智联网传感器安全综述：机理、攻击和防护

徐文渊
浙江大学

关键词：传感器安全　换能攻击

智联网与传感器安全

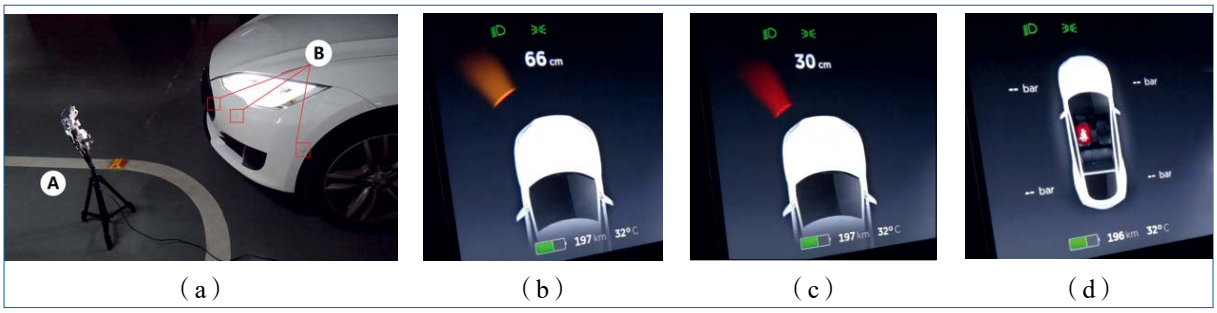
智联网的发展驱动智能进入实体环境。越来越多的智能设备具备了“眼睛”和“大脑”，能够动态感知、智能处理并自主回应来自周围环境的信息。而在这其中，智能设备的“眼睛”正是摄像头、雷达、麦克风、加速度计、陀螺仪、热电偶等各类传感器。目前，传感器已广泛应用于个人设备、智能家居、工业控制、自动驾驶、医疗诊断等智联网领域，并且随着智能设备种类和数量的显著增加（预计2022年将达到430亿[1]），传感器的数量也将突破1万亿大关[2]，在智联网中拥有举足轻重的地位。例如，随着汽车越来越智能化，一辆汽车中的传感器数量有望超过200个[3]，应用于从引擎控制到自动驾驶的各个环节。

传感器是一类通过换能（transduction）器件将各类物理能量信号转换为电能信号的设备，如图1所示。例如，热电偶可以产生与温度相关的电动势，麦克风可以通过压电效应将声音转换为相应的模拟电信号，加速度计可以将受力转换为电容变化。除了换能器，传感器中通常还有信号处理电路，对换能器产生的电信号进行放大、滤波等处理，并输出模拟或数字形态的测量值。智能设备通过热电偶和麦克风输出的测量值便可感知其所在环境的温度和声音，进而实现智能家居控制和语音助手等功能。

图1　传感器换能和感知原理

然而，传感器的安全风险在过去常被忽视。作为一类复杂度和智能化程度普遍较低的测量设备，传感器容易遭受来自实体环境的恶意攻击，发生故障或产生攻击者预期的错误测量结果，造成危及整个系统的安全后果。例如，已有研究[4]表明，通过发射恶意的电磁波信号可以使热电偶产生错误的测量结果，例如将正常室温（20°C）篡改为0°C

图2　通过电磁波攻击温度热电偶，使其产生错误的温度测量结果



（a） （b） （c） （d）

图 3 （a）攻击超声波避障传感器；（b）正常情况下的测量结果；（c）车检测到虚假障碍物；（d）汽车无法检测障碍物

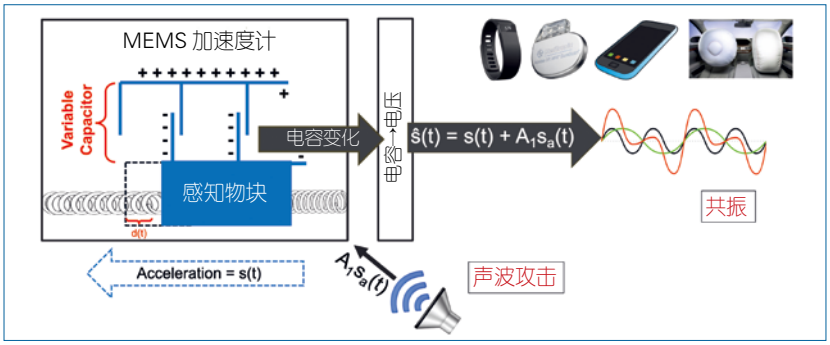


图 4 通过共振频率的声波攻击加速度计，使其在静止时测量到运动下的加速度

甚至 -800°C 等任意读数，如图 2 所示，此类攻击可能对核反应堆等基于温度的工业控制造成灾难性的影响。我们的研究 [5] 发现自动驾驶依赖的雷达等避障传感器可以被声波、电磁波和激光攻击，使得汽车无法有效检测障碍物并发生碰撞，或检测到虚假障碍物导致停车，如图 3 所示。传统设计通常只考虑传感器的测量噪声和有限误差，难以避免由攻击导致的感知结果篡改。如果上层智能系统无法主动发现感知失真，针对传感器的攻击则可能引发错误的系统决策和控制结果。虽然业界普遍会对传感器进行震动和电磁兼容等可靠性测试，但目前尚无针对传感器的安全测试，大量的传感器仍然缺乏安全设计。

目前，学术界已提出针对各种传感器的攻防方法，包括雷达、摄像头、加速度计、陀螺仪、麦克风、触摸屏、磁编码器等十余种最常见的传感器类型，涉及汽车、智能手机、医疗设备、能源等应用领域。然而，现有的攻击和防护方法大多为点对点（ad-hoc），与目标传感器和应用场景强相关，不同的工作之间不具备明显的关联性。为此，我们将此类针对传感器的攻击统一定义为“换能攻击”（Transduction Attack）[6, 7]，并使用统一的形式描述此类攻击的机理、方法和防护。

换能攻击机理

换能攻击是一类利用传感器的设计缺陷，通过物理攻击信号影响传感器测量结果的攻击方式。攻击者产生的物理攻击信号可以通过传感器中设计或非设计的换能过程转换为传感器内部的模拟电信号，从而影响传感器的输出。此类攻击通常不需要与传感器有实际的物理接触，因此具有较高的隐蔽性。图 4 展示了一个声波攻击的案例 [8]，攻击者通过声波可以引发 MEMS[1] 加速度计内部微型机械结构的共振，从而使加速度计在静止时测量到运动状态下的加速度值。此外，通过利用加速度计内部放大器、滤波器或模数转换器的缺陷，攻击者可以使传感器输出可控的加速度值。值得注意的是，换能攻击是利用传感器的设计缺陷造成不可信的测量值，因此不包括改变被测对象的攻击形式，例如使用热风枪加热温度传感器。虽然被加热的传感器无

[1] MEMS：Micro-Electro Mechanical System，微机电系统。

法可信地反映大环境的真实温度，但它依然可以可靠地测量它周围的温度。

攻击目标

换能攻击可能对传感器的测量结果进行两种形式的控制。

• 拒绝服务（Denial of Service，DoS）：此类攻击的目的是阻止传感器输出可用的测量值。例如，一个频率与陀螺仪固有谐振频率一致且非常强的声音干扰可能导致陀螺仪输出随机的角速度 [9]，使无人机无法维持稳定的飞行甚至发生坠毁。在拒绝服务攻击下，传感器的测量值通常是攻击者无法预测和控制的。

• 欺骗（Spoofing）：此类攻击的目标是造成传感器产生看似正常实则错误的测量值。例如，一段特殊制作的声音信号可以欺骗手机中的加速度计，从而控制与手机绑定的遥控汽车 [8]。与拒绝服务攻击的区别是，欺骗攻击通常可以部分或完全控制传感器的测量结果。

攻击信号

换能攻击可利用的物理信号包括但不限于以下几类：

• 电磁波是由互相垂直且同相振荡的电场与磁场产生的波，它的传播不需要依靠介质，在真空中的传播速度为光速。按照频率可以将电磁波分为无线电波、微波、红外线、可见光、紫外线、X 射线和伽马射线。

• 声波是一种由于介质的振动而产生的机械波，可以在气体、液体、固体中传播。按照声波的频率，可以将其分为次声波、可听声波和超声波。

• 磁场是由磁体、运动的电荷或电场的变化而产生的物理场。处于磁场中的磁性物质或电流会因为磁场的作用而感受到磁力。

• 电场是一种存在于电荷周围的物理场，对场中的其他电荷有作用力。

换能攻击机理

换能攻击的核心机理在于如何通过以上物理攻击信号影响传感器内部的模拟电信号，并利用传感器的设计缺陷实现攻击者预期的拒绝服务或欺骗攻击结果。由于传感器的多元异构（多于 387 类）且换能交互形式多（多于 61 种），不同的攻击使用来源广泛的攻击信号和方法，互相之间不具有明显的相似性，难以直接进行比较和归纳。为此，我们提出了基于传递函数的换能攻击机理模型 [7]。该模型可以描述已有换能攻击的通用性机理，预测可能的新攻击方法，并且帮助传感器工程师更好地将可测量的安全融入到传感器设计中。

如图 5 所示，我们可以将传感器的换能器、信号处理电路（放大器、滤波器等）和模数转换器等组件分别抽象表示为级联的传递函数 f_i，该函数可以将对应组件的输入输出信号描述为：$x_{i+1}=f_i(x_i)$。其中传递函数的输入为 x_i，输出信号为 x_{i+1}。在此模型的基础上，换能攻击机理可以描述为攻击者通过产生物理攻击信号 a_i，向传感器组件 f_i 中注入恶意噪声干扰，并利用 f_i 的固有特性构造预期的组件输出 x_{i+1} 和传感器测量值 y。因此，我们可以统一地将不同的换能攻击分解为信号注入和信号整形两个步骤。

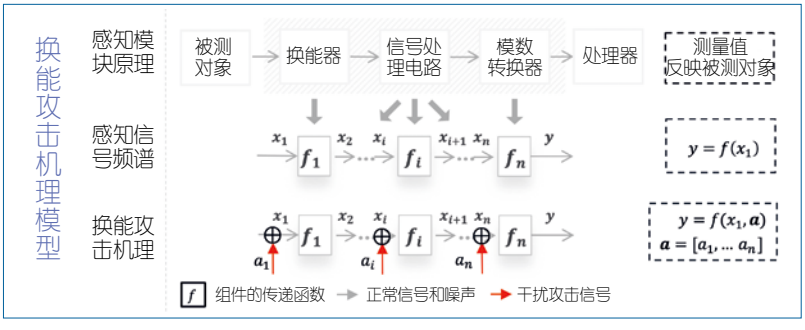


图 5　换能攻击机理模型

• 信号注入：该步骤对应模型中攻击信号 a_i 转换为传感器组件 x 的输入。攻击者在这一步骤需要考虑的因素为如何利用传感器中设计或非设计的换能过程将攻击信号转换为传感器内部的模拟电信号，主要包括信号注入点（可

注入的传感器组件）和攻击信号参数等。

• 信号整形：简单的信号注入不一定能产生攻击者期待的传感器测量值。因此，信号整形步骤对应模型中攻击者将注入的信号整形为预期的组件输出 x_{i+1}，主要涉及如何构造可以利用传感器固有特性的恶意物理信号，使其可以不被传感器滤除并造成预期的传感器测量值。

换能攻击方法

我们对已有换能攻击使用的通用性信号注入和信号整形步骤进行介绍。

信号注入

为成功且高效地向传感器中注入恶意信号，攻击者需要同时考虑信号的注入点以及信号的类型、幅度和频率等影响信号注入效率的参数。

注入点与信号类型 由于换能器是传感器在设计上唯一接收物理信号的组件，因此我们将信号注入点分为“换能器前”与“换能器后”两类。传感器中存在的信号注入点决定了攻击信号的类型。（1）换能器前的注入点：攻击者可以利用换能器在设计上可以接收至少一种物理信号的特性来注入同种类型的恶意物理信号。大部分已有的换能攻击都是利用此类注入点，例如通过激光攻击激光雷达[10]、通过超声波攻击超声波避障传感器[5]、通过声波（一类机械波）攻击测量机械振动的陀螺仪[9]和加速度传感器[8]等。（2）换能器后的注入点：换能器之后的传感器组件也可能成为无意的换能器，尽管这并非它们在设计之初的用途。例如，电路中的导线可能由于电感耦合或电容耦合将环境中的电磁波转换为导线中的电信号，成为无意的天线[11]。

信号幅度与频率 信号注入的有效性和效率通常取决于物理信号的频率、幅度等参数。（1）幅度：更高幅度的攻击信号可以实现更好的注入效果和更远的攻击距离。在实际攻击中，攻击者不可能无限地增加功率。大功率的电磁波、激光和声音可能会对人体造成损伤；无声的超声波在大功率时会因为非线性声学产生可听的声音，使原本无声的攻击变得有声[12]。（2）频率：MEMS 结构的换能器和导线（天线）等许多器件都有共振频率，当攻击信号处于这些频率时，可以在相同的强度下实现更强的注入结果。攻击者可以通过寻找传感器的共振频率实现更高效的信号注入。除了注入效率，某些频率范围的攻击信号（如超声波和红外线）不会被人类感知，相比其他频率具有更隐蔽的攻击效果。

信号整形

信号整形步骤利用传感器组件的特性对注入的信号进行进一步修改，从而构造攻击者预期的测量结果。本节以饱和、交调失真（intermodulation distortion）、混叠（aliasing）等特性为例进行简要介绍。

饱和 指某种物理量无法超过一定的阈值，是模拟电路中常见的一种现象。例如，当输入超过一定阈值时，一个放大器就可能进入饱和状态，此时放大器的输出变为与供电电压相关的常量，不会继续随着输入线性增长。类似的饱和现象也可能发生在其他组件上，例如换能器和模数转换器。攻击者可以通过注入较强的恶意信号有意地造成某个组件进入饱和状态，从而掩盖正常信号造成拒绝服务攻击，或者引入直流分量造成欺骗效果。

交调失真 放大器、二极管、换能器等常见的器件都具有一定的非线性，即其输入输出不以线性规律变化。当一个含有多个频率分量的信号经过一个非线性器件时就会发生交调失真。交调失真会导致输出信号中出现输入信号里不包含的频率分量，例如输入信号频率的和、差与倍数频率。攻击者可以利用传感器的非线性器件，对经过调幅调制的攻击信号实现类似解调的效果。我们提出的“海豚音攻击”[13]，正是利用麦克风的非线性特性将调制的超声波攻击信号解调，实现无声语音指令的攻击效果。类似地，通过高频调制电磁波也可以向录音笔中注入低频的语音信号[11]。

混叠 根据“奈奎斯特－香农采样定理”，如果一个信号的频率高于采样率的一半，那么这个信号将与特定频率的信号无法区分。这种现象被称为

混叠，一般是信号处理中需要避免的。然而，攻击者可以利用传感器中滤波器的一些不足，制造模数转换时的混叠现象，从而将高频的攻击注入信号转换为低频、可控的采样结果。例如，特里佩尔（Trippel）等人[8]和Tu等人[14]通过调整声音信号的频率、幅度和相位控制混叠结果，实现了对MEMS加速度计和陀螺仪输出的精准控制。

攻击者可以根据目标传感器和预期的攻击效果，将不同信号注入和信号整形步骤串联组合，形成换能攻击的攻击链。在设计换能攻击时，攻击者可以通过分析目标传感器的信号处理链路确定可利用的信号注入和信号整形步骤，并基于此构造恶意的物理信号。

传感器安全防护方法

换能攻击的防护思路可以分为两类，分别是攻击检测与攻击抵御。攻击检测旨在检测换能攻击的存在，而攻击抵御则是避免攻击对传感器的测量值造成影响，让传感器在攻击下也有可信的输出。

攻击检测

攻击检测方法仅检测是否有换能攻击存在，并不能抵御攻击对传感器测量值的影响。然而，此类方法可以成为更加鲁棒的系统级防御的起点，例如作为采取攻击抵御方法或开启失效安全（fail-safe）机制的前提条件。攻击检测可以分别针对信号注入和信号整形进行设计。

检测信号注入步骤　防御者可以利用额外的换能器有针对性地检测环境中存在的攻击信号。例如，使用额外的麦克风可以检测到攻击MEMS加速度计和陀螺仪的共振频率声音[8]。同时具有传感器和执行器的系统可以主动检测当攻击发生时测量结果的某些特性与预期不符的情况。

检测信号整形步骤　饱和使得器件进入异常的工作状态，这种异常通常比较容易通过硬件或软件进行检测。交调失真可能会在模拟信号中留下可识别的特征，例如在500 Hz~1 kHz频段或50 Hz以下频段产生异常的信号强度或与高频信号相关的特性[13, 15]。

攻击抵御

抵御方法确保传感器即使在换能攻击下也能输出可信的测量值。此类方法通常需要在传感器外部或内部衰减恶意信号，包括屏蔽、滤波、随机化、改进组件质量和传感器融合。

屏蔽　通过减少传感器对外部信号的暴露来抵御信号注入，主要包括物理隔离和攻击面缩减。物理隔离的目的是衰减进入传感器的外部物理干扰，例如电磁屏蔽、隔音、光屏蔽等。攻击面缩减指有选择地在空间、时间或频域上限制换能器对外部的暴露，在确保换能器接收正常物理信号的同时增加攻击的难度。

滤波　旨在不影响正常信号的情况下衰减恶意信号。例如通过设计合适的低通滤波器去除高频攻击信号，避免交调失真或混叠的发生。当简单的滤波器不足时，防御者可以捕获环境中的攻击信号并使用自适应滤波有针对性地去除电路中的恶意信号。此外，防御者还可以设计异相采样（out-of-phase sampling）等特殊的采样模式，抵御利用混叠效应的攻击。

随机化　增加传感器的随机性通常可以弱化攻击的影响。随机性可以添加在换能器、模数转换器和后端控制器等传感器组件中，包括输入随机化与输出随机化。输入随机化将随机性增加在传感器输入信号流的控制上，例如将模数转换器的采样率进行随机变化，由于随机参数对于传感器来说是已知的，因此不会影响传感器的正常测量。输出随机化利用相似的思路增加传感器探测波形的随机性，适用于主动式传感器。

改进组件质量　通过重新设计传感器硬件，改进传感器中有缺陷的组件质量可以从根本上抵御一些换能攻击。例如，使用具有足够动态范围的放大器可以避免攻击者利用饱和现象；重新设计MEMS陀螺仪可以缩减共振频率的范围，降低其影响，从而避免声波注入。

传感器融合 由于攻击者难以同时攻击所有的传感器，因此通过融合多个或多种传感器在不同空间、时间或频率上的测量结果可以在一定程度上抵御攻击的影响。

总结

在智联网的发展下，越来越多的智能设备通过传感器感知实体环境，成为了与物理世界紧密交互连接的复杂信息物理系统。因此，智联网安全的关注点应该从组件转为系统的各个层面，并且确保系统在具有不可信的组件时也能够可靠工作。本文探讨了传感器特有的安全问题及其对系统层的影响，从机理、攻击和防护的角度简要梳理了传感器的安全风险与攻防方法，希望可以为读者提供有价值的参考。 ■

参考文献

[1] Statista Research Department. Global IoT forecast: sensors market breakdown by segment 2022[OL]. https://www.statista.com/statistics/480114/global-internet-of-things-enabled-sensors-market-size-by-segment/.

[2] EETimes. A Trillion Sensors Are on the Way[OL]. https://www.eetimes.com/a-trillion-sensors-are-on-the-way/.

[3] Els Verlinden. The sense of virtual sensors[OL]. https://blogs.sw.siemens.com/simcenter/the-sense-of-virtual-sensors/.

[4] Tu Y, Rampazzi S, Hao B, et al. Trick or Heat? Attack on Amplification Circuits to Abuse Critical Temperature Control Systems[J]. 2019.

[5] Yan C, Xu W, Liu J. Can You Trust Autonomous Vehicles: Contactless Attacks against Sensors of Self-Driving Vehicles[J]. *Technical Report*, 2016.

[6] Fu K, Xu W. Risks of trusting the physics of sensors[J]. *Communications of the ACM*, 2018, 61(2):20-23.

[7] Chen Y, Hocheol S, Connor B, et al. SoK: A Minimalist Approach to Formalizing Analog Sensor Security[J]. 2020.

[8] Trippel T, Weisse O, Xu W, et al. WALNUT: Waging Doubt on the Integrity of MEMS Accelerometers with Acoustic Injection Attacks[C]// *2017 IEEE European Symposium on Security and Privacy* (EuroS&P). IEEE, 2017.

[9] Son Y . Rocking Drones with Intentional Sound Noise on Gyroscopic Sensors[M]. 2015.

[10]H. Shin, D. Kim, Y. Kwon, and Y. Kim, ‘Illusion and Dazzle: Adversarial Optical Channel Exploits Against Lidars for Automotive Applications’, in Cryptographic Hardware and Embedded Systems – CHES 2017, Cham, 2017, vol. 10529, pp. 445–467, doi: 10.1007/978-3-319-66787-4_22.

[11]D. F. Kune et al., ‘Ghost Talk: Mitigating EMI Signal Injection Attacks against Analog Sensors’, in 2013 IEEE Symposium on Security and Privacy, Berkeley, CA, May 2013, pp. 145–159, doi: 10.1109/SP.2013.20.

[12]C. Yan, G. Zhang, X. Ji, T. Zhang, T. Zhang, and W. Xu, ‘The Feasibility of Injecting Inaudible Voice Commands to Voice Assistants’, IEEE Trans. Dependable Secure Comput., pp. 1–1, 2019, doi: 10.1109/TDSC.2019.2906165.

[13]G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, ‘DolphinAttack: Inaudible Voice Commands’, in Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security - CCS ’17, Dallas, Texas, USA, 2017, pp. 103–117, doi: 10.1145/3133956.3134052.

[14]Y. Tu, Z. Lin, I. Lee, and X. Hei, ‘Injected and Delivered: Fabricating Implicit Control over Actuation Systems by Spoofing Inertial Sensors’, p. 19, 2018.

[15]N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury, ‘Inaudible Voice Commands: The Long-Range Attack and Defense’, 2018, p. 14.

徐文渊

CCF 高级会员。浙江大学教授。主要研究方向为无线网络安全、嵌入式系统安全、高可靠性传感器网络、医疗器械安全。wyxu@zju.edu.cn