $\mathcal{N}_{pdf}([200,$

500]，$\mu * 250$，80)，其中 μ 为最大概率对应的延迟时间的缩放系数，如图 6 中不同颜色虚线所示.

(7) 攻击设置

编码器测速精度高于雷达、摄像头，为更好地验证各算法的抗攻击能力，本文所有攻击均发生在精度较高的第 1 组传感器上. 设定通过直接物理干扰或者在通信过程中拦截的方式进行攻击，最终结果为控制器接收到的第 1 组传感器(即前车编码器 1)测量值与真实测量值存在偏差. 为验证本文算法对不同形式攻击的抵抗性，设置了普通攻击和自适应攻击两种攻击方式分别进行验证，在仿真结果图上方用红色箭头和线段表示攻击触发的对应时间段. 车辆控制系统实现闭环控制有“反馈”和“控制”两个环节，虽然攻击的效果在控制中会逐渐被消除，但是在相同的控制系统中，反馈值越接近真实值，攻击带来的影响也会越小. 本文重点在多传感器融合过程中消除攻击者对反馈值带来的影响，通过观测攻击过程中的车速和车距等变量可以对传感器融合算法的抗攻击性进行有效评估.

普通攻击. 如果攻击者不知道本文所提出的感知融合算法，在车辆运行期间篡改第 1 组传感器测量值，具体设置如下：攻击 1，5 s～8 s 期间，第 1 组传感器输出值为实际测量值加 0.5 m/s；攻击 2，16 s～19 s 期间，第 1 组传感器输出值为实际测量值加 0.5 m/s. 上述攻击设定模拟场景为，在自动驾驶车队正常行驶过程中，攻击者篡改头车的速度，给尾车造成头车突然加速的假象，而车队控制要求车距恒定，车距控制程序由于收到“头车加速”的虚假信号，为了预先消除头车加速造成的车距增大，将头车车速通过前馈环节给到车速控制器，进而造成实际的车距减小.

自适应攻击. 如果攻击者完全知道系统内的攻击检测和信息融合算法，在车队系统运行期间能够监视所有传感器数据，并控制第 1 组传感器输出任意值，则遵循以下 2 个原则：原则 1，在攻击过程中，不被攻击检测算法识别；原则 2，在原则 1 基础上，使最终融合值最大限度偏离真实值. 攻击者预先知道控制器如何对多传感器进行攻击检测和融合，对 1 个(小于 $N/2$)被攻击传感器进行控制，同时能够预先判断篡改后的第 1 组传感器是否会被当前抗攻击融合算法检测出来. 如果会被检测出来，则该传感器测量值会被丢弃，不能对融合值造成影响，违背原则 2，因此攻击者会放弃当前的攻击方案，调整攻击向量直到不会被检测出来，且能对融合值产生尽可能大的影响.

(8) MATLAB 实现说明

本文所有仿真均在 MATLAB/Simulink 平台上实现. 首先在 Simulink 平台上搭建基础的车队控制模型，单个车辆模型简化为直流电机数学模型，传递函数为 $1.786/(0.0088\,s^2+0.58\,s+1)$；车队控制采用车距-车速双闭环 PID 控制，内环车速 PID 参数为 4/25/0.05，外环车距 PID 参数为 1000/800/0. 所有感知融合算法由 S-Function 实现，输入 4 组原始传感数据，在 S 函数内部进行各类感知融合算法，输出最终的前车车速估测值，输入到车队控制器实现恒车距控制. 仿真求解器为 ode1(Euler)，仿真步长设置为定步长 0.0001 s.

6.2　普通攻击仿真结果

在控制器获得真实通信延迟概率信息的条件下，对不同算法在普通攻击下的工况进行仿真验证.

(1) 估测精度及控制效果对比

5 种算法车队控制效果对比如图 8，第 1～5 行分别为 KF、KF-E、IT-W、KF-P、KF-PW 融合算法仿真结果，用下角标$_{0,1,2,3,4}$标识；第 1～3 列分别为头车车速估测值、头车车速估测误差、车距实际控制结果. 对应的误差累积值 $\int e^2$ 分别为 0.16、0.15、0.26、0.12 和 0.05，本文提出的随机延迟补偿算法相对于传统卡尔曼滤波算法，累积误差减小了 25%；概率时间窗融合算法相对于传统卡尔曼滤波，累积误差减小了 67%.

在攻击 1 触发期间，KF 算法对头车速度估测值由真实值增加到 1.3 m/s 左右，估测误差约为−0.3 m/s，造成的车距波动约为 0.02 m；KF-E 效果与 KF 基本一致，只不过由于欧拉距离攻击检测机制的存在，在一定程度上限值了攻击的效果，但是该算法的灵敏度较差；IT-W 虽然在攻击期间速度扰动与卡尔曼滤波基本类似，但是在非攻击区间由于算法自身缺陷，存在比较大的高频抖动，幅值约为 0.15 m/s；KF-P 算法的表现比 KF 稍好，主要体现在估测误差积累值较小，但是估测误差幅值基本与 KF 相同，为−0.3 m/s；KF-PW 在 5 个算法中表现最优，估测误差控制在±0.02 m/s 的范围内.

在攻击 2 触发期间，KF 算法对头车速度估测误差约为±0.15 m/s，造成的车距波动约为 0.04 m；KF-E 效果与 KF 基本一致，攻击造成的误差因欧拉距离检测的存在被处理成高频抖动，造成的波动比

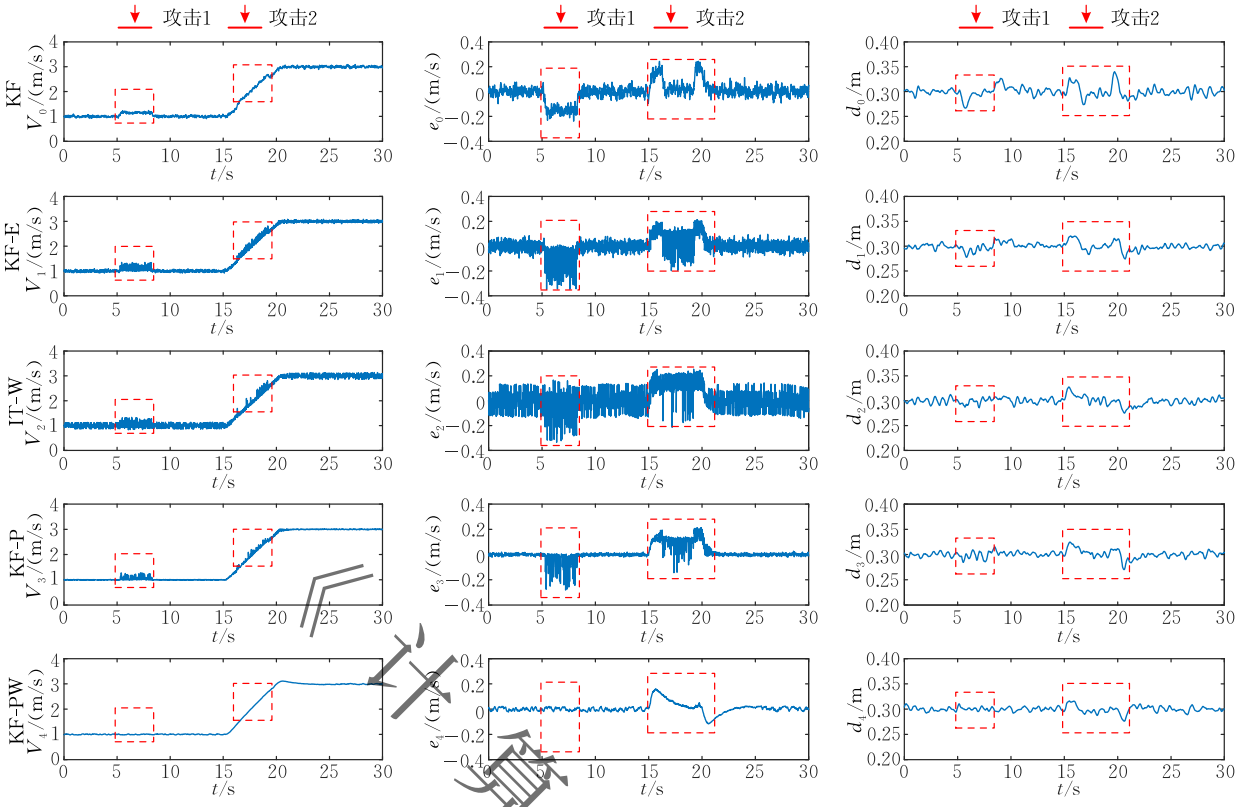

图 8 普通攻击下不同算法车队控制效果对比

KF 稍小;IT-W 在攻击期间速度扰动约为 0.15±0.06 m/s,存在比较大的高频抖动;KF-P 算法同样受到攻击的一定影响,约有 0.10±0.05 m/s,车距几乎未受影响;KF-PW 在 5 个算法中表现最优,误差的幅值和积累值都最小,车距稍有波动,约为 0.02 m,而重新恢复稳定时间最短.

(2) 攻击检测结果

算法 KF-PW 在系统运行期间利用概率时间窗对各组传感器可能受到攻击的行为进行实时检测,对第 1 组传感器数值的概率时间窗检测结果如图 9、图 10 所示. 在第 1 组传感器未受到攻击期间,由图 9 观察,最终的融合值始终在其估测的最小值和最大值的范围内小幅度波动,对应的图 10 中概率时间窗数值与其他 3 组传感器基本接近;在第 1 组传感器受到攻击期间,最终的融合值超出其估测最小值和最大值的范围,对应的图 10 中概率时间窗数值迅速降低到 0 左右的水平,而其他 3 组传感器的概率时间窗数值迅速上升,表明在此期间,控制器检测到第 1 组传感器数据的异常,并逐渐降低了更新估测值时该组传感器对应的权重系数,直至相当于完全剔除了受攻击数据. 表明 KF-PW 算法对普通攻击具有较强的抵抗能力.

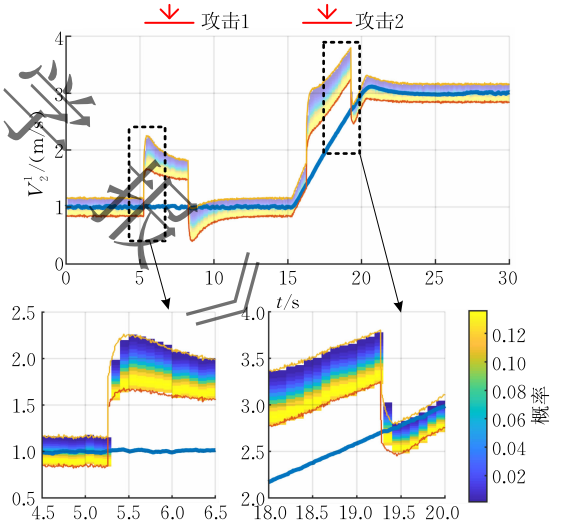

图 9 普通攻击下第 1 组感知融合过程

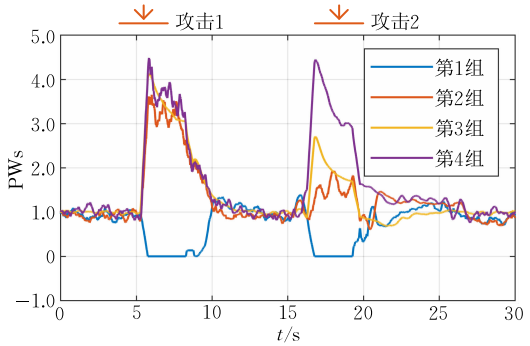

图 10 普通攻击下各组传感器概率时间窗

6.3　概率信息误差的影响

本文所提出的算法需要利用随机噪声和随机延迟的概率分布信息，因此验证算法在不同概率信息误差下的容错性十分必要. 根据 6.1 节(6)、(7)中对概率信息误差所做的设置，验证 5 种融合算法在 6.2 节所述普通攻击下的效果. 对比不同误差对应的融合误差累积值.

(1) 随机噪声误差

随机噪声的概率信息主要表现为高斯分布中协方差的大小，对应卡尔曼滤波算法中的 R_k^i 和基于间隔算法中的限定值. 基于卡尔曼滤波的融合算法对协方差概率信息误差的仿真结果如图 11 所示. 相比于图 12 中基于间隔算法的仿真结果，基于卡尔曼滤波的 4 个融合算法对噪声信息误差具有更高的容错性，KF 对协方差的容错性是 4 种 KF 算法中最差的，在缩放系数 $\gamma=1.0$ 附近有最小的累积误差. KF-E 由于欧拉检测算法的存在，其能够在一定程度上降低攻击造成的影响，在不同缩放系数 γ 下，累积误差相似，这是由于不同组别传感器误差在欧拉距离的机制的调节性中获得了一定的相互补偿. KF-P 由于随机噪声补偿，获得了更高的精度，但是由于没有攻击检测处理机制，因此呈现出与 KF 相似的趋势，在缩放系数 $\gamma=1.0$ 附近有最小的累积误差. KF-PW 在 KF-P 基础上加入时间窗攻击检测处理，在不同缩放系数 γ 下均具有最小的累积误差. 基于间隔的融合算法具有明显区别于 KF 算法的趋势. 在图 12 中，当缩放系数 $\gamma<1.0$ 时，IT-W 算法累积误差随缩放系数的减小而迅速增大，已经超出了系统正常工作允许的最大误差. 这是因为 IT-W 算法要求间隔必须覆盖所有可能范围，当获得的数据范围比实际范围小时，不能保证间隔内包含真实数据，违背了该算法的基本条件. 而基于 KF 的算法，自身对协方差具有一定的容错性.

(2) 通信延迟误差的影响

考虑到实际系统中的通信延迟与控制器获得的信息存在一定误差，因此对通信延迟误差分别从范围和概率分布误差对其影响进行仿真验证.

① 时间范围误差

控制器获得的随机延迟概率分布函数为 $\mathcal{N}_{pdf}(\lambda*[200,500],250,80)$，其中 λ 为延迟时间缩放系数，当 $\lambda=1.0$ 时即为真实通信延迟. 不同缩放系数对应的估测误差积分 $\int e^2$ 曲线如图 13、图 14 所示. KF 和 KF-E 不根据随机延迟信息进行补偿，直接将该分量视作传感器自身噪声，因此其累积误差与 λ 没有明显的趋势关系，而 KF-P 和 KF-PW 的算法很依赖于正确的延迟概率信息，累积误差和缩放系数呈现明显的抛物线关系，在 $\lambda=1.0$ 时具有最小的累积误差，说明在控制器完全获得准确的延迟分布信息时，能够达到最优的补偿效果，符合常理分析. 同时，与 KF 算法的累积误差值 0.15 相比，当延迟范围缩放系数在[0.5,1.5]的范围内时，KF-P 均能实现更高精度的感知融合；当延迟范围缩放系数在[0.4,1.6]的范围内时，KF-PW 均能实现更高精度的感知融合. 这也说明所提出算法对延迟的时间范围误差具有较高的容错性. 与图 12 中噪声概率误差的趋势相似，如图 14 所示，IT-W 算法对比实际值

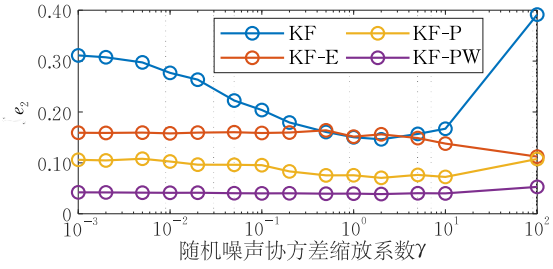

图 11　基于卡尔曼滤波算法受噪声概率的影响

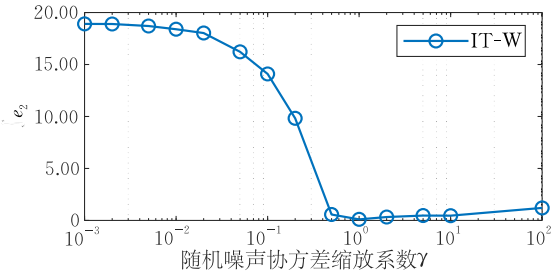

图 12　基于间隔算法受噪声概率的影响

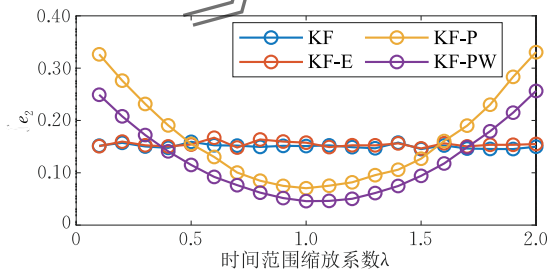

图 13　基于卡尔曼滤波算法受延迟范围的影响

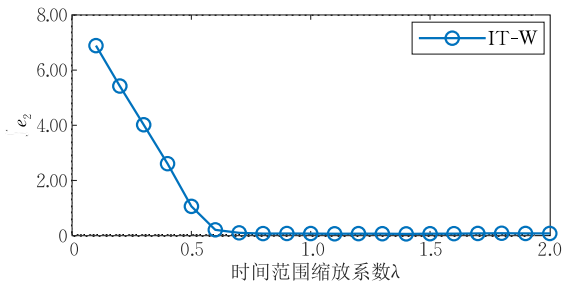

图 14　基于间隔算法受延迟范围的影响

小的时间范围 $\lambda<1.0$ 容错性很差，在 $\lambda<0.6$ 时无法满足控制要求.

② 概率分布误差

控制器获得的随机延迟概率分布函数为 $\mathcal{N}_{pdf}([200,500],\mu*250,80)$，其中 μ 为最大概率对应的延迟时间缩放系数，当 $\mu=1.0$ 时即为真实通信延迟. 不同最大概率延迟时间对应的估测误差积分 $\int e^2$ 曲线如图 15 所示. 累积误差和 μ 呈现明显的相关关系，在 $\mu\in[0.67,1.33]$ 范围时，具有较小的累积误差，说明在控制器获得相对准确的延迟分布信息时，能够达到最优的补偿效果，符合常理分析. 同时，与 KF 算法的累积误差值 0.15 相比，只要通信延迟范围准确，各种概率分布误差条件下所提出的 KF-P、KF-PW 算法均有一定的优化效果，这也在一定程度上说明所提出算法对延迟的概率分布误差具有较高的容错性.

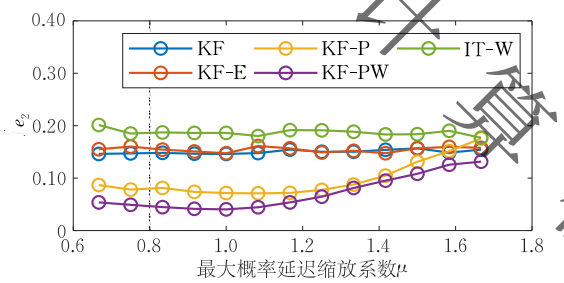

图 15 延迟时间概率分布误差的影响

(3) 总结分析

通过分析上述仿真结果，当累积误差小于 0.20 时，系统可以实现较好的控制效果. 以累积误差 0.20 为界，对上述误差仿真结果进行总结，得到不同算法对不同误差容错性的对比表格，如表 1 所示. 容错区间越大，表明系统容错性越高，在实际系统中的适用性越好. 基于卡尔曼滤波的 4 种算法对随机噪声的容错性都很高，由于 KF 和 KF-E 未利用延迟概率信息，因此对延迟的误差容错率也很高，但是在延迟的范围和概率分布误差较小时，KF-P 和 KF-PW 均能获得更高准确性的融合估测值. 基于间隔的算法对噪声和延迟的误差都非常敏感，在获得的限定值范围小于实际范围时，会产生很大的累积误差，远高于系统正常工作的要求.

表 1 $\int e^2$ 小于 0.20 对应的误差容错区间

算法	噪声协方差缩放系数 γ	延迟时间范围缩放系数 λ	延迟概率分布缩放系数 μ
KF	$[0.1,10.5]$	$[0.1,2.0]$	$[0.7,1.7]$
KF-E	$[10^{-3},10^{2}]$	$[0.1,2.0]$	$[0.7,1.7]$
IT-W	$[0.6,1.1]$	$[0.6,2.0]$	$[0.7,1.7]$
KF-P	$[10^{-3},10^{2}]$	$[0.4,1.7]$	$[0.7,1.7]$
KF-PW	$[10^{-3},10^{2}]$	$[0.2,1.9]$	$[0.7,1.7]$

6.4 自适应攻击仿真结果

根据自适应攻击保证不被发现和误差最大化两条原则，攻击者对第 1 组传感器发起自适应攻击持续整个系统工作过程. 因为在 6.2 小节已经验证了 KF 和 KF-P 算法无法识别、抵抗普通攻击，KF-E 和 IT-W 算法对普通攻击的抵抗性较差，受篇幅限制，此处省略这 4 种算法在自适应攻击下的控制效果及相应分析.

(1) 估测精度及控制效果

图 16 为自适应攻击下车队控制效果，此时估测误差在保证攻击不被发现的原则下保持最大值抖动，累积误差远高于普通攻击行为. 但是其攻击效果同时也被限制在可接受范围内，对头车速度的估计值即便存在较大误差，仍能满足车队控制的基本要求，车距控制误差在 ±0.03 m 范围内.

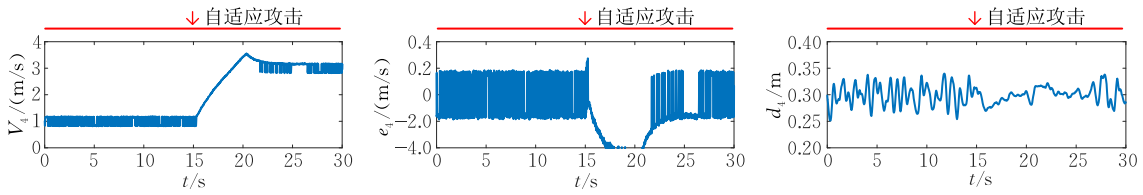

图 16 自适应攻击下车队控制效果

(2) 攻击检测结果

图 17 和图 18 为自适应攻击的检测过程. 在系统运行过程中，攻击者能够获得所有传感器数据并完全了解本文所提出的算法，可以对第 1 组传感器进行任意攻击. 在融合算法限制下，自适应攻击能够计算出攻击行为不被检测出来的最大范围，在该范围内选择使估测值误差最大的值作为第 1 组传感器输出. 因此在个别时间段，输出融合值在最大、最小值之间频繁跳动，如图 17 中 4.5 s～6.5 s 时间段局部放大图所示；在个别时间段内，输出融合值始终保持为估测范围中的一个，如图 17 中 18 s～20 s 时间段局部放大图所示. 在整个系统工作过程中，由于自适应攻击下，第 1 组传感器输出值的范围始终保持在不被发现的范围内，因而在图 18 中的概率时间窗

结果始终维持在正常状态，期间没有某一传感器概率时间窗的值迅速下降的情况，说明自适应攻击保证了不被发现原则.

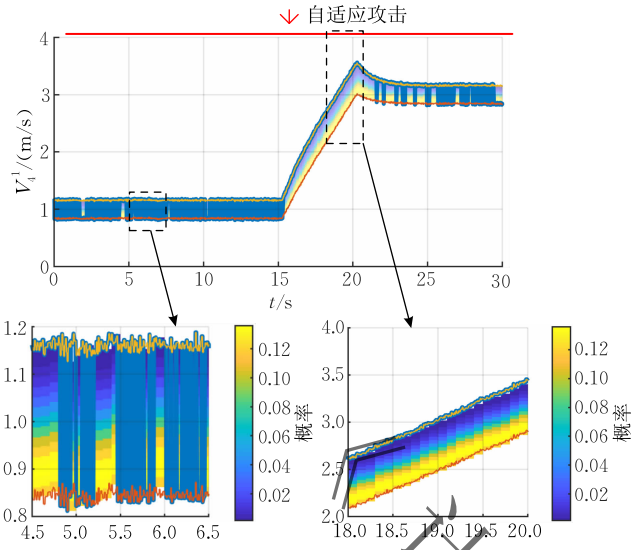

图 17　自适应攻击下第 1 组感知融合过程

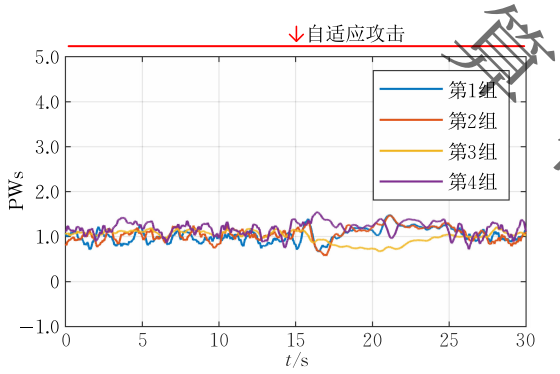

图 18　自适应攻击下各组传感器概率时间窗

7　总结及展望

本文对多感知融合领域中随机通信延迟的影响进行了分析，利用延迟的概率分布信息对传感器测量值进行了补偿. 为了监测系统运行过程中传感器的行为，提出了概率时间窗的传感器受攻击评估算法. 仿真结果表明，所提出的算法对随机通信延迟具有较强的补偿效果，累积误差比传统卡尔曼滤波降低 67%，对间歇触发的普通攻击行为具有很好的抵抗性. 同时，本文还假设控制器获得的概率信息存在误差，在不同误差情况下进行了仿真，结果表明所提出的算法对噪声和延迟的概率误差具有较强的容错性，在通信延迟范围和真实延迟有 60%相对误差时仍能取得比传统卡尔曼滤波器更好的效果. 最后，即使攻击者掌握了本文所提出算法，在不被发现的前提下进行自适应攻击，本文算法能够将自适应攻击的影响控制在可接受误差范围内.

参　考　文　献

[1] Wolf M, Serpanos D. Safety and security in cyber-physical systems and Internet-of-Things systems. Proceedings of the IEEE, 2018, 106(1): 9-20

[2] Chen K, Zhang S, Li Z, et al. Internet-of-Things security and vulnerabilities: Taxonomy, challenges, and practice. Journal of Hardware and Systems Security, 2018, 2(1): 97-110

[3] Chattopadhyay A, Lam K Y, Tavva Y. Autonomous vehicle: Security by design. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(11): 7015-7029

[4] Boloor A, Garimella K, He X, et al. Attacking vision-based perception in end-to-end autonomous driving models. Journal of Systems Architecture, 2020, 110: 101766

[5] Jha S, Cui S, Banerjee S, et al. ML-driven malware that targets AV safety//Proceedings of the 2020 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). Valencia, Spain, 2020: 113-124

[6] Li B, Xiao G, Lu R, et al. On feasibility and limitations of detecting false data injection attacks on power grid state estimation using D-FACTS devices. IEEE Transactions on Industrial Informatics, 2020, 16(1): 854-864

[7] Panice G, Luongo S, Gigante G, et al. A SVM-based detection approach for GPS spoofing attacks to UAV//Proceedings of the 2017 23rd International Conference on Automation and Computing. Huddersfield, UK, 2017: 1-11

[8] Akowuah F, Kong F. Physical invariant based attack detection for autonomous vehicles: Survey, vision, and challenges//Proceedings of the 2021 4th International Conference on Connected and Autonomous Driving. Detroit, USA, 2021: 31-40

[9] Cao Y, Xiao C, Cyr B, et al. Adversarial sensor attack on LiDAR-based perception in autonomous driving//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. London, UK, 2019: 2267-2281

[10] Tsai T, Yang K, Ho T Y, et al. Robust adversarial objects against deep learning models//Proceedings of the AAAI Conference on Artificial Intelligence. California, USA, 2020, 34(1): 954-962

[11] Merdrignac P, Shagdar O, Nashashibi F. Fusion of perception and V2P communication systems for the safety of vulnerable road users. IEEE Transactions on Intelligent Transportation Systems, 2016, 18(7): 1740-1751

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

[13] Park J, Ivanov R, Weimer J, et al. Sensor attack detection in the presence of transient faults//Proceedings of the ACM/IEEE 6th International Conference on Cyber-Physical Systems. Seattle, USA, 2015: 1-10

[14] Park P, Ergen S C, Fischione C, et al. Wireless network design for control systems: A survey. IEEE Communications Surveys & Tutorials, 2017, 20(2): 978-1013

[15] Khalajmehrabadi A, Gatsis N, Akopian D, et al. Real-time rejection and mitigation of time synchronization attacks on the global positioning system. IEEE Transactions on Industrial Electronics, 2018, 65(8): 6425-6435

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

[17] Zhao D, Ding S X, Karimi H R, et al. On robust Kalman filter for two-dimensional uncertain linear discrete time-varying systems: A least squares method. Automatica, 2019, 99: 203-212

[18] Wang Y, Liu Y, Fujimoto H, et al. Vision-based lateral state estimation for integrated control of automated vehicles considering multirate and unevenly delayed measurements. IEEE/ASME Transactions on Mechatronics, 2018, 23(6): 2619-2627

[19] Ravi A, Narasimhan S, Kaisare N S. Sampled output augmentation method for handling measurement delays in multirate Kalman filter. Chemical Engineering Science, 2020, 224: 115763

[20] Dorigoni D, Fontanelli D. An uncertainty-driven analysis for delayed mapping SLAM//Proceedings of the 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). Ottawa, Canada, 2021: 1-6

[21] Bai C Z, Gupta V, Pasqualetti F. On Kalman filtering with compromised sensors: Attack stealthiness and performance bounds. IEEE Transactions on Automatic Control, 2017, 62(12): 6641-6648

[22] Chang Y H, Hu Q, Tomlin C J. Secure estimation based Kalman filter for cyber-physical systems against sensor attacks. Automatica, 2018, 95: 399-412

[23] Manandhar K, Cao X, Hu F, et al. Detection of faults and attacks including false data injection attack in smart grid using Kalman filter. IEEE Transactions on Control of Network Systems, 2014, 1(4): 370-379

[24] Zeng T, Semiari O, Saad W, et al. Joint communication and control for wireless autonomous vehicular platoon systems. IEEE Transactions on Communications, 2019, 67(11): 7907-7922

[25] Yang T, Lv C. A secure sensor fusion framework for connected and automated vehicles under sensor attacks. IEEE Internet of Things Journal, 2021, 9(21): 22357-22365

[26] Chen Y, Zhang T, Kong F, et al. Attack-resilient fusion of sensor data with uncertain delays. ACM Transactions on Embedded Computing Systems, 2022, 21(4): 1-25

[27] Zhu Y, You Z, Zhao J, et al. The optimality for the distributed Kalman filtering fusion with feedback. Automatica, 2001, 37(9): 1489-1493

[28] Chong C Y, Chang K C, Mori S. Distributed tracking in distributed sensor networks//Proceeings of the 1986 American Control Conference. Seattle, USA, 1986: 1863-1868

[29] Song E, Zhu Y, Zhou J, et al. Optimal Kalman filtering fusion with cross-correlated sensor noises. Automatica, 2007, 43(8): 1450-1456

CHEN Yan-Feng, Ph. D. candidate. His research interest is embedded system.

DENG Qing-Xu, Ph. D., professor. His research interest is real-time embedded system and internet of things security.

ZHANG Tian-Yu, Ph.D.. His research interest is embedded system.

SUN Lei, Ph. D. candidate. His research interest is embedded system.

Background

Security issues have attracted much attention in cyber-physical systems (CPSs), and sensor attack is one of the most common ways to impact security of CPS. Multiple sensors are equipped with CPSs to measure physical state, then the measurements are transmitted to controllers to realize different functions. There is random sensing noise during the generation of measurements and random communication delays when transported to controller. Attacks can take charge of some of the sensors to generate fault measurements, leading controller to make wrong state estimation. Given multiple sensors with random noise and random communication delay, making high-precision physical state estimation under sensor attacks is a

challenging problem. Existing studies of process sensing noise with Kalman Filter methods to make accurate estimation of physical state. And some work focus on dealing with random communication delays with extra process, such as delay estimation and compensation. However, fusion methods based on Kalman Filter assume that sensing noise obeys zero-mean Gaussian distribution. The random delays are all positive value and not obeys Gaussian distribution. Thus, processing random delay with Kalman Filter will introduce extra errors. In this way, specific method should be proposed to handle random noise and random delay simultaneously to make high-precision fused value and resist malicious attacks. To tackle malicious attacks, some studies are proposed to identify sensors behaviors, which all follow a guideline that normal sensors have consistency between each other. However, to the best of our knowledge, no existing attack-resilient fusion method simultaneously considers random noise and random communication delay.

We propose multi-sensors fusion method based on Kalman Filter, combining a probability time window to detect sensor attacks. Given the probability distribution information of random delays, we compensate random delays and obtain the corresponding probability of a range of possible state estimation through the basic Kalman Filter. The confidence of the sensor working normally is evaluated by the cumulative value of the probability within the time window. Finally, the fused value is generated through Kalman Filter and the confidence, where the lower confidence sensors have lower influence of the fused value. The simulation validations show that the proposed method can simultaneously handle malicious attacks with sensing noise and random delay. The cumulative estimation error is 67% lower than the traditional Kalman filter algorithm under attacks and random delays. In addition, if the attacker obtains the full-knowledge of the proposed method and launches adaptive sensor attacks, the fused errors are limited to acceptable range, and the system can still work normally.

This work is partially supported by the National Natural Science Foundation of China (Nos. U1908212, 62072085, 62102073) and the Talent Project of Revitalizing Liaoning (No. XLYC1902017).