


浙江大学电气工程学院
College of Electrical Engineering Zhejiang University

网络安全导论

智能无人系统安全

1. 概述、基础知识
2. 加密与认证技术
3. 软件与通讯安全
4. 电力工控系统安全
5. 物联网终端安全
6. **智能无人系统安全**

目录

6.1 定义、组成、分类、主要特性

智能无人系统介绍

6.1 智能无人系统介绍

1. 智能无人系统的定义

2. 智能无人系统的组成

3 .智能无人系统的分类

4. 智能无人系统的特性

5.1 智能无人系统概述

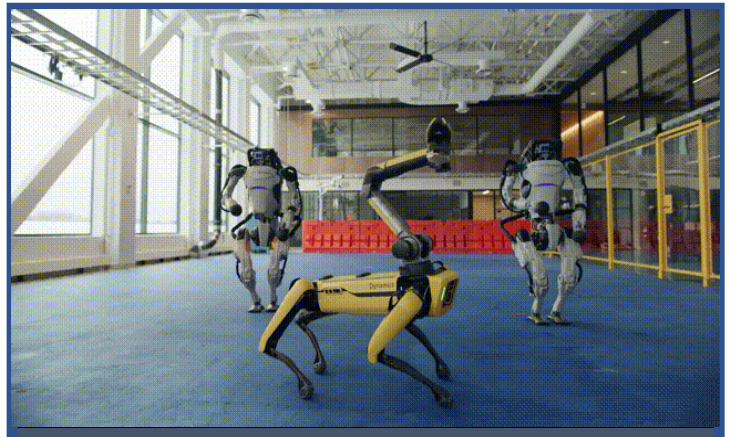
- **智能无人系统**是一种集成了先进的人工智能技术，能够自主执行任务，而无需人类直接干预的系统。
- 随着人工智能（AI）、技术的发展，以机器人、无人驾驶汽车和无人机为代表无人系统开始代替人类从事各种场景中简单或者复杂的工作。
- 随着物联网的快速发展，智能无人系统在发展经济、推动社会进步方面的重要作用越来越明显，但其存在的安全问题也不容忽视。

无人驾驶系统的硬件组成

机器人失控伤人

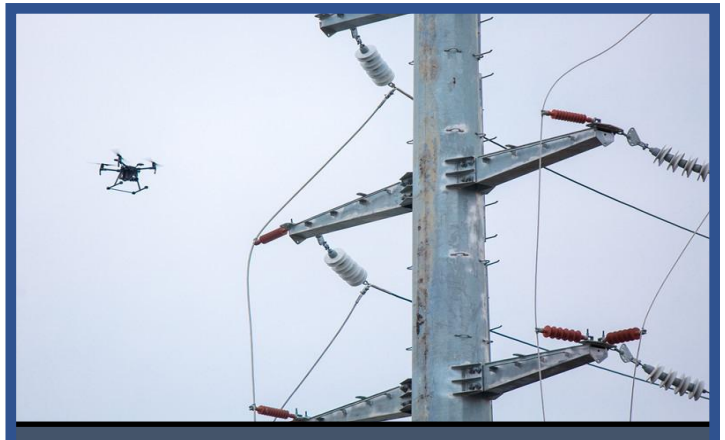
5.1.1 智能无人系统的定义

- **定义**：利用各类传感器、执行器和智能/控制算法实现自主功能的无人系统，含有“感知-计算-控制”环节。
- **实例**：机器人、自动驾驶汽车、无人机……
- **相关**：信息物理系统、自主无人系统……



机器人舞蹈

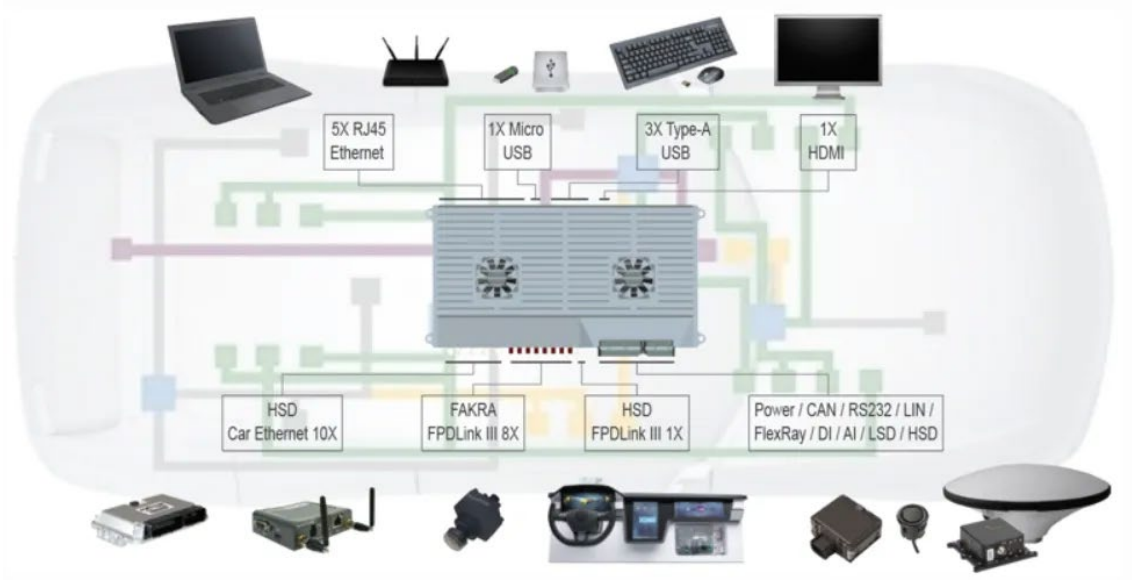
自动驾驶



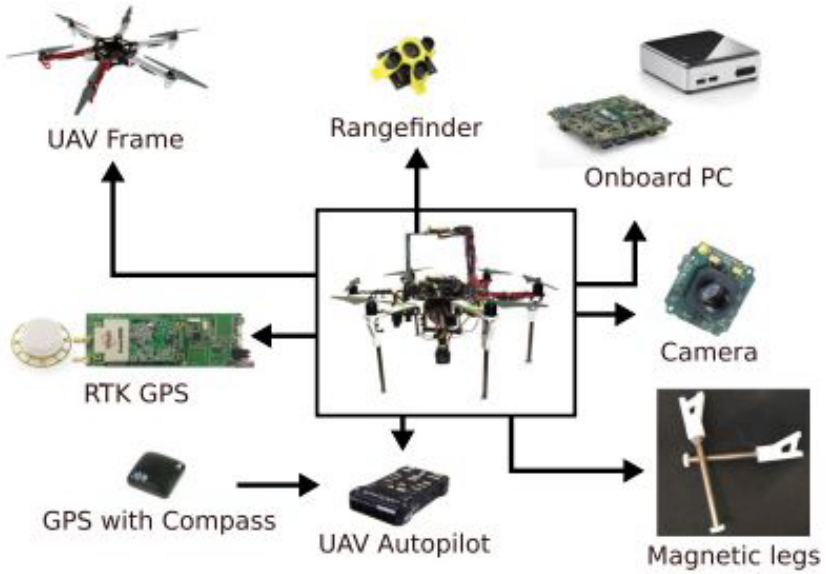
无人机巡检

5.1.2 智能无人系统的组成

- **智能无人系统**一般由**传感模块、数据处理模块、算法决策模块、执行模块、通信模块、电源模块、导航模块、故障处理模块、人机交互界面、机械结构等**部分组成。
- **算法和决策模块**是智能无人系统任务执行过程的**核心**。



自动驾驶汽车的执行与控制模块



无人机的硬件结构图

5.1.3 智能无人系统的分类

- 由于传感器应用领域众多，适用范围又广，其品种和规格繁多，根据不同的原则可以将传感器分成不同类型。比较常用的分类方法有以下几种。

1 按应用场景分类

- 自动驾驶载具
- 机器人系统
- 农业智能系统
- 救援和灾害响应系统
- 物流和仓储系统
- 无人飞行系统
- 水下和水面系统
- 建筑和施工系统
- 医疗保健系统
- 环境监测系统

2 按操作环境分类

- 陆地无人系统
- 空中无人系统
- 水下无人系统
- 水面无人系统

3 根据规模和大小分类

- 微型无人系统
- 小型无人系统
- 大型无人系统

5.1.3 智能无人系统的特性

- **自主性（Autonomy）** 智能无人系统具备自主决策和执行任务的能力，能够在特定环境中感知、理解和应对变化，而无需持续的人工控制。
- **感知能力（Sensing Capabilities）**： 这类系统通常配备各种传感器，以感知周围环境的信息。传感器的种类可以包括视觉、声音、雷达、激光等，这些信息帮助系统理解其工作环境。
- **决策能力（Decision-making Capabilities）**： 智能无人系统具备处理感知信息、做出决策的能力。这可能包括使用预定的算法、模型或深度学习技术来解释感知数据并做出相应的行动。
- **执行能力（Execution Capabilities）**： 该系统能够根据其决策自主执行任务。这可能涉及到机械执行，如移动、抓取物体，或者其他与系统设计目标相关的操作。
- **通信能力（Communication Capabilities）**： 智能无人系统通常能够与其他系统或者中央指挥中心进行通信，以便接收指令、报告任务进展，或者获取实时的环境信息。
- **适应性（Adaptability）**： 智能无人系统可能具备适应不同环境和任务的能力，能够灵活应对各种情境。
- **安全性和可靠性（Safety and Reliability）**： 智能无人系统需要被设计为安全可靠，以确保其在执行任务时不会对人类和环境造成危害。
- **多模态性（Multimodality）**： 这类系统可能整合多种感知模态，如视觉、声音和其他传感器，以更全面地理解其周围环境。

6.2 安全事件、“感-算-控”架构、安全问题

智能无人系统安全概述

6.2 智能无人系统安全概述

1. 安全事件

2. 智能无人系统“感-算-控”通用架构

3. 智能无人系统存在的安全问题

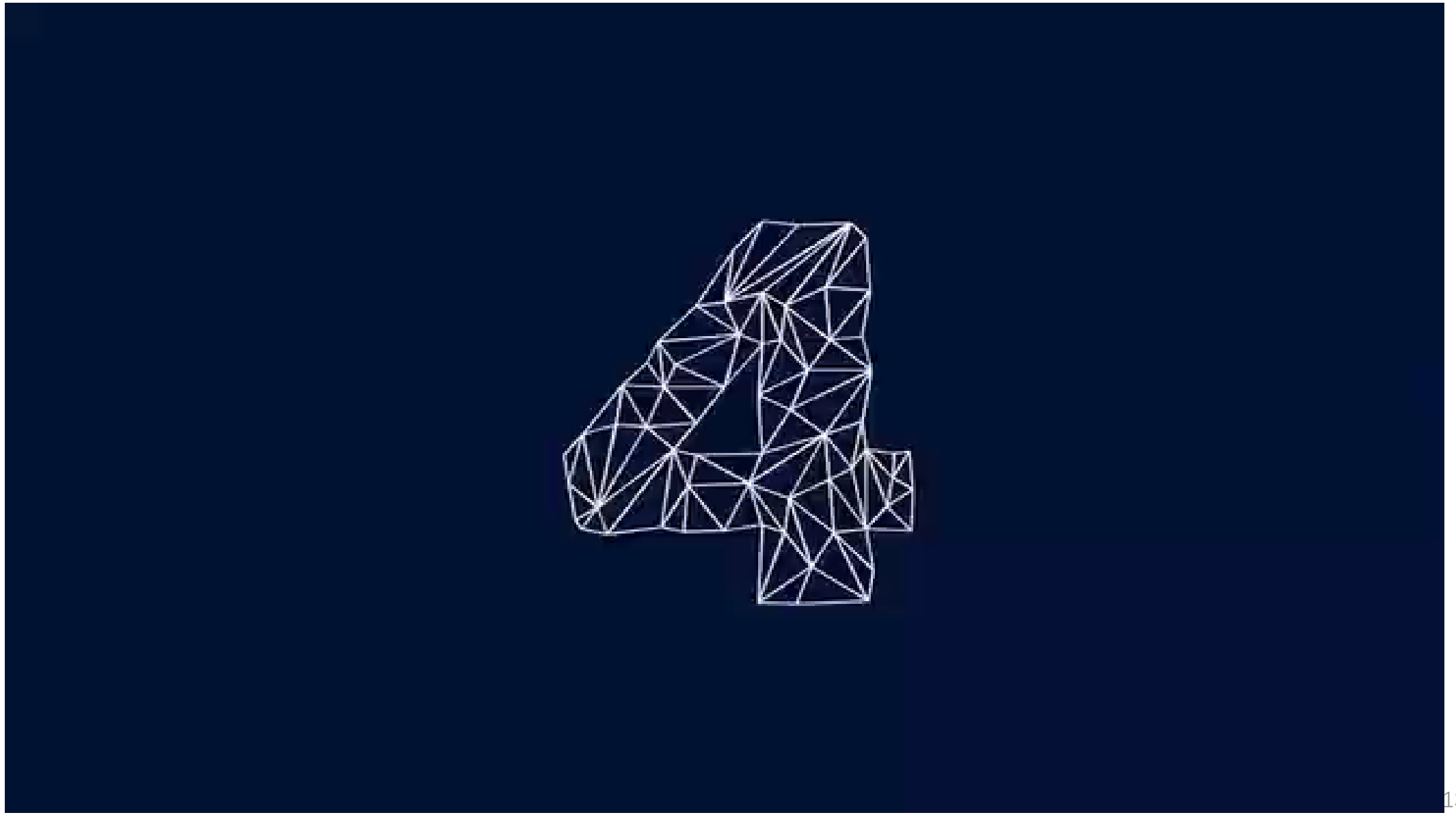
6.2.1 安全事件

- 近年来智能无人系统安全事件频发，以自动驾驶汽车为例，特斯拉的自动驾驶系统的安全可靠性引发广泛关注。

案例1：2020年6月，台湾一特斯拉撞上一辆侧翻在地的白色货柜车。

案例2：一辆特斯拉汽车在一片无人墓地中检测出许多“鬼魂”。

6.2.1 安全事故



6.2.1 安全事件

俄乌战场无人机武装袭击

无人机群失控导致社会人员受伤

软件漏洞可导致网联汽车被远程控制引发失控

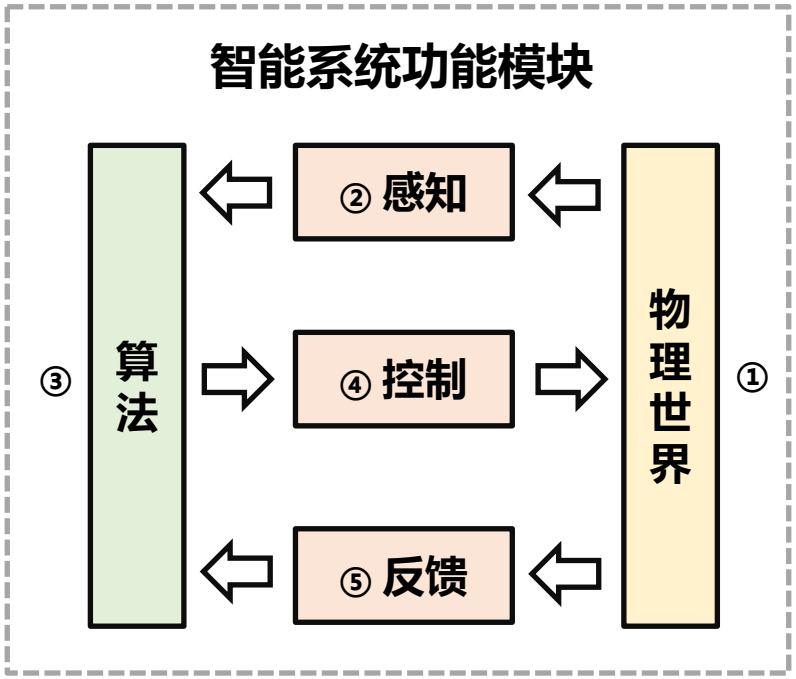
2011年信号感知错误导致甬温线特大铁路交通事故

2019年迎角传感器输出错误导致埃航波音客机坠毁

1996年惯性基准无输出导致长三乙火箭失控爆炸

6.2.2 智能无人系统“感 - 算 - 控”通用架构

- **感知**：将人类直观感知的信息变换成系统可处理的信息如电信号等，是系统与物理世界沟通的桥梁
- **算法**：具有知识学习、逻辑推演等能力的信息处理手段，是实现系统智能化的关键，AI（人工智能）
- **控制**：按照系统指令，精准操纵执行机构
- **反馈**：将执行结果送回系统，形成闭环回路



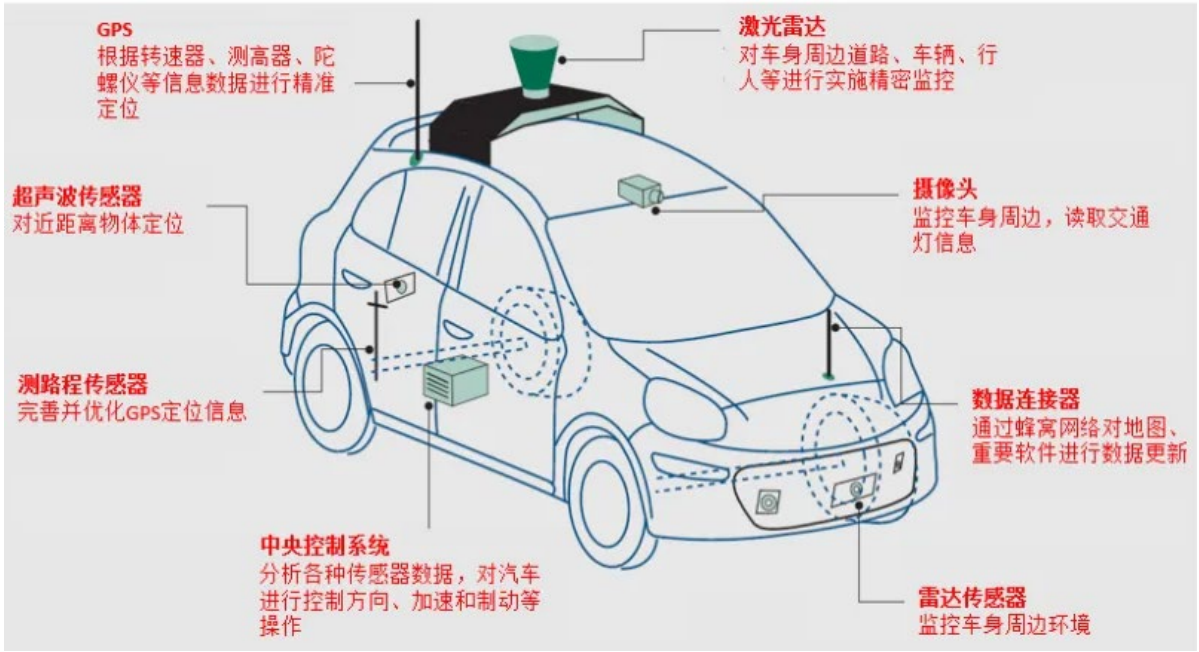
6.2.2 智能无人系统“感-算-控”通用架构

“感 - 算 - 控”通用架构

这种"感-算-控"的通用架构反映了智能无人系统的基本工作原理，强调了从环境感知到决策制定再到实际控制执行的流程。

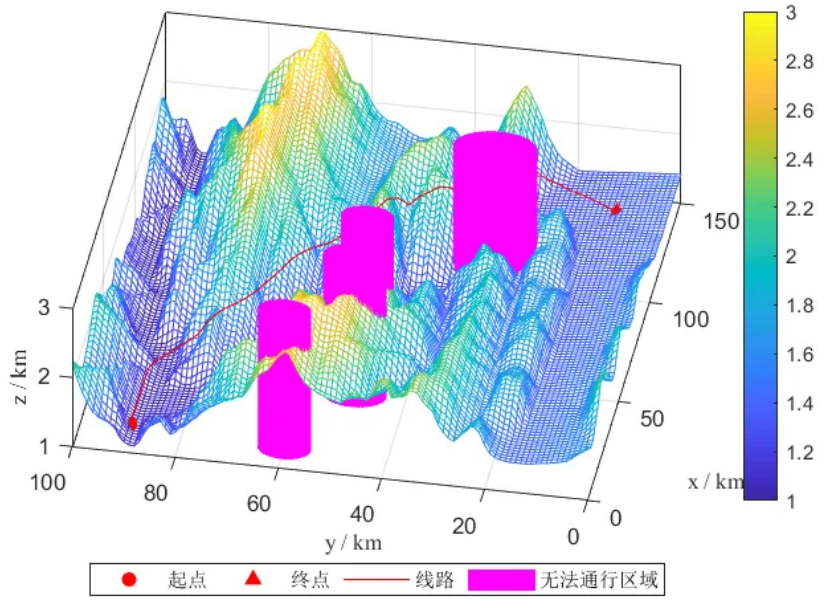
感知（Perception）

- **传感器**： 包括摄像头、激光雷达、超声波传感器、雷达、红外线传感器等，用于获取系统周围环境的信息。
- **数据融合**： 将来自不同传感器的数据融合在一起，形成对环境的整体理解。
- **环境建模**： 利用感知数据构建环境模型，描述系统周围的物体、障碍物、道路等。

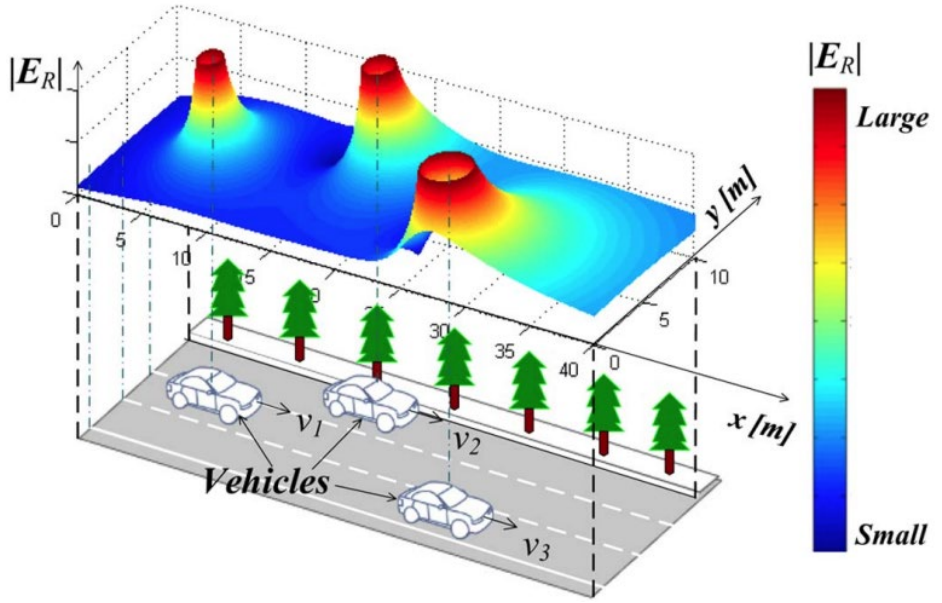


6.2.2 智能无人系统“感-算-控”通用框架

- **计算（Cognition）：**
  - **决策制定：** 利用感知数据和环境模型，采用算法和模型进行决策，确定系统的下一步行动。
  - **路径规划：** 对于移动系统，规划合适的路径，避开障碍物，确保安全到达目的地。
  - **智能算法：** 利用智能算法进行模式识别、行为预测等任务，以适应不断变化的环境。



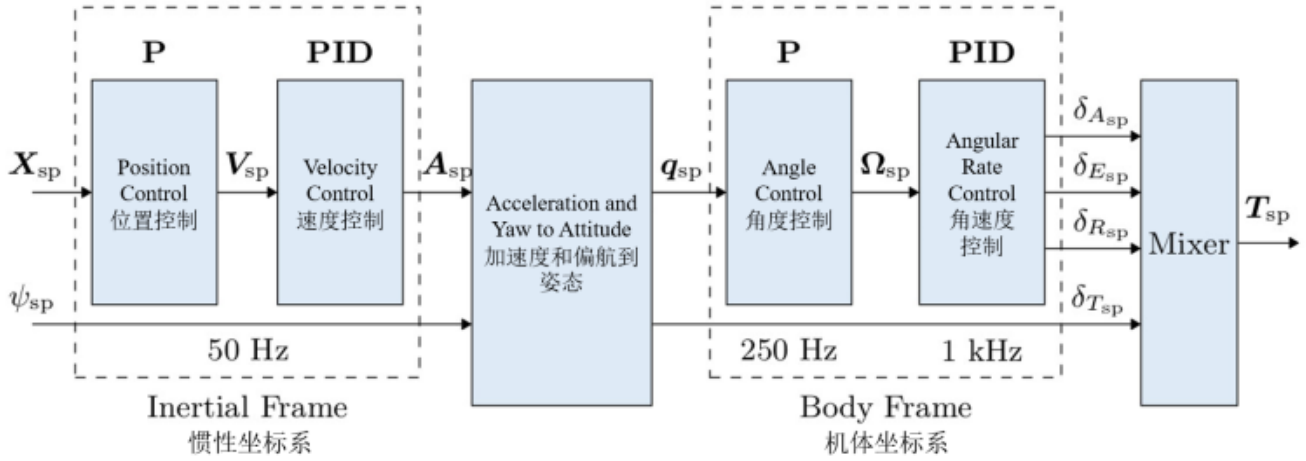
无人机路径规划仿真示意图



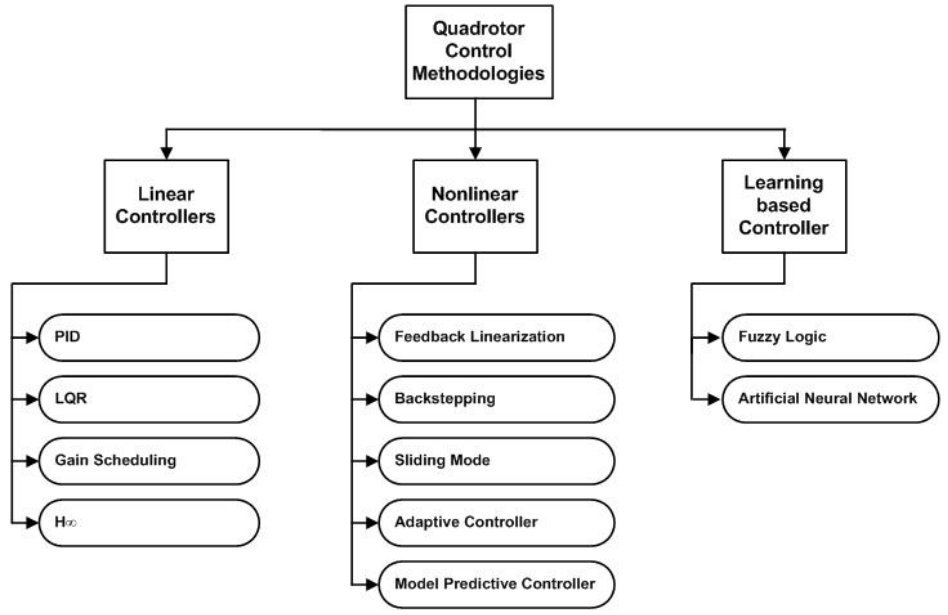
自动驾驶场景中利用人工势场法进行安全域划分

6.2.2 智能无人系统“感-算-控”通用框架

- **控制（Control)**
  - **执行控制指令**： 根据计算模块的决策，执行相应的控制指令，例如调整航向、速度、姿态等。
  - **动作执行**： 执行实际的动作，例如机器人的移动、手臂的操作、车辆的转向等。
  - **状态实时调整**： 根据实时的感知数据和环境变化，调整控制策略，保持系统的稳定性和适应性。



无人机串级控制环路



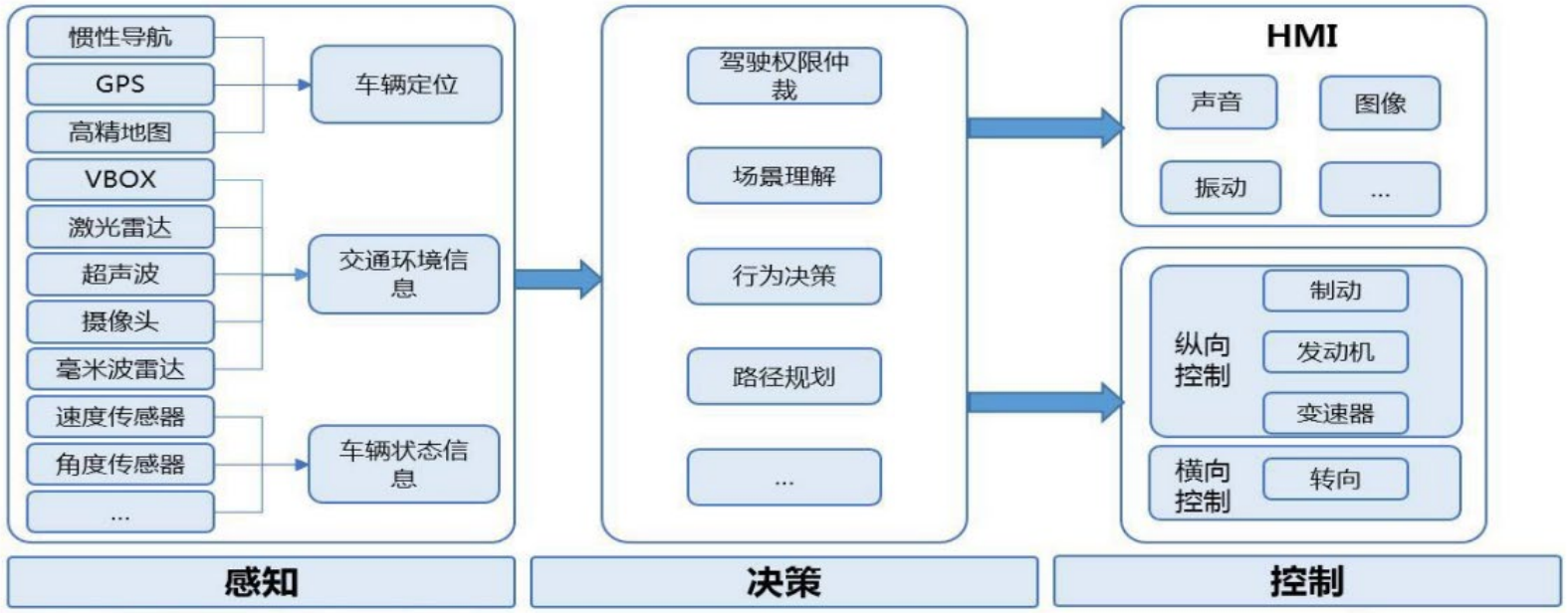
四旋翼无人机控制方法

6.2.2 智能无人系统“感-算-控”通用框架

- 不同的应用领域和系统中，这个框架可能会有一些变化，根据具体的需求和技术要求进行调整。

➢ 自动驾驶汽车的“感-算-控”架构可能会包括更多关于车辆动力学的控制方面；

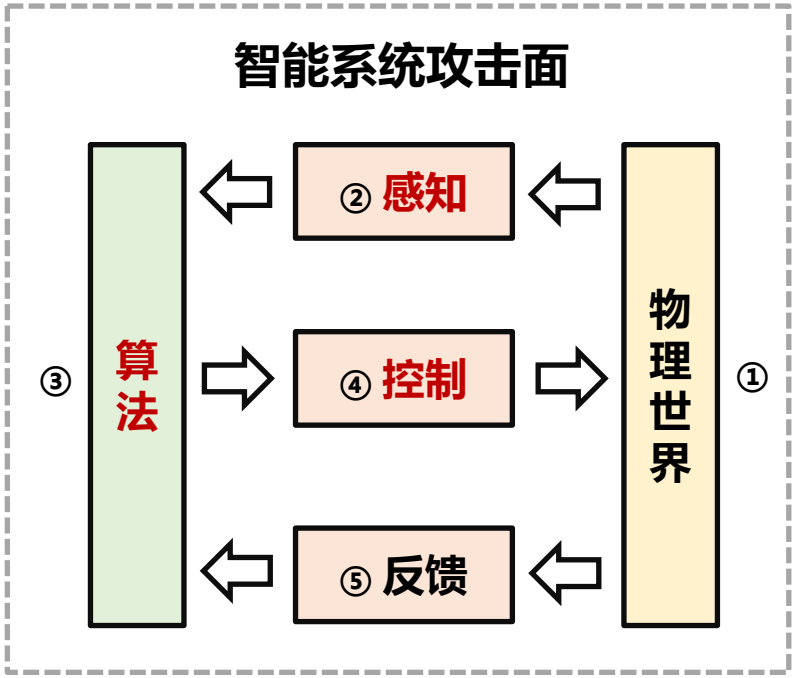
➢ 机器人系统可能更注重对周围环境的高度敏感的感知系统。



自动驾驶汽车的“感知-决策-控制”框架

6.2.3 智能无人系统的安全问题

- **1. 感知安全：**感知器件安全设计欠缺，面临恶意信号注入等安全威胁。
- **2. 算法安全：**例如AI算法具有不可解释性，面临恶意对抗样本等安全威胁。
- **3. 控制安全：**执行器件设计欠缺，控制逻辑漏洞，“由感到控”。
- **4.通信安全**：智能无人性系统与其他设备、中央控制中心或云服务进行通信过程中存在的潜在威胁。



6.3 定义与组成、安全事件、安全问题、攻击案例

感知安全

智能无人系统中的感知安全

6.3 感知安全

1. 定义和组成
2. 安全事件
3. 安全问题
4. 攻击案例

6.3.1 定义和组成

- **感知环节在智能无人系统中的定义：**
  - 在智能无人系统中，感知环节是系统获取外部环境信息的过程，通常通过各种传感器来实现。
  - 智能无人系统的常通过特定的传感器采集相关数据，并在完成预处理和融合后，实现对感知数据的利用，完成状态估计、环境建模、异常检测等任务。

- **感知环节在智能无人系统中的作用：**

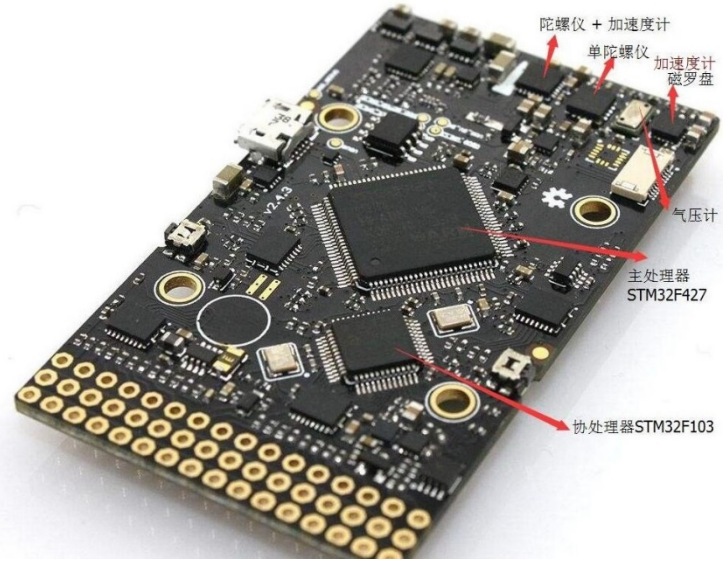
感知环节在智能无人系统中扮演着关键角色，通过使用各种传感器获取外部环境信息，包括图像、声音、距离等。其主要作用包括实时环境感知、建立环境模型、障碍物检测与避障、导航与路径规划、环境变化检测、人体检测与识别等。感知环节为系统提供了对周围环境的全面理解，为后续决策和控制阶段提供基础，直接影响系统的适应性和安全性。一个有效的感知环节是智能无人系统高效、安全运行的关键之一。

6.3.1 定义和组成

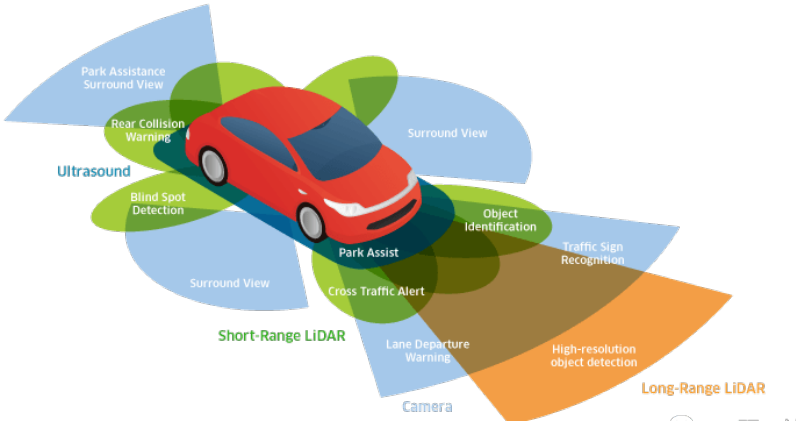
感知环节的软硬件组成：

一般来说，感知模块由传感器及其配套处理模块组成的。

- **传感器：**
  1. **摄像头：** 用于捕捉视觉信息，例如图像和视频。
  2. **激光雷达：** 用于测量距离和检测物体的位置。
  3. **超声波传感器：** 用于距离测量和障碍物检测。
  4. **GPS（全球定位系统）：** 用于确定系统的位置和方向。
  5. **IMU（惯性测量单元）：** 测量系统的加速度和角速度。
- **计算设备：**
  1. **嵌入式处理器：** 用于实时处理传感器数据，执行感知算法和模型。
  2. **图形处理单元（GPU）：** 用于加速图像处理和计算密集型任务。
  3. **中央处理单元（CPU）：** 执行感知算法和决策制定过程。
- **通信模块：**
  1. **无线通信设备：** 用于与其他系统、中央控制中心或云进行通信。
  2. **数据传输设备：** 负责将感知数据传输到其他部分进行处理或决策。



飞控板载传感器组成



自动驾驶汽车上的传感器

6.3.2 感知出错导致的安全事件

- 智能无人系统以大数据为支撑，传感器数据如各类视频、图片、声音、加速度等数据是算法的基石；
- 跨媒体AI将更加依赖传感器数据的输入，如无人机、自动驾驶汽车；
- 智能无人系统的**感知安全**是安全的**决定性因素**，而目前对其研究不充分。


Boeing admits full responsibility for 737 Max plane crash in Ethiopia

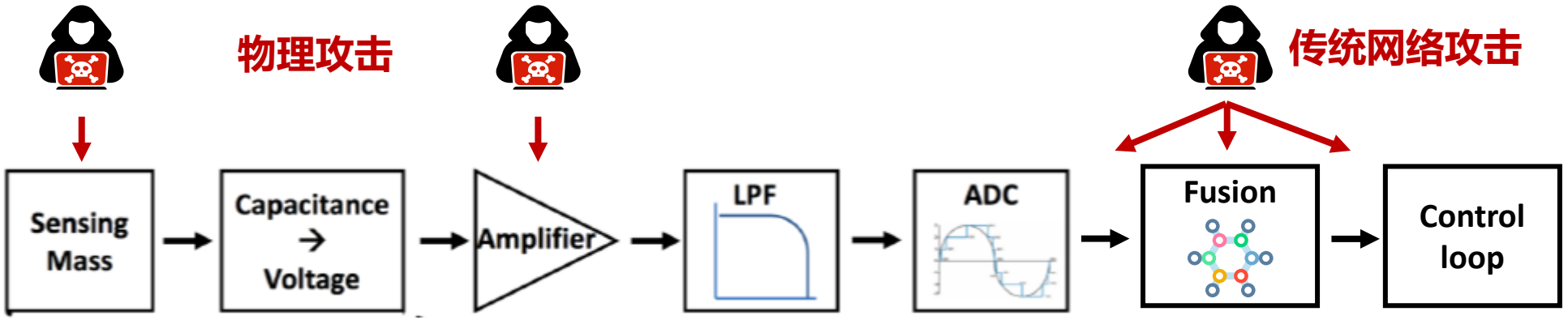
'Significant milestone' paves way for families of 157 victims of 2019 crash to seek compensation, say lawyers

波音737空难中迎角传感器的安全问题
迎角传感器反馈错误数据，触发客机“死亡俯冲”

自动驾驶汽车中传感器问题导致安全事故（特斯拉）

6.3.3 感知环节安全问题

- 感知环节的安全问题可以描述为，**感知环节的输出信号是否是可信的。**
- 感知环节安全问题包括：**传统网络攻击**和**物理攻击。**



6.3.3 感知环节的安全问题

- **传统网络攻击**
  - **数据篡改和注入**
    - **中间人攻击**： 攻击者可能截获传感器数据，篡改数据后再传递给系统，使系统基于错误的信息做出决策。
    - **数据注入**： 攻击者可能通过恶意设备注入虚假的传感器数据，引导系统做出错误的判断。
  - **网络监听和隐私窃取**
    - **窃听攻击**： 攻击者可能通过监听传感器数据的通信链路，获取敏感信息，了解系统的工作状态和环境信息。
    - **隐私窃取**： 分析传感器数据的通信流量，推断系统的行为和决策，窃取控制信息。
  - **虚假传感器**
    - **模拟传感器攻击**： 攻击者可能模拟传感器信号，向系统发送虚假的传感器数据，欺骗系统对环境的理解。
    - **传感器冒充**： 攻击者可能冒充合法传感器设备，向系统提供虚假的身份和数据。

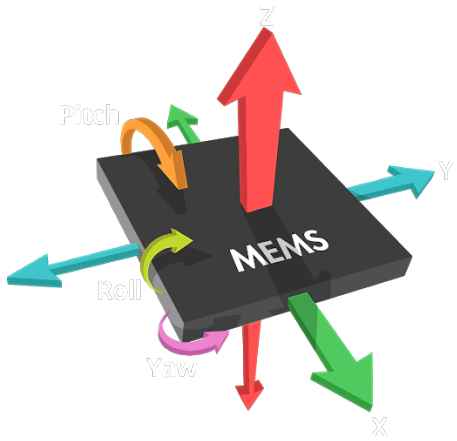
6.3.3 感知环节的安全问题

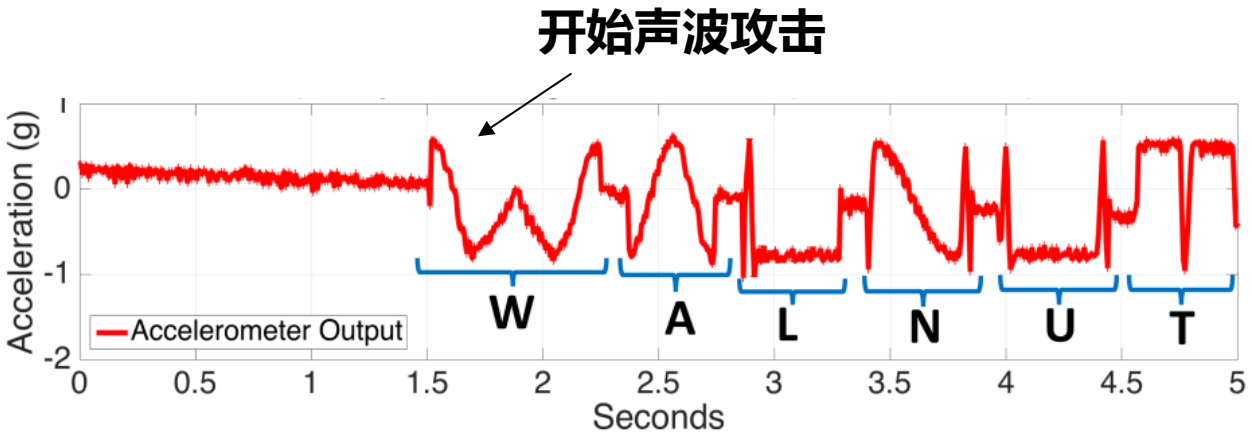
- **物理攻击**
  - 利用传感器自身脆弱性
    - **电磁干扰**： 攻击者可能使用电磁干扰设备干扰传感器的正常运行，导致传感器数据不准确。
    - **信号屏蔽**： 对无线传感器使用信号屏蔽设备，阻止其正常通信，导致数据丢失或延迟。
    - **换能攻击**：一类利用传感器设计缺陷，通过对被感知物理信号作出某种影响来改变传感器测量结果的攻击。
  - 物理攻击感知算法：
    - **对抗样本**： 攻击者可能使用反光材料或特殊图案来欺骗计算机视觉系统，误导其对环境的认知。
    - **光学攻击**： 使用强光或激光来干扰摄像头或激光雷达，影响感知设备的性能。

6.3.4 攻击案例

攻击案例1：声波攻击智能小车的加速度计

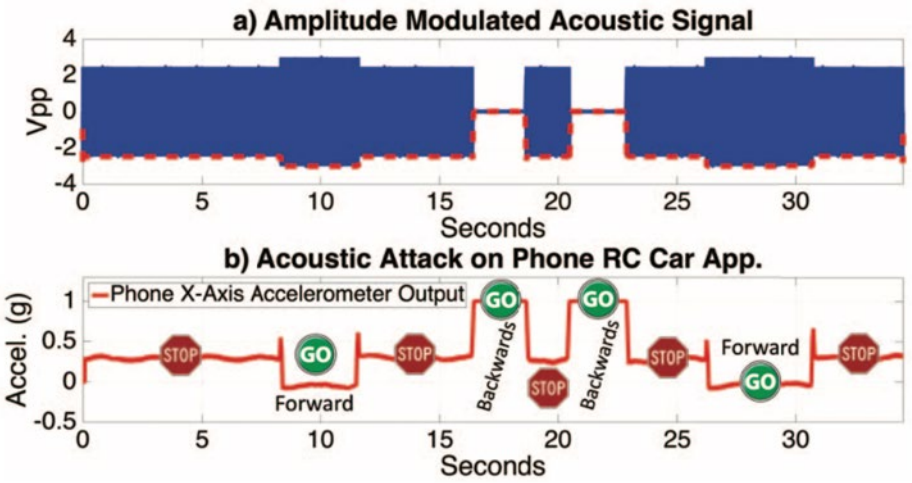
- **攻击原理**：超声波引发运动传感器内部结构**共振**，产生错误或特定读数
- **攻击对象**：IMU中的加速度计、陀螺仪





6.3.4 攻击案例

攻击DEMO：基于传感器的机器人运动控制

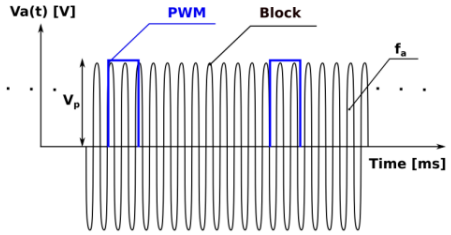


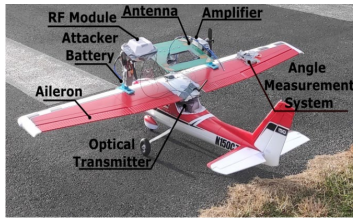
在正常情况下，用户可以倾斜手机至不同的角度来控制小车运动方向。通过声波攻击，小车可以在无需移动手机的情况下运动。

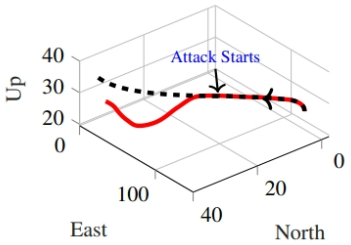
6.3.4 攻击案例

攻击案例2：电磁信号攻击执行器控制信号

场景：无人机的控制信号PWM遭恶意篡改





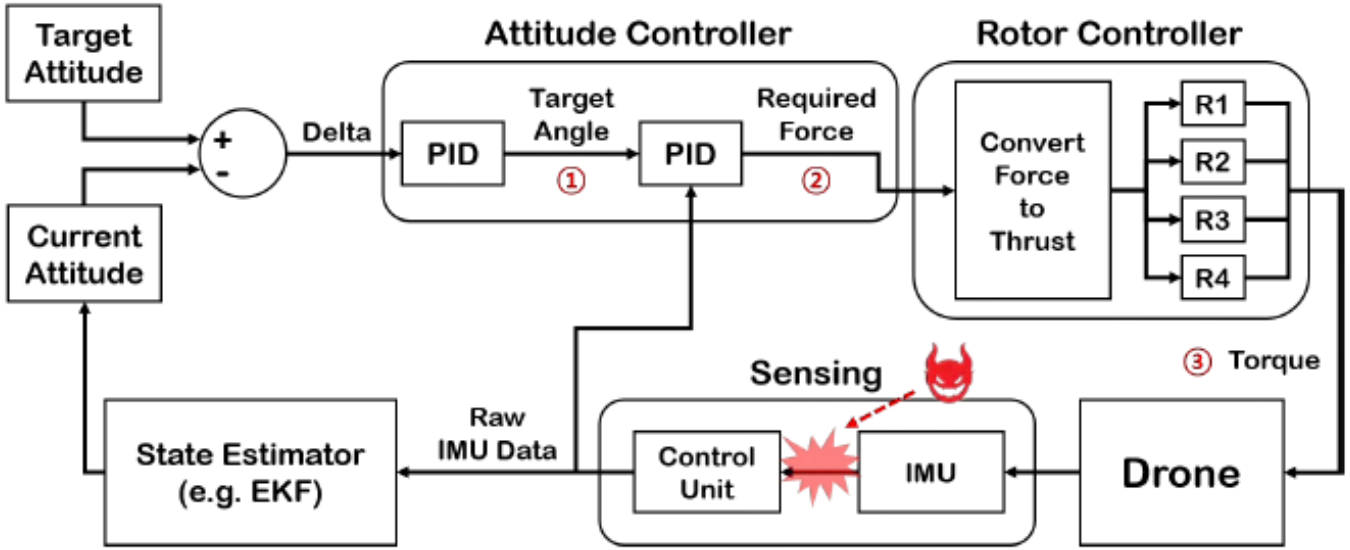


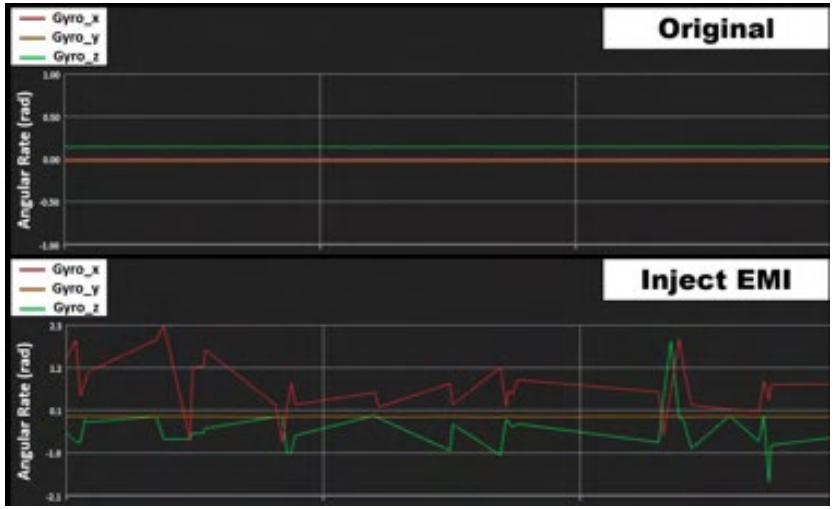
通过电磁信号影响控制器输出的PWM信号，导致无人机坠毁。

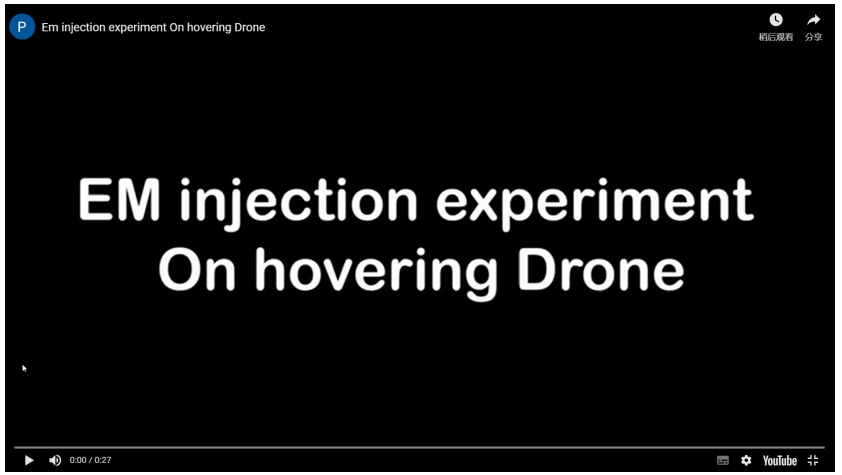
6.3.4 攻击案例

攻击案例3：EMI攻击无人机感知环节输出

通过**电磁干扰(EMI)信号**注入，有效地扭曲了IMU和无人机控制单元之间的通信信道。引起UAV控制出错，系统宕机。







6.4 定义与组成、安全问题、攻击案例

算法安全

6.4 算法安全

1. 定义和组成

2. 安全问题

3. 攻击案例

6.4.1 定义和组成

- **对于算法安全最广泛的定义是：**

　　算法安全在智能无人系统中是指确保系统中使用的算法在设计、实施和运行过程中，不受到潜在攻击或滥用的威胁，并能够保护系统的数据、决策和功能免受未经授权的访问、篡改或破坏。这包括对算法本身和其在整个系统中的集成进行全面的安全考虑。

- **主要目标**是保护系统的算法**免受攻击**和满足**隐私保护**需求。

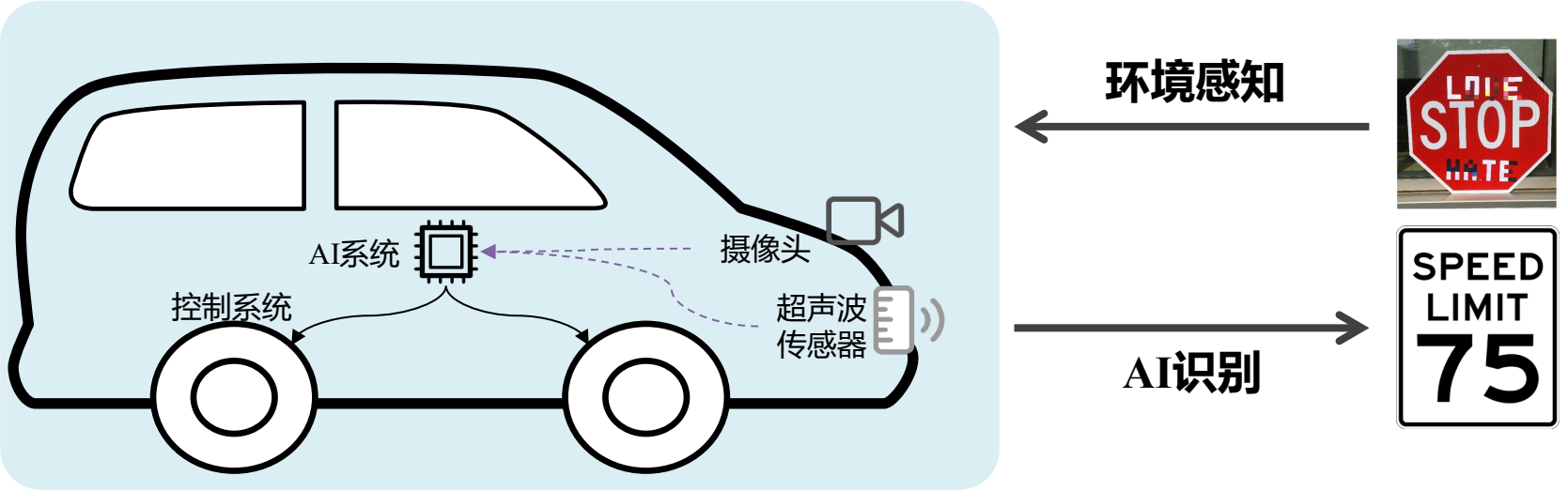
- **智能算法的安全和隐私威胁：**
  - **对抗攻击**
  - **后门攻击**
  - **成员推理攻击**
  - **模型反演攻击**
  - **模型提取攻击**


MIT News
ON CAMPUS AND AROUND THE WORLD
SUBSCRIBE
Machine learning speeds up vehicle routing
Strategy accelerates the best algorithmic solvers for large sets of cities.
Becky Ham | Department of Civil and Environmental Engineering
December 10, 2021

6.6.2 安全问题

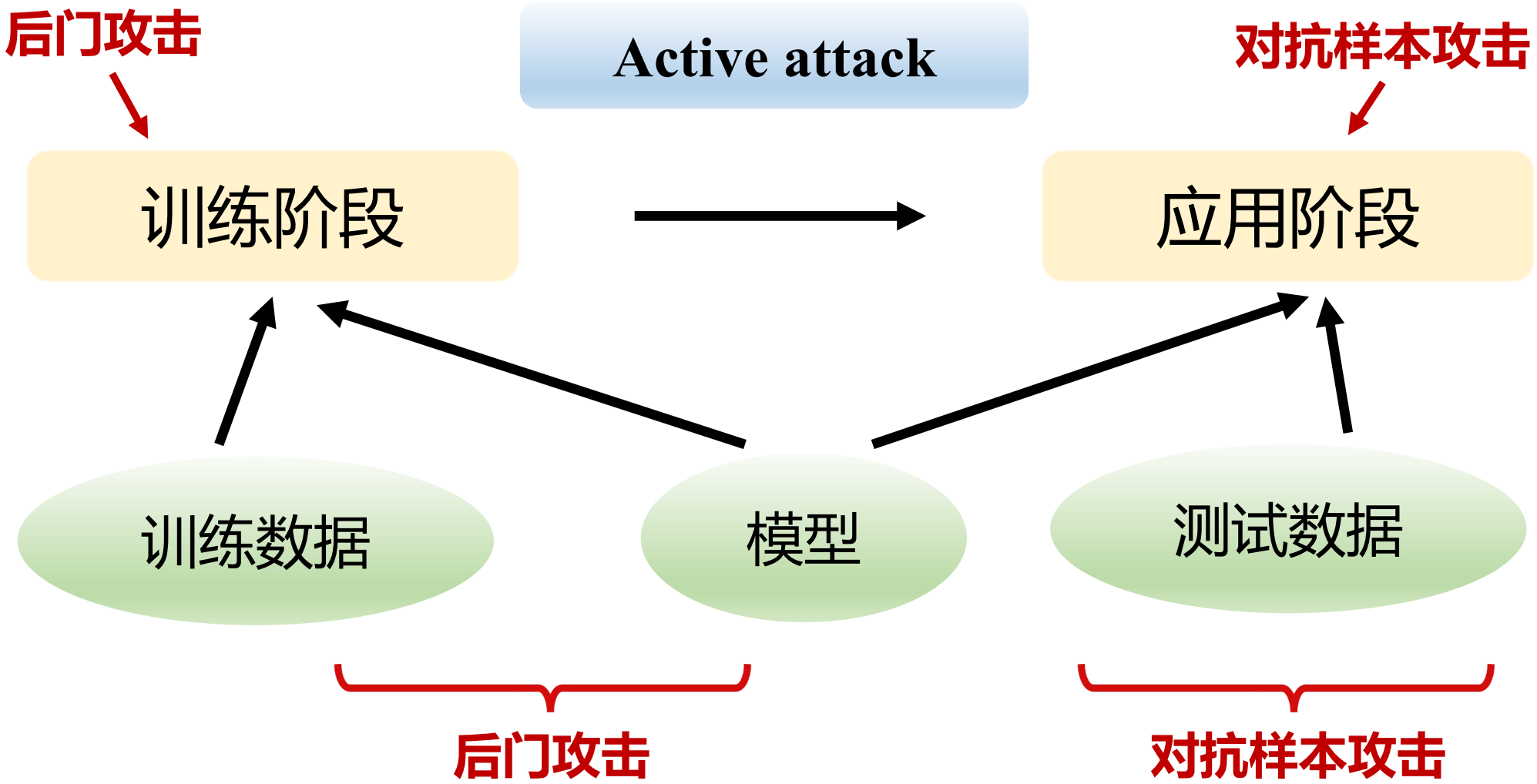
智能无人系统面临的新问题：由感到控

- **智能系统 = AI算法 + 传感器 + 执行器**
- **传感器脆弱性**导致感知结果受恶意攻击信号影响
- **AI算法脆弱性**放大感知错误，影响系统决策，造成严重后果



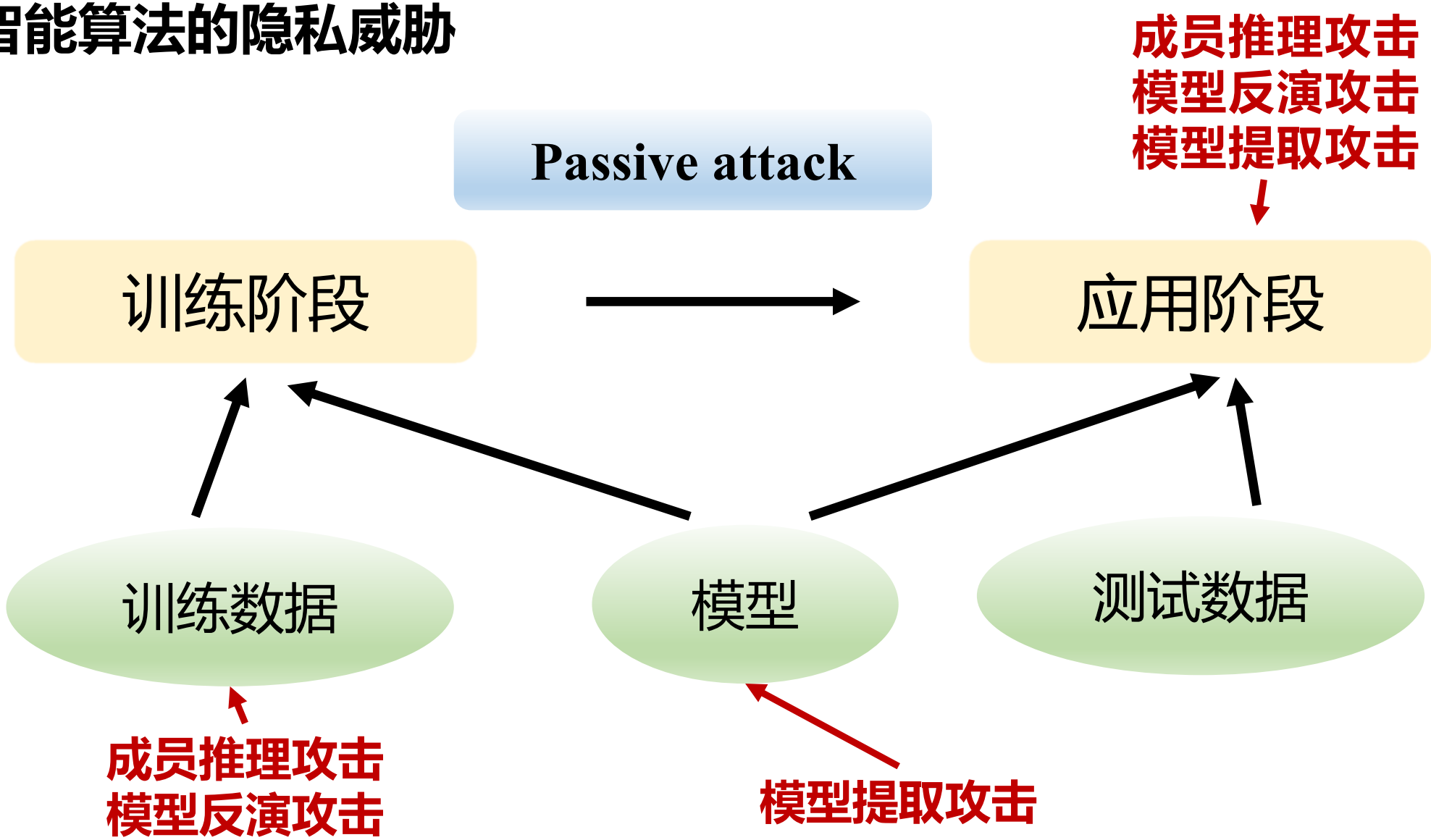
6.4.2 安全问题

智能算法的安全威胁



6.4.2 安全问题

智能算法的隐私威胁



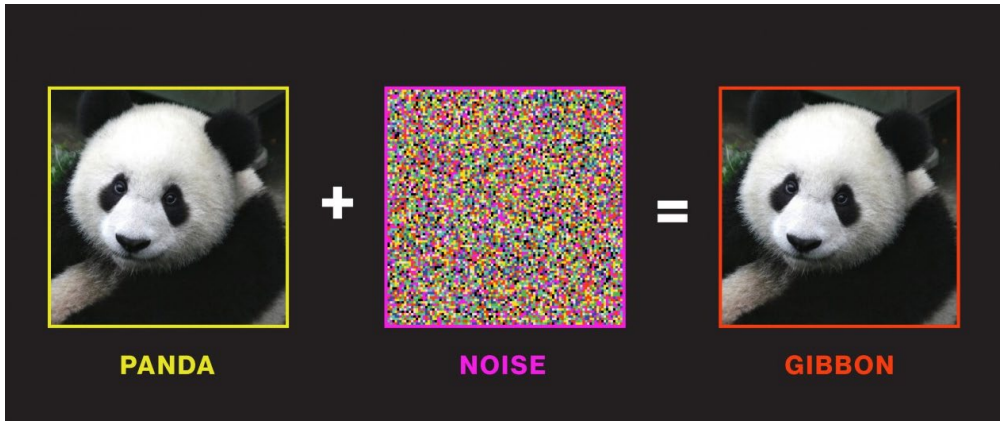
6.4.2 安全问题

安全威胁：对抗样本

对抗样本：通过添加特定扰动使得AI算法错误输出的恶意样本。

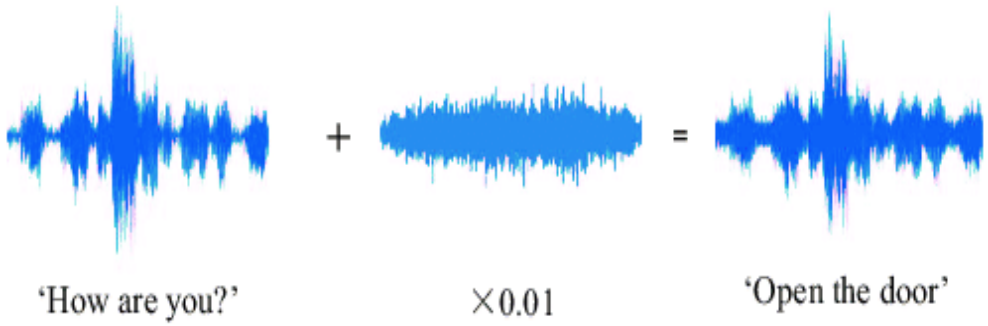
典型攻击示例：

图像：将熊猫图片误分类为长臂猿



图像对抗样本

语音：将How are you误识别为Open the door

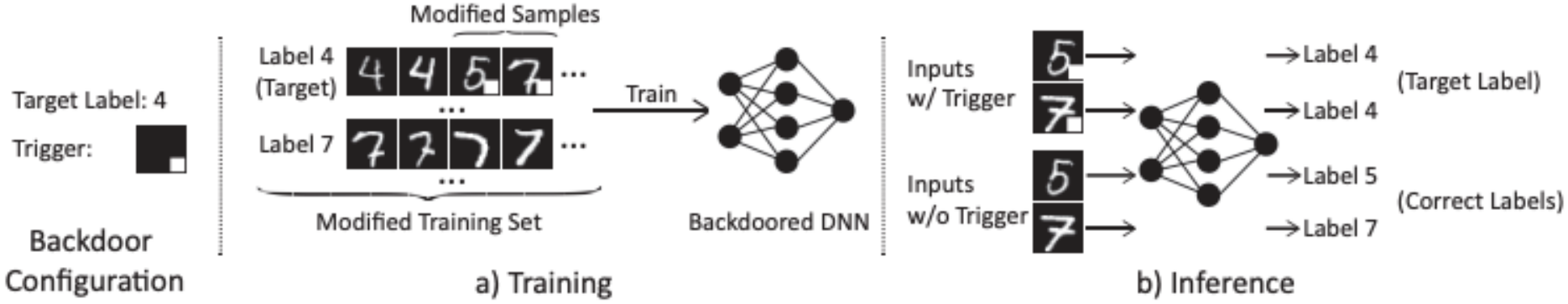


语音对抗样本

6.4.2 安全问题

安全威胁：后门攻击

后门攻击：后门攻击将后门植入到AI模型中，这样后门模型就可以学习攻击者选择的子任务和良性的主任务。



后门模型：

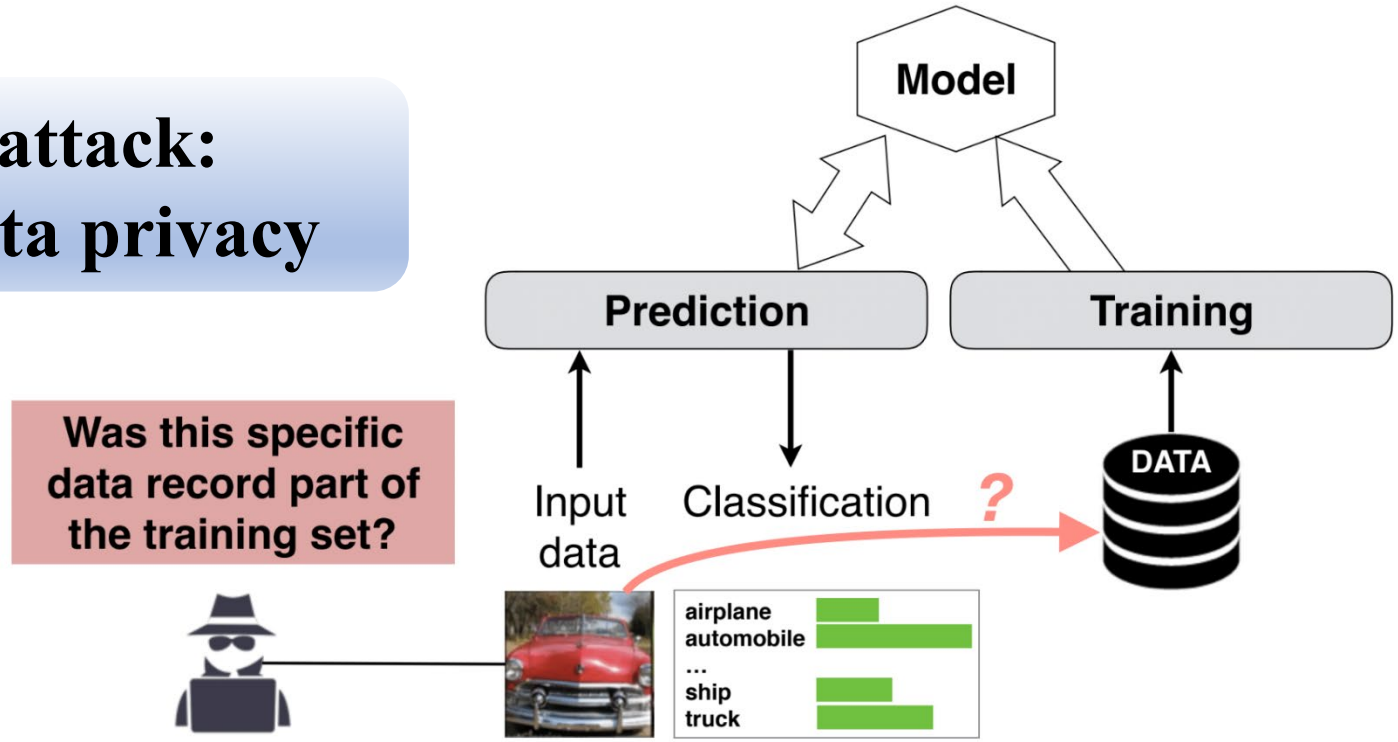
- 平时作为干净输入的干净模型正常运行；通过特制的触发器触发，执行恶意子任务。
- Untargeted attack：所有带有触发器的输入都被错误地分类为任何类；
- Targeted attack：所有带有触发器的输入都被错误地分类为特定的类。

6.4.2 安全问题

隐私安全：成员推理攻击

成员推理攻击(MIA)：给定一个模型和一个样本，确定样本是否在训练数据集中。

Passive attack:
training data privacy

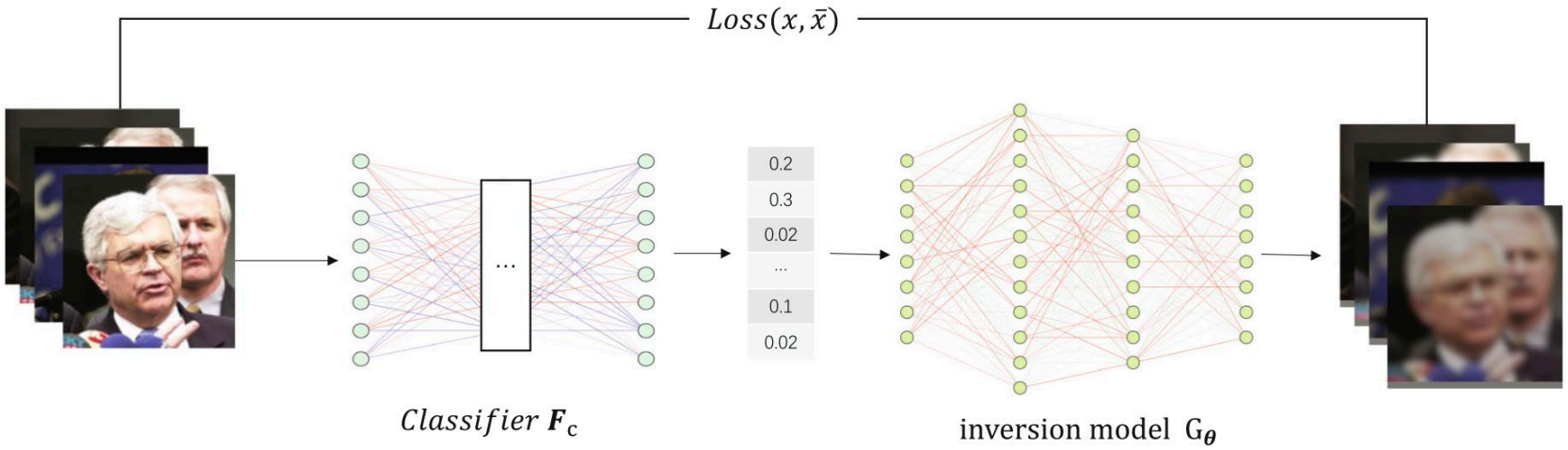


6.4.2 安全问题

隐私安全：模型反演攻击

模型反演攻击的目的是重建训练数据样本。

Passive attack:
training data privacy

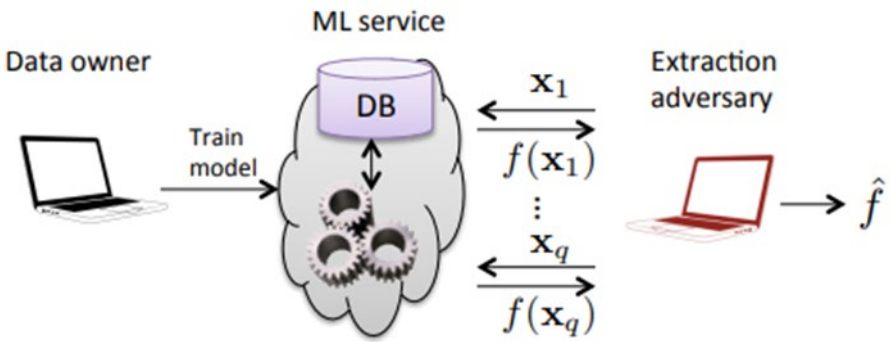


6.4.2 安全问题

隐私安全：模型提取攻击

模型提取攻击的目的是学习目标模型的私有信息，或者建立一个功能与目标模型相似的替代模型。

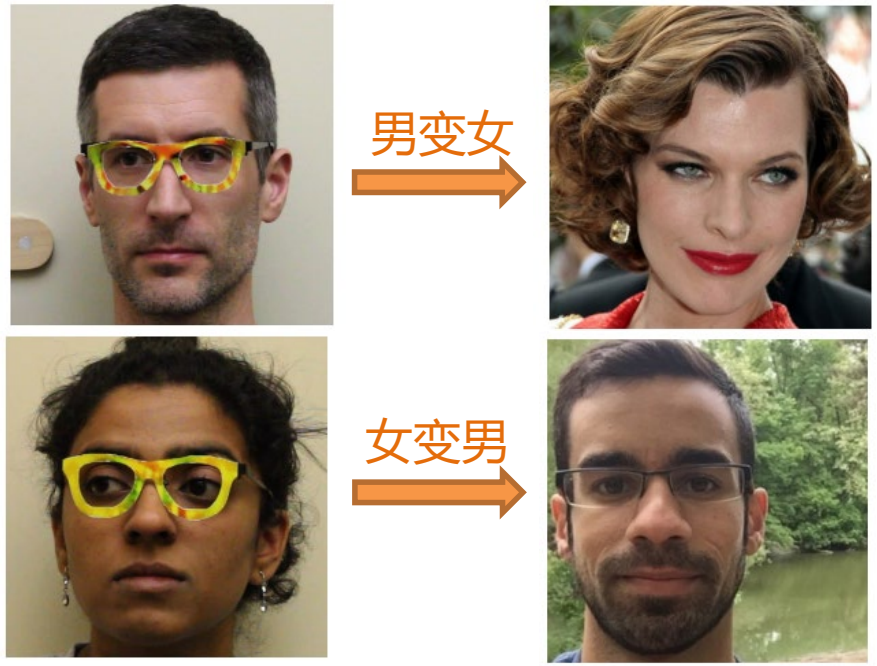
Passive attack:
training data privacy



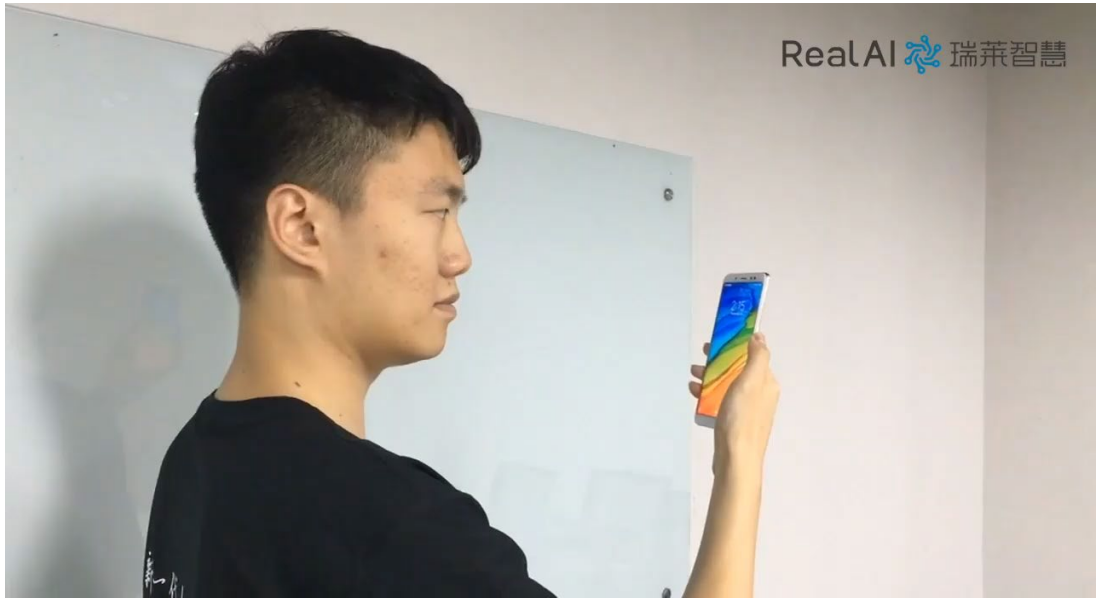
- 参数提取
- 超参数提取
- 模型架构提取
- 模型功能提取

6.4.3 算法安全的案例

图像对抗样本举例：人脸识别



通过佩戴**具有对抗噪声的3D打印眼镜**，可欺骗人脸识别系统，包括身份和性别。

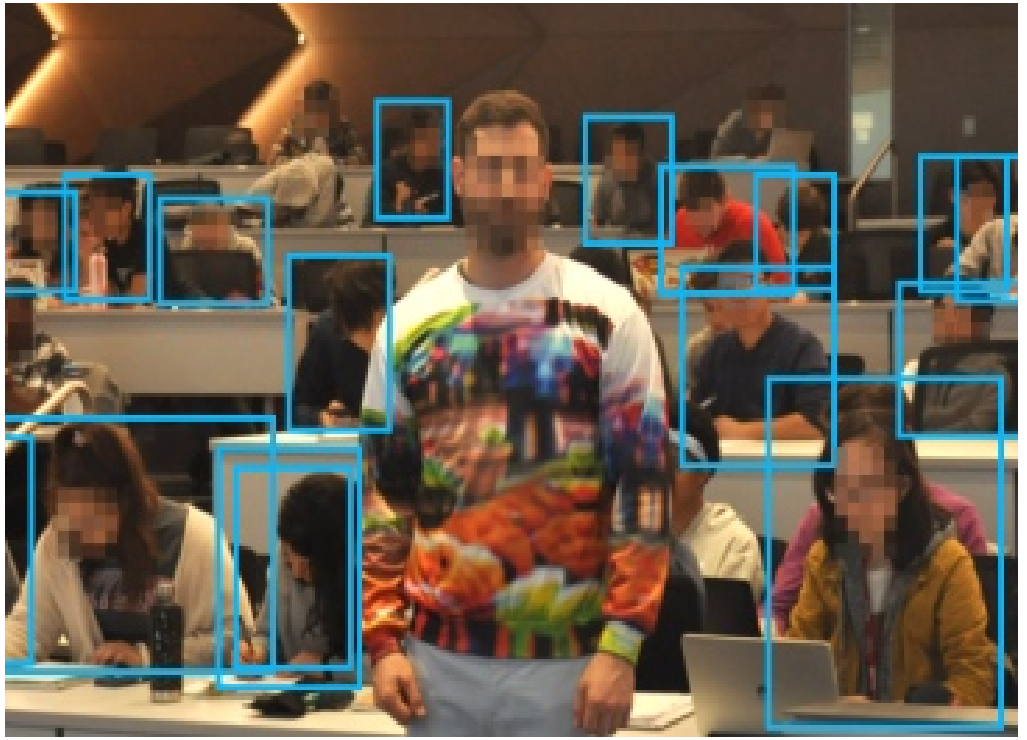


通过佩戴**贴有机主眼部特征的眼镜**，可以破解多款手机的人脸识别解锁系统。

6.4.3 算法安全的案例

图像对抗样本举例：目标检测

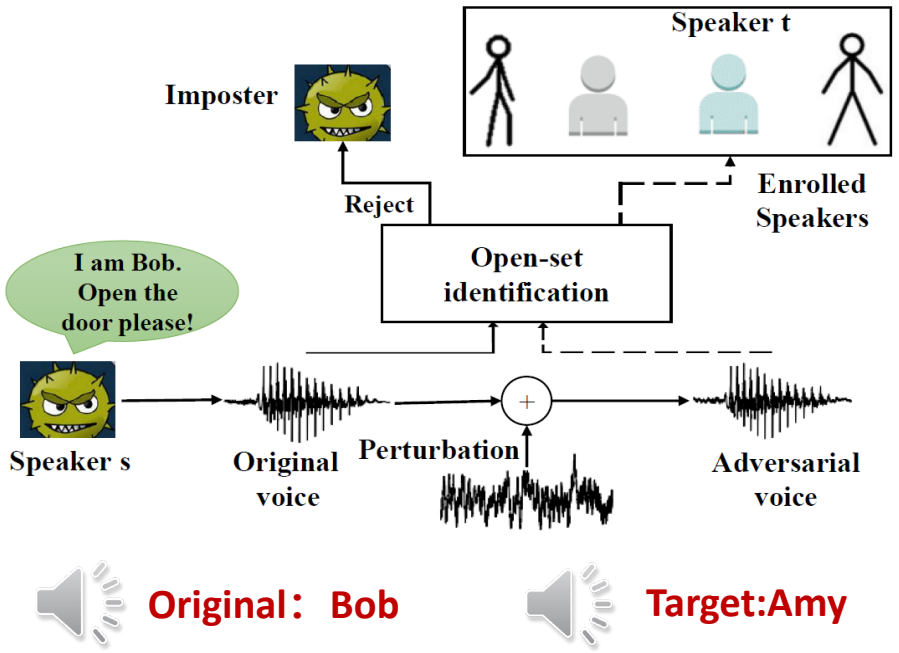
使用**带有对抗样本的纸板**遮挡部分身体，可以在目标检测系统眼中“消失”。



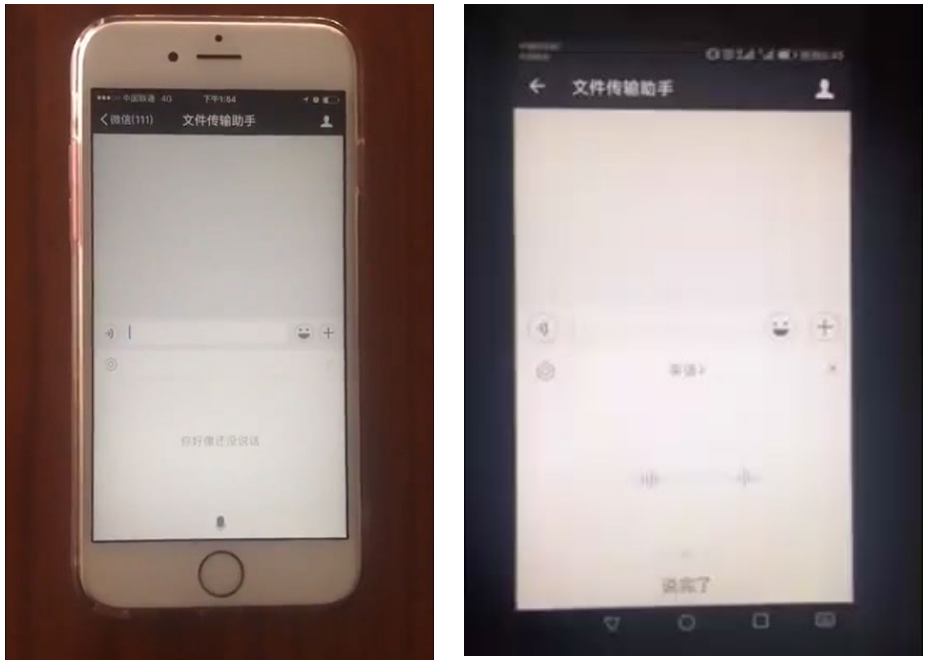
通过穿着**印有对抗样本的衣服**，可以在检测人类的目标检测系统前“隐身”。

6.4.3 算法安全的案例

语音对抗样本举例



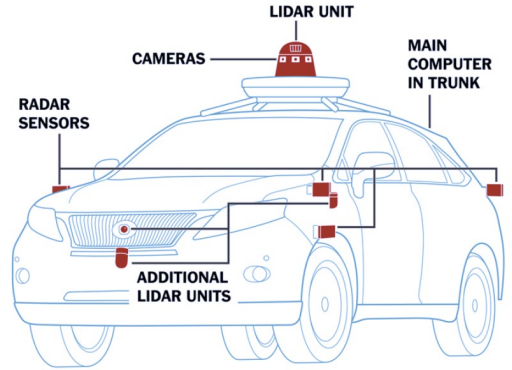
在Bob的原始音频中加入**对抗噪声**，可欺骗**声纹识别系统**——即人耳识别为Bob，AI算法识别为Amy。



两段被**添加对抗噪声的歌曲**分别被**语音识别系统**识别为“make credit card”（左）和“open the door”（右）。

6.4.3 算法安全的案例

以无人驾驶系统安全为例

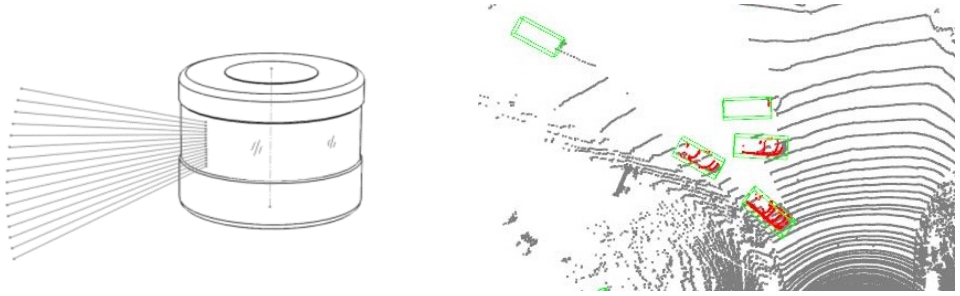


- **研究背景：** 自动驾驶汽车使用各类传感器感知环境
- **攻击对象：** **激光雷达、摄像头、**超声波雷达

激光雷达

组成：多对激光收发模组、旋转部件

用途：障碍物测距、3D物体检测



摄像头

组成：镜头、CMOS感光元件

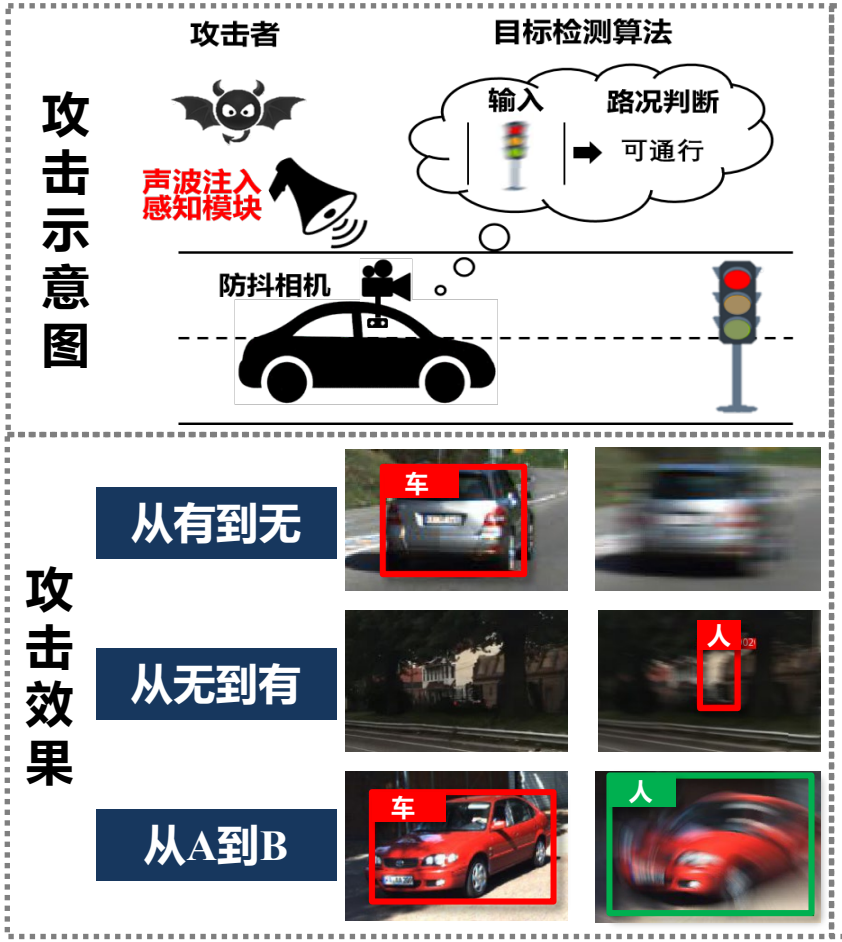
用途：车道线检测、障碍物探测、交通标志识别、红绿灯识别

6.4.3 算法安全的案例

案例1：基于声波注入的摄像头对抗攻击

- **攻击原理：**将声波注入感知模块，相机防抖模块产生反向补偿，形成模糊图像对抗样本。

- **攻击效果：**可以实现消失、出现和替换三种攻击，产生安全隐患。

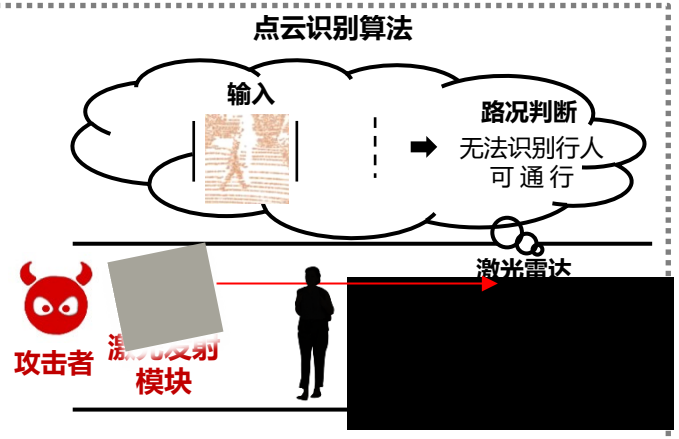


6.4.3 算法安全的案例

6.4.3 算法安全的案例

案例2：激光雷达对抗攻击

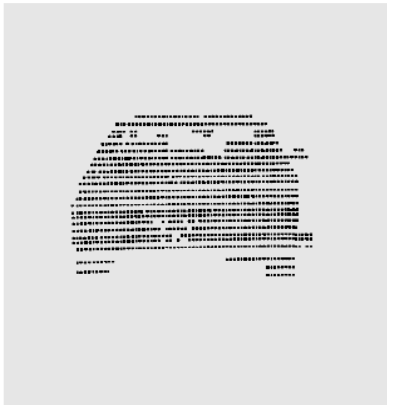
攻击原理



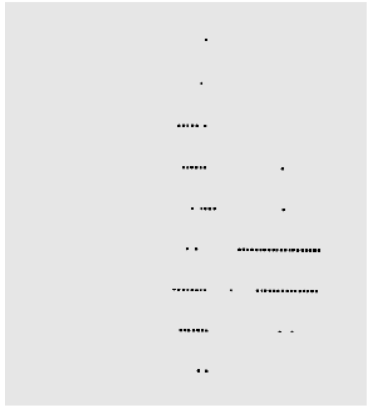
攻击原理：攻击者可通过在合适的时间向激光雷达发射特制的激光信号，使得激光雷达的探测结果中生成对抗样本。

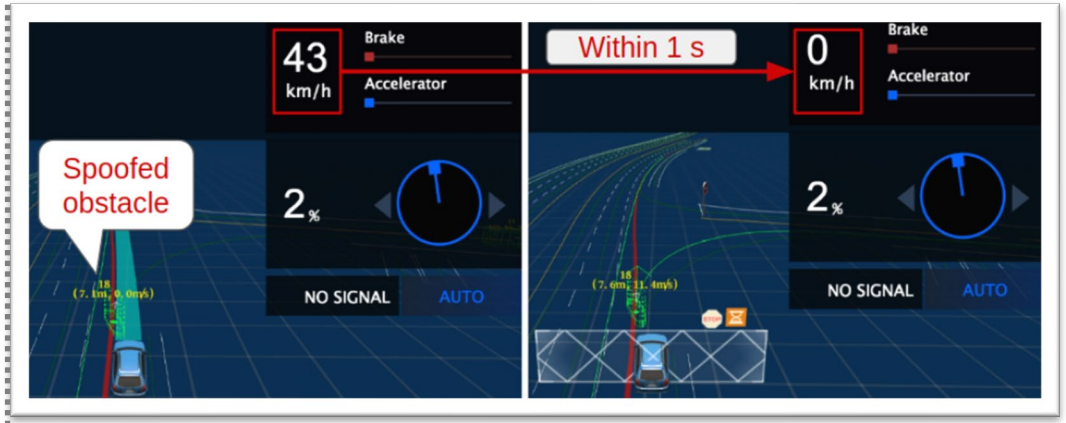
攻击效果：使被攻击车辆紧急刹车或冻结在原地。

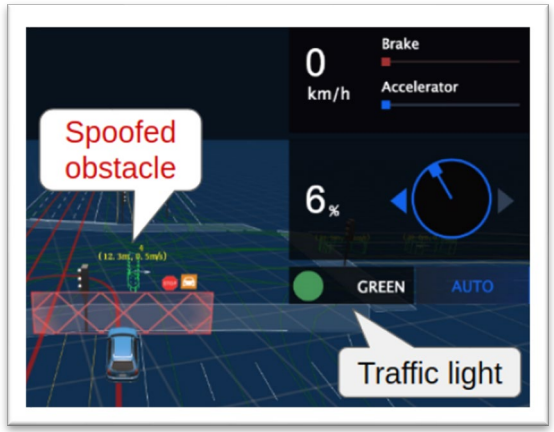
正常点云车辆图像



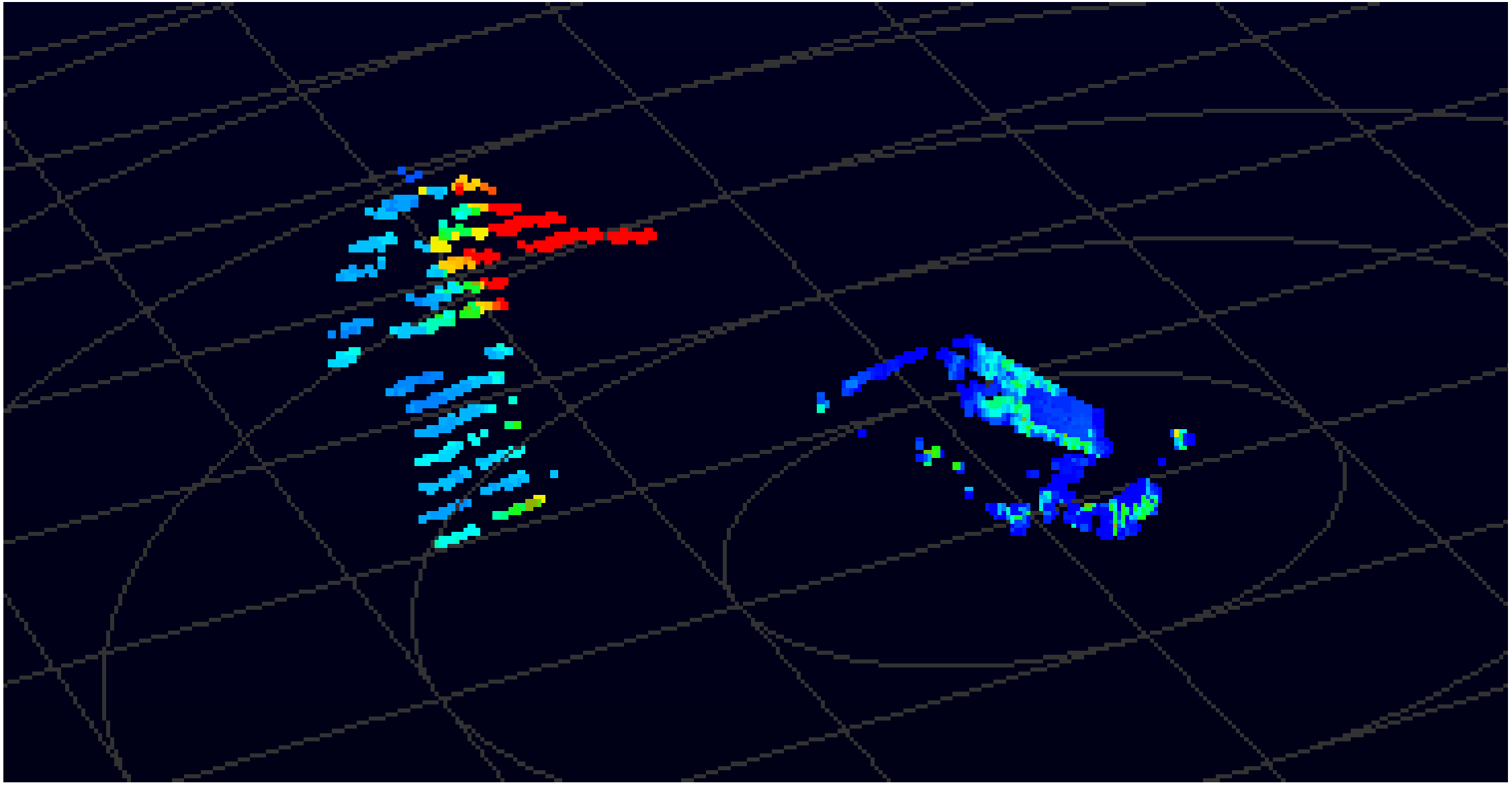
恶意点云车辆图像







6.4.3 算法安全的案例



6.4.3 算法安全的案例

自动驾驶中的后门攻击举例

Dirty-label attacks

Normal run

Trojaned run

Liu, Yingqi, Shiqing Ma, Yousra Aafer, Wen-Chuan Lee, Juan Zhai, Weihang Wang, and Xiangyu Zhang. "Trojaning attack on neural networks." *NDSS*, 2017.

6.5 定义与组成、安全问题、攻击案例

通信安全

6.5 通信安全

1. 定义和组成

2. 安全问题

3. 攻击案例

6.5.1 定义和组成

- **对于通信安全最广泛的定义是：**

智能无人系统通信安全是指在无人系统中确保通信过程中的数据保密性、完整性和可用性，以及防范潜在的网络攻击和威胁的一系列安全措施。通信安全涵盖了无人系统与其他设备、中央控制中心、云服务等进行的数据传输和信息交换，以保护系统的运行、任务执行和用户隐私。

- **通信安全的组成：**
  - **数据保密性、完整性、可用性**
  - **身份认证**
  - **访问机制**
  - **安全的远程访问管理**
  - **防御网络攻击**
  - **……**

6.5.2 安全问题

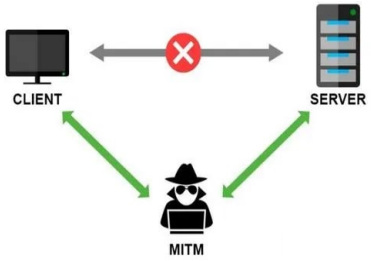
通信过程脆弱性

- **无加密/弱加密通信**： 当通信过程中的数据没有经过加密保护时，攻击者能够轻松地截获、窃听和篡改传输的信息。缺乏加密机制使得数据容易受到未经授权的访问。使用弱加密算法或者采用过时的加密标准可能导致密钥的破解。攻击者可以通过密码破解技术或其他攻击手段来获取加密数据。
- **身份验证薄弱**：攻击者可能伪造通信参与者的身份，冒充合法用户或设备。这种攻击形式可能导致未经授权的访问和信息泄露。使用弱密码、缺乏多因素身份验证或未采用适当的身份验证机制会增加身份被盗用的风险。攻击者可能通过猜测密码、社会工程学等手段获得合法用户的凭证。
- **协议漏洞**： 未及时更新软件、协议或操作系统可能存在已知漏洞，攻击者可以利用这些漏洞进行恶意活动，例如拒绝服务攻击或远程执行代码攻击。
- **访问控制不当**：如果通信系统没有实施严格的访问控制，攻击者可能能够轻松地进入系统并获取敏感信息，或对通信进行干扰。

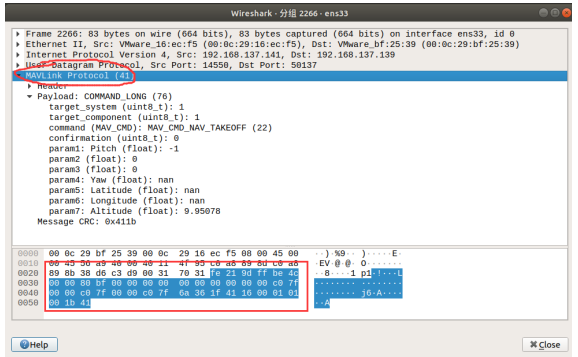
6.5.3 通信安全的案例

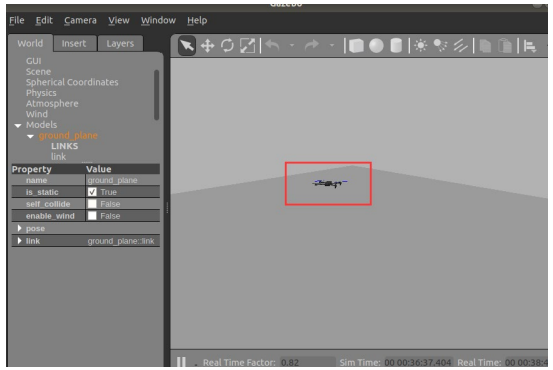
MAVLINK 中间人攻击

- MAVLINK是一种用于飞行器和地面站之间通信的轻量级通信协议。
- 中间人攻击可以通过欺骗协议的数据流，从而实施多种恶意行为。
- 攻击者可以窃听敏感信息，如航迹、传感器数据等，也可以篡改飞行控制命令，导致危险行为。
- 中间人攻击可能采用欺骗手段，伪造MAVLINK消息，导致误导性信息的传输，影响飞行器的正常操作。



案例一：Mavlink中间人攻击





中间人攻击实现未发送指令情况下的异常起飞

6.6 定义与组成、安全问题、攻击案例

通信安全

6.6 控制安全

1. 定义和组成

2. 安全问题

3. 攻击案例

6.6.1 定义和组成

- **对于控制安全最广泛的定义是：**

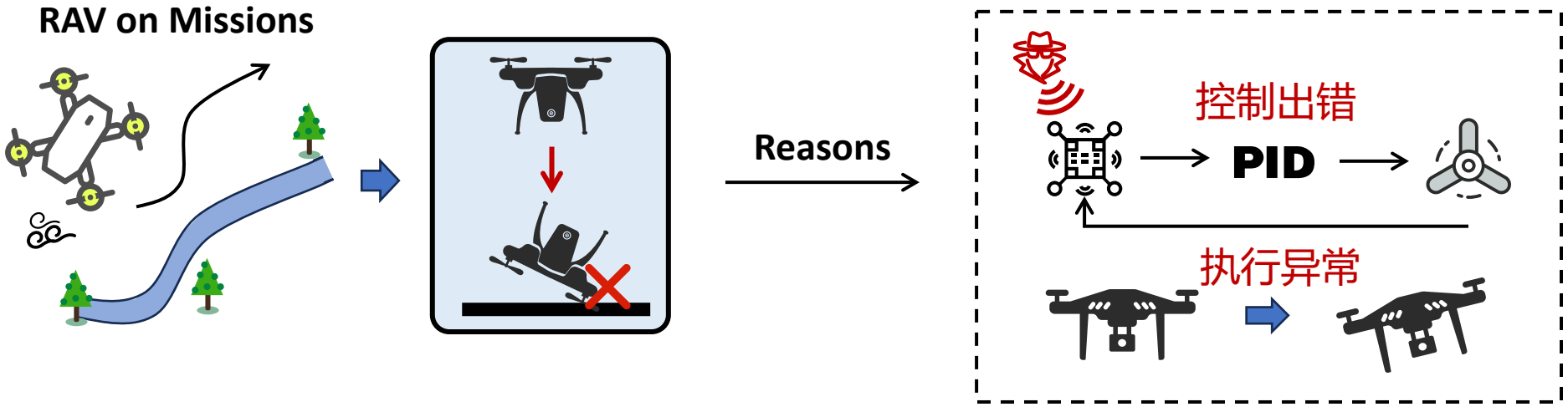
智能无人系统控制安全是指确保无人系统中的控制过程安全可靠的一系列措施。这包括防范潜在的攻击和威胁，以确保系统的控制指令、决策和执行过程不受到未经授权的干扰、篡改或破坏。控制安全的关键目标是保护系统免受恶意行为、误操作或技术故障的影响，确保系统能够稳定、可靠地执行任务。

- **对于控制安全的组成：**
  - 控制系统软硬件安全
  - 网络通信安全
  - 代码固件安全
  - 语义执行过程安全

6.6.2 安全问题

智能无人系统面临的新问题：由感到控

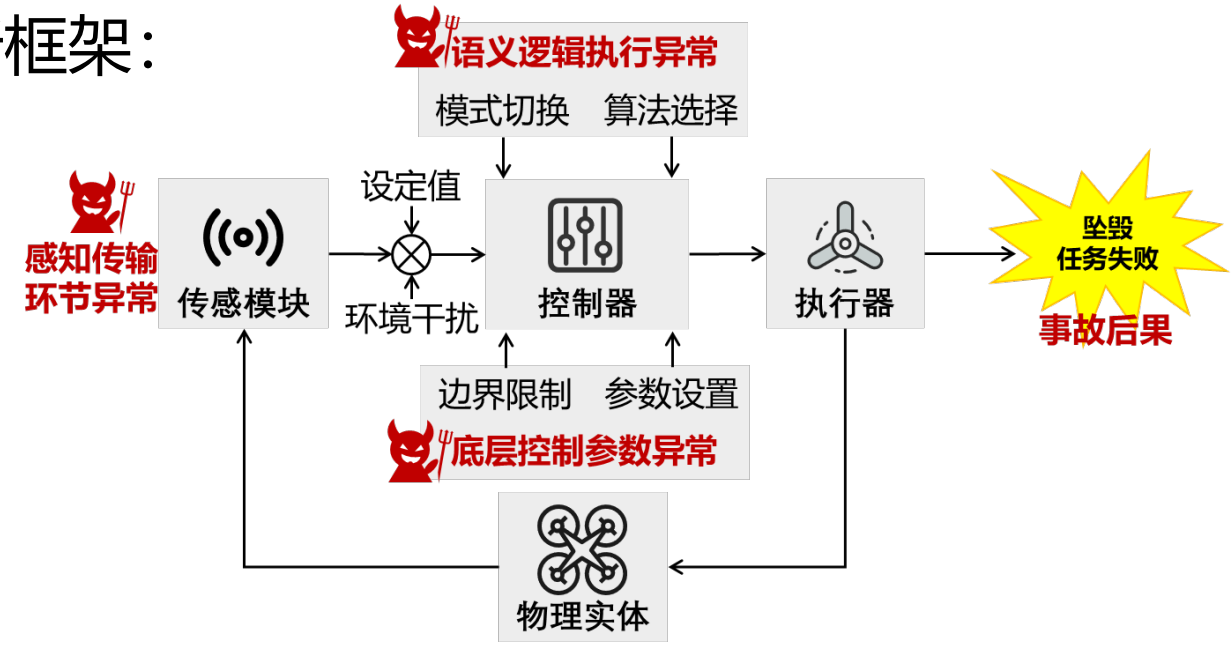
- **无人系统 = 控制算法 + 传感器 + 执行器**
- **传感器脆弱性**导致感知结果受恶意攻击信号影响
- **控制安全的脆弱性**由错误感知结果触发控制的错误，直接影响执行



6.6.2 安全问题

控制安全的整体分析框架

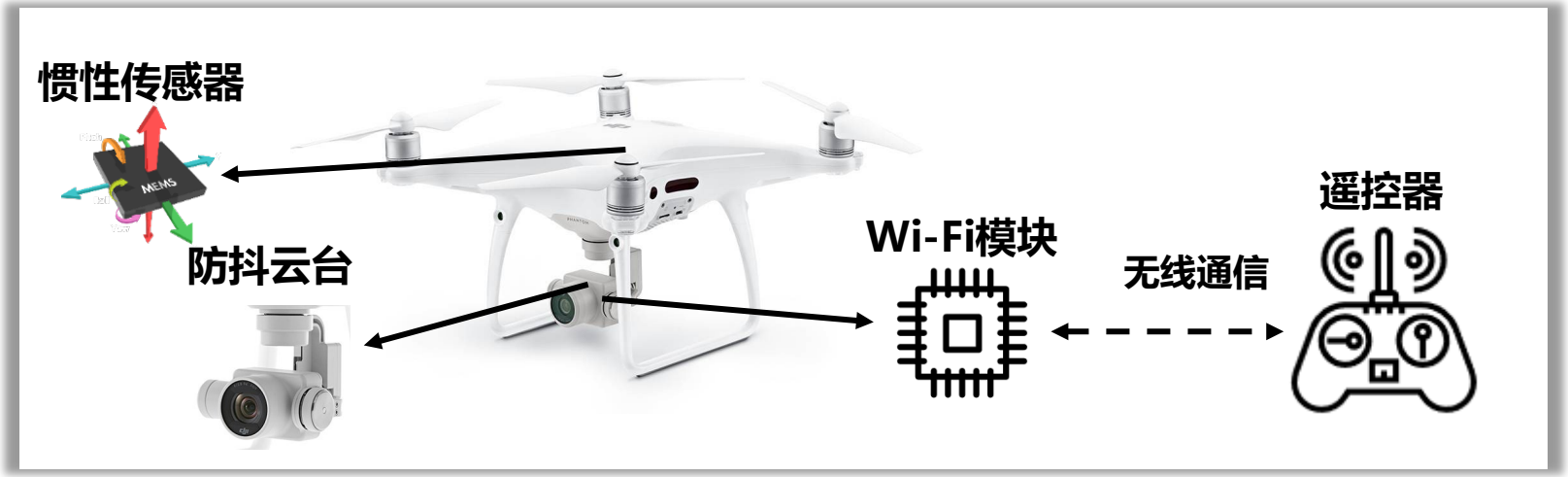
- 智能无人系统的**控制安全**是确保系统在操作和执行任务时能够保持稳定、可靠、安全的关键方面。
- 智能无人系统控制安全的整体分析框架：
  - 感知传递环节异常
  - 底层控制参数异常
  - 语义逻辑执行异常



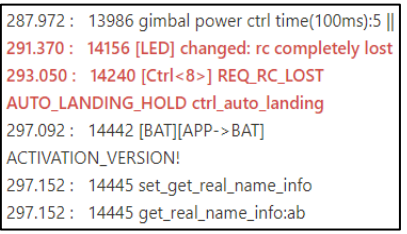
6.6.3 控制安全案例

案例1：超声波攻击无人机MEMS陀螺仪

通过**发射声波信号**干扰无人机摄像头云台防抖系统，使电机剧烈抖动发热，由于系统对温度过于敏感，进行过热重启，**无人机与遥控器断开连接**，触发保护机制并自动降落。



图像传输断开



无人机降落

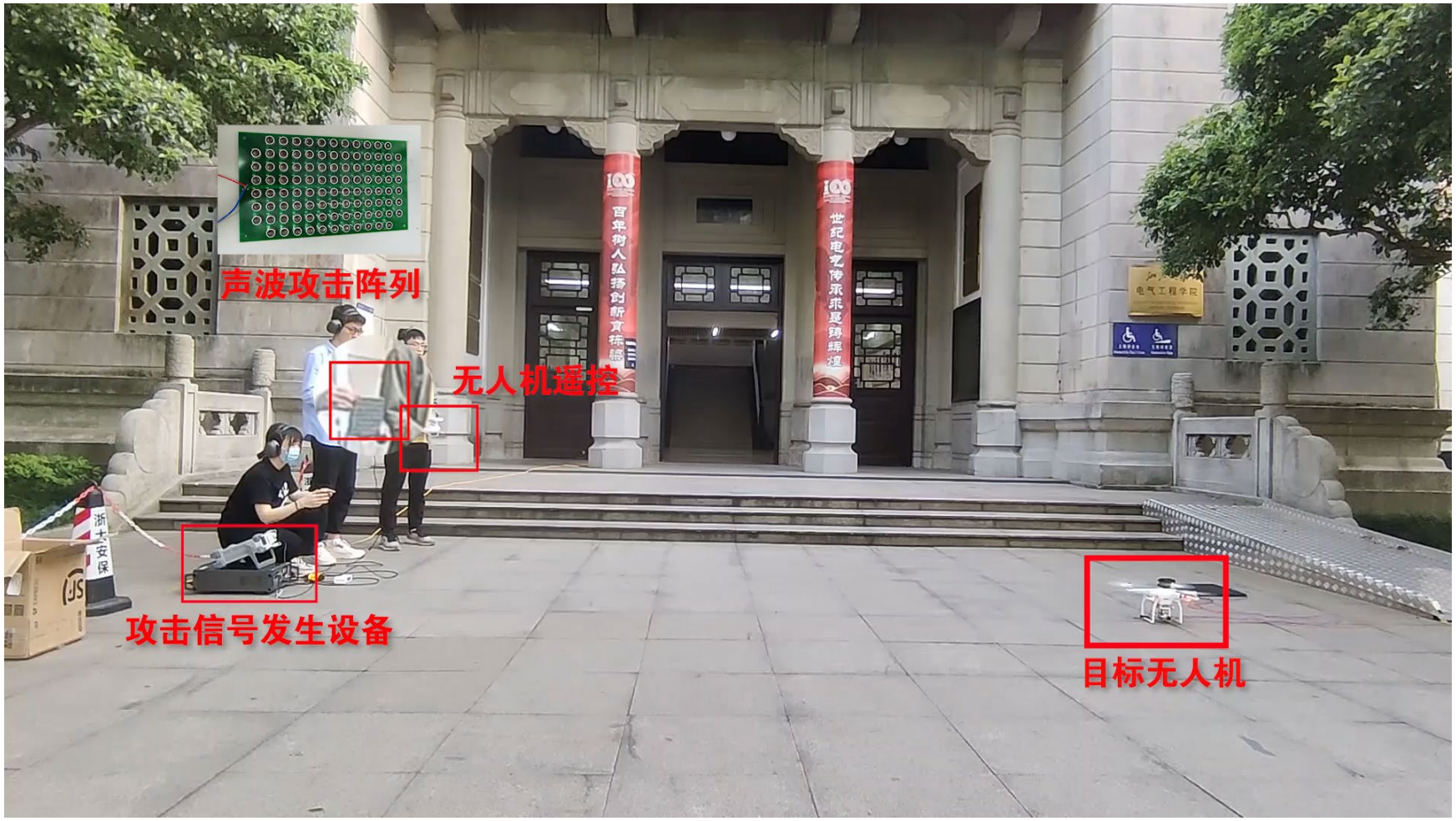
声波攻击信号 → 惯性传感器输出异常 → 云台电机剧烈抖动 → 系统过热重启 → 触发保护机制

6.6.3 控制安全的案例

案例2：超声攻击无人机并躲避故障检测器

- **攻击原理：**超声攻击无人机惯性传感器，利用故障检测器阈值设计漏洞，躲避检测与保护。

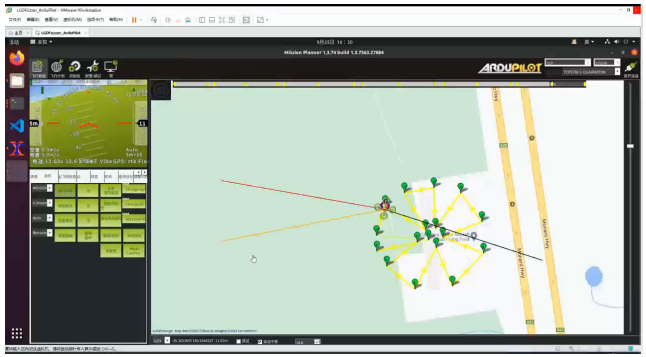
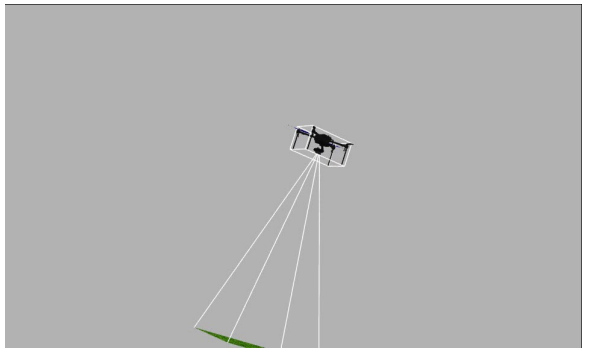
- **攻击效果：**躲避故障检测器检测从而不触发保护机制，导致直接坠机而非迫降。

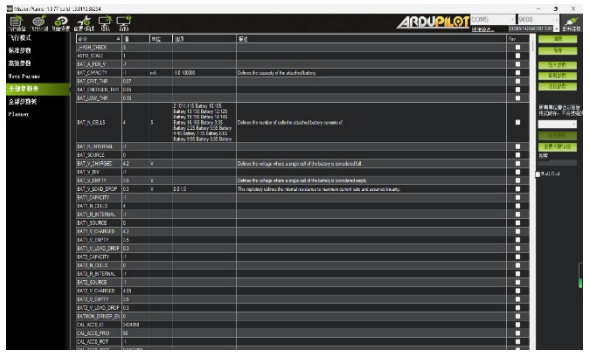


6.6.3 控制安全的案例

案例3：不合理的参数配置导致无人机姿态异常

- 飞行冻结：无人机在航点处意外冻结，小范围内徘徊
- 飞行偏离：无人机飞行轨迹与预期轨迹严重偏差



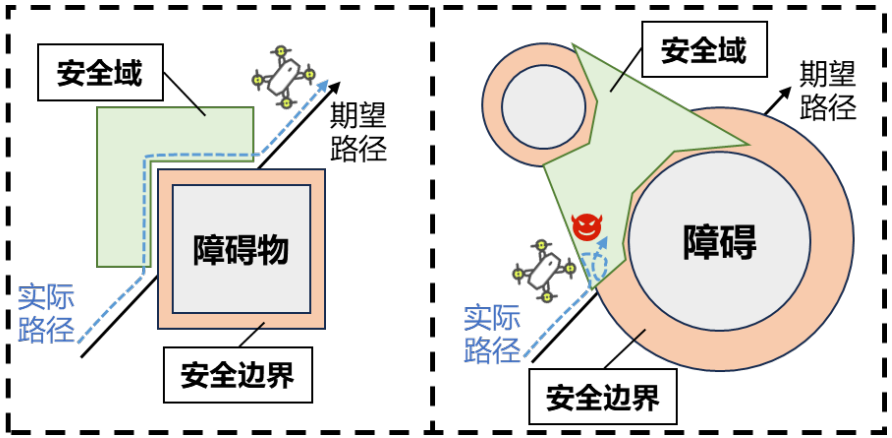
不合理的参数配置导致坠机的情况复现

6.6.3 控制安全的案例

案例4：无害人为障碍影响无人机任务执行

- 由于人为障碍物的影响，使得安全区间和地理围栏设置重叠，导致Dijkstra算法与“简单避障”算法间逻辑出现冲突，导致车辆在靠近地理围栏的区域上来回往复运动，无法完成其任务。

Demo: A faulty patch discovered by PatchVerif in ArduPilot's object avoidance with Dijkstra's algorithm

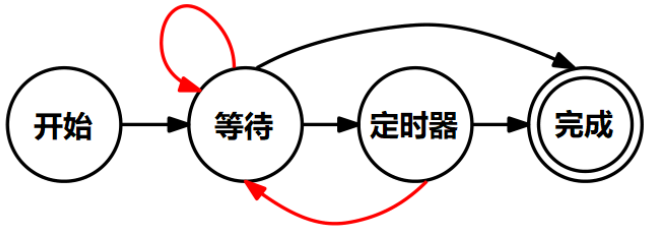


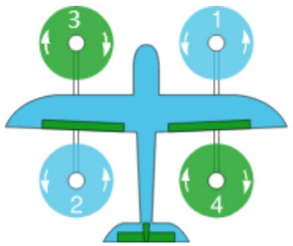
成因原理分析

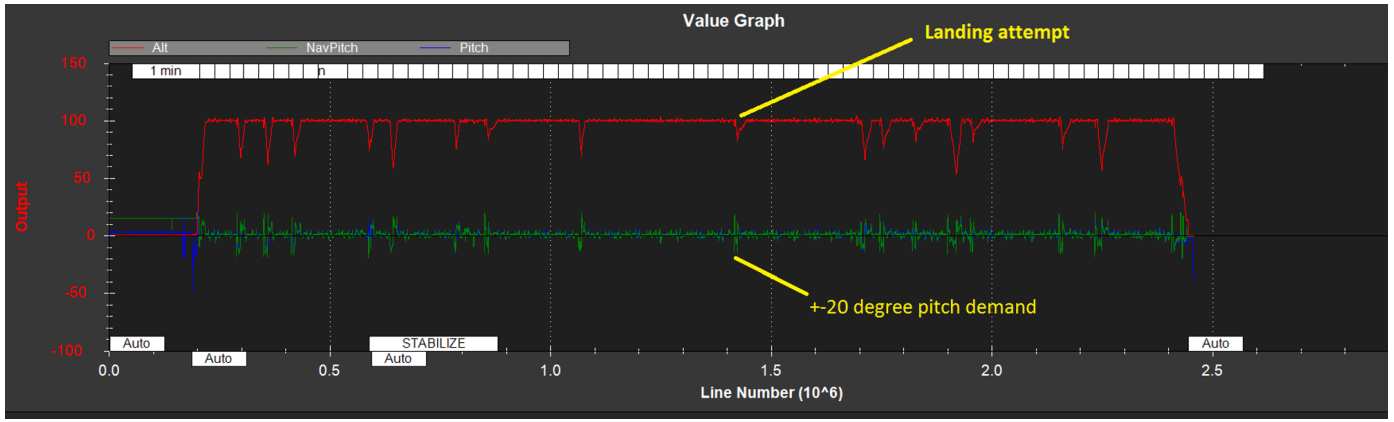
6.6.3 控制安全的案例

案例5：Bug7062 控制优先级冲突

- 无人机为到达较低的新航点，电机转速需要下降，所以控制算法减少了油门出力；但与此同时，无人机需要增加前进速度，以便从悬停过渡到前进飞行。这两种控制算法输出的矛盾关系导致油门来回震荡，无人机将会被卡死在等待状态，而无法继续完成任务。







6.7 从攻击链视角出发的安全防护

智能无人系统的安全防护

6.7 智能无人系统的安全防护

6.6 智能无人系统的安全防护

1. 脆弱性挖掘

2. 攻击检测

3. 实时防御

4. 溯源定位

6.7.1 脆弱性挖掘

智能无人系统脆弱性挖掘技术：

脆弱性挖掘是指对智能无人系统中可能存在的安全漏洞、弱点和潜在风险进行主动探测和分析的过程。该过程旨在发现系统中的潜在脆弱性，以便及时采取修复和改进措施，提高系统的安全性和稳定性。脆弱性挖掘是安全评估的一部分，通过模拟攻击者的行为来评估系统的安全性。

脆弱性挖掘的定义：

脆弱性挖掘是一种系统化的方法，用于主动识别和分析智能无人系统中可能存在的漏洞、弱点和潜在威胁，以揭示系统中可能被攻击者利用的安全缺陷。通过模拟潜在攻击场景和采用安全工具，挖掘系统的潜在漏洞，为系统管理员和开发者提供有针对性的改进建议，从而提高系统的整体安全性。

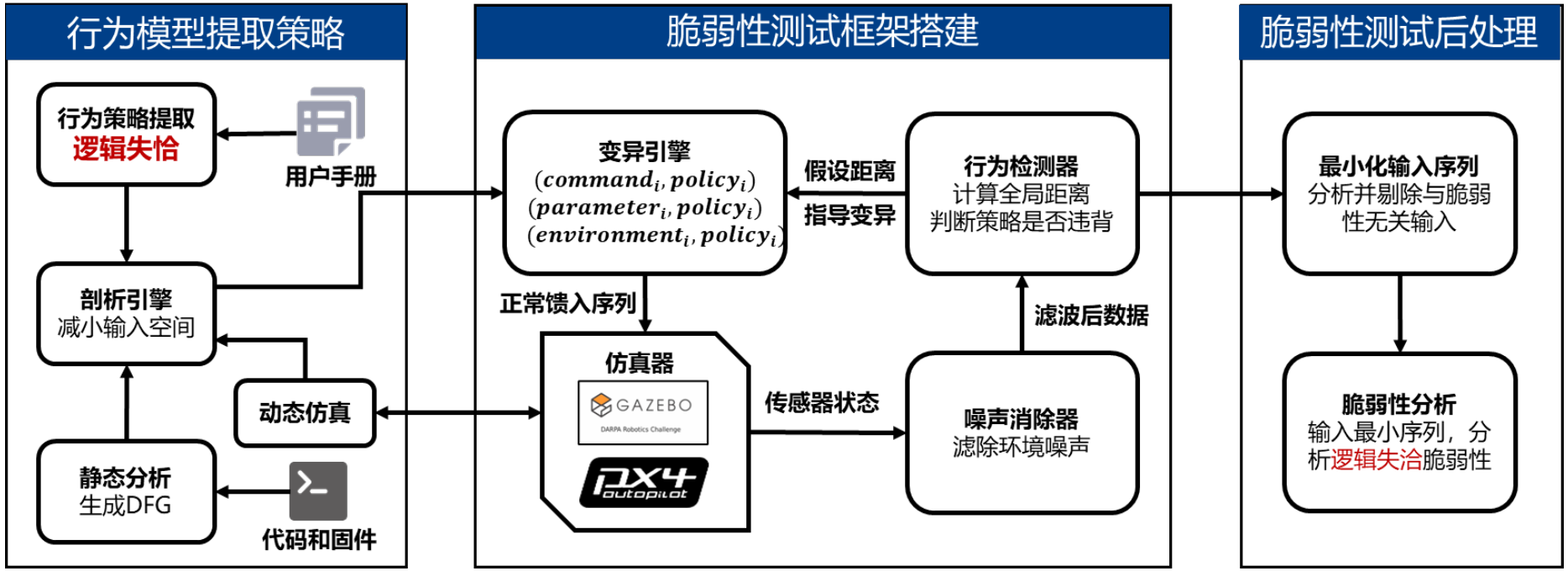
6.7.1 脆弱性挖掘

- **脆弱性挖掘的常用方法：**
  - **静态代码分析**：分析系统源代码或二进制代码，寻找潜在的编程错误、安全漏洞和不安全的编码实践。工具如Checkmarx、Fortify等可用于自动静态代码分析。
  - **安全代码审查**：对系统的源代码进行仔细审查，寻找潜在的漏洞、缺陷和不安全的编码实践。通过手动审查和自动工具辅助，发现潜在问题。
  - **模糊测试**：使用模糊测试技术，向系统输入模糊、异常或非法的数据，以观察系统的响应并检测潜在的崩溃或漏洞。
  - **手动渗透测试**：
    - 黑盒测试： 模拟攻击者的角色，对系统进行手动测试，发现可能的漏洞。测试人员没有关于系统内部结构和代码的详细信息。
    - 白盒测试： 在拥有系统内部结构和代码的情况下，进行深入的手动测试，以更全面地发现潜在的漏洞。

6.7.1 脆弱性挖掘

案例1：基于策略引导的固件模糊测试技术

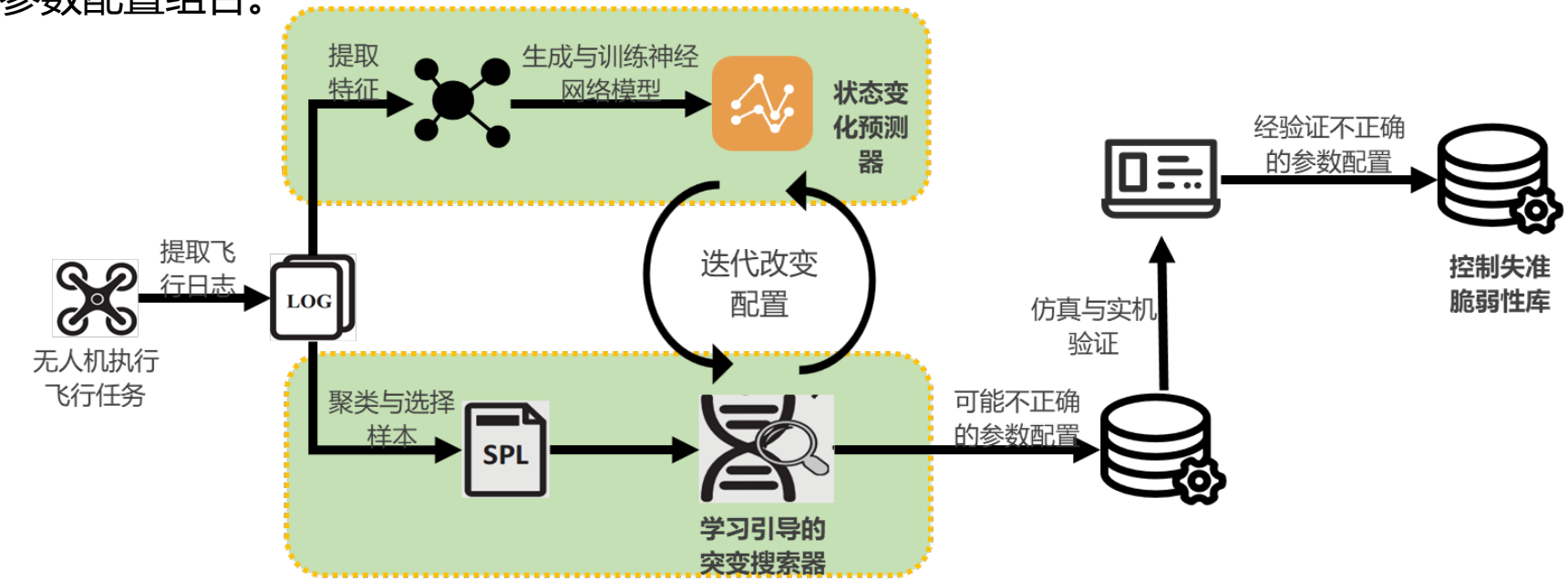
- **原理：**验证无人机是否遵循已识别的安全和功能策略，使用模糊测试哪些指令的操作会导致安全策略的违背，从而分析代码和固件中的语义和逻辑脆弱性。



6.7.1 脆弱性挖掘

案例2：通过学习引导的搜索来检测无人机的参数配置错误

- **漏洞根源**：所有单一参数的取值都在推荐取值范围中，但某些合理的参数组合会影响到无人机的物理稳定性。
- 基于日志训练状态预测器，用于估计无人机在下一个时间戳达到的参考状态。运行基于遗传算法的突变搜索器来迭代生成配置。
- 检测器根据参考状态与预期状态的偏差来推断配置是否正确。推断不正确的配置会通过仿真进行验证，最终检测出一系列不合理的参数配置组合。



6.7.2 攻击检测

- **智能无人系统攻击检测技术：**

攻击检测是指通过监控和分析智能无人系统的行为，及时识别和响应潜在的恶意活动或攻击尝试的过程。这包括使用各种技术手段和工具，以便及时发现异常行为、入侵或其他安全威胁，并采取相应的措施来保护系统免受损害。

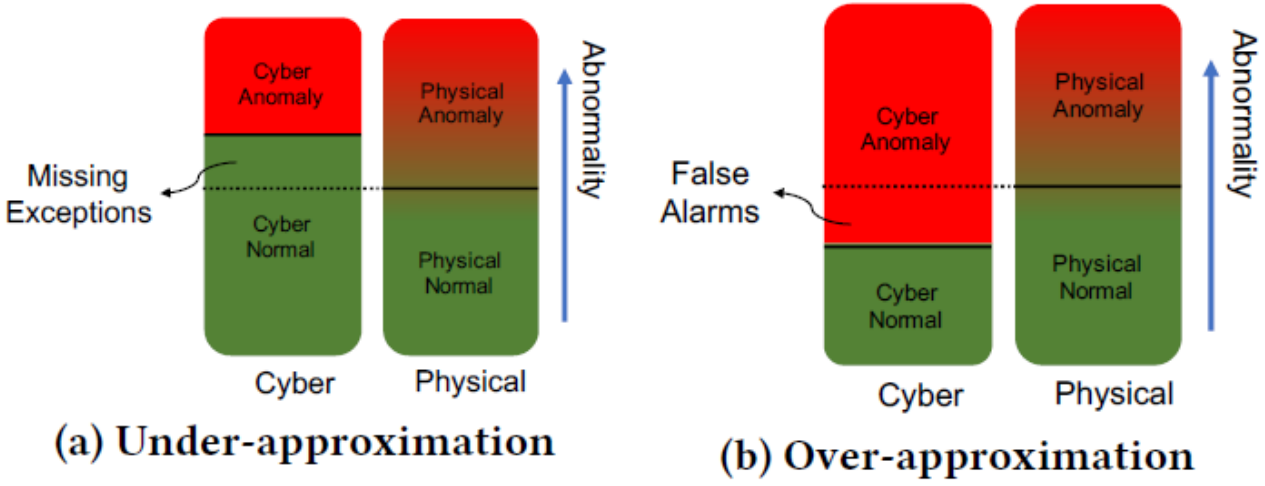
- **脆弱性挖掘的定义：**

攻击检测是一种安全机制，通过实时监控系统的活动、分析网络流量和行为模式，以及使用各种技术手段，识别可能的攻击行为或异常活动。其目标是在攻击者能够对系统造成实际损害之前，及时发现并采取防御措施，以保障系统的安全性。

6.7.2 攻击检测

攻击检测技术常用方法

- **阈值检测法：** 设计系统边界，即定义一个表示正常行为的区域，并将数据中不属于该正常区域的任何观测结果声明为异常，从而确定特征量的阈值。



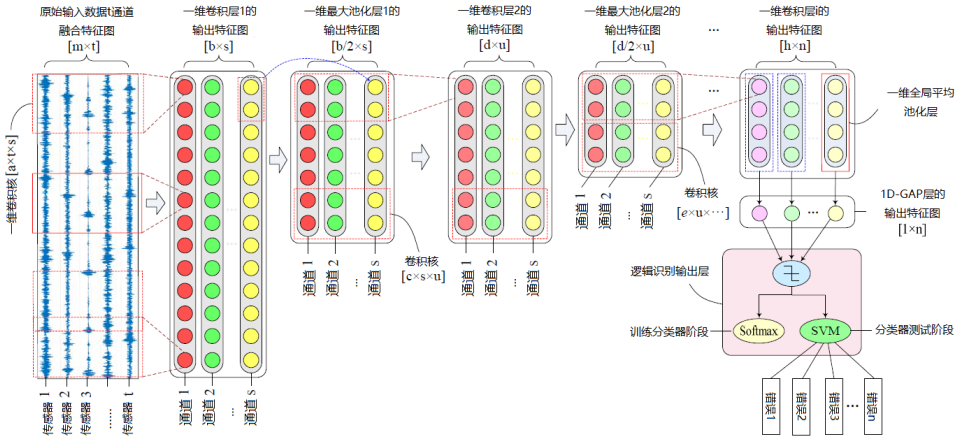
- **存在的问题**

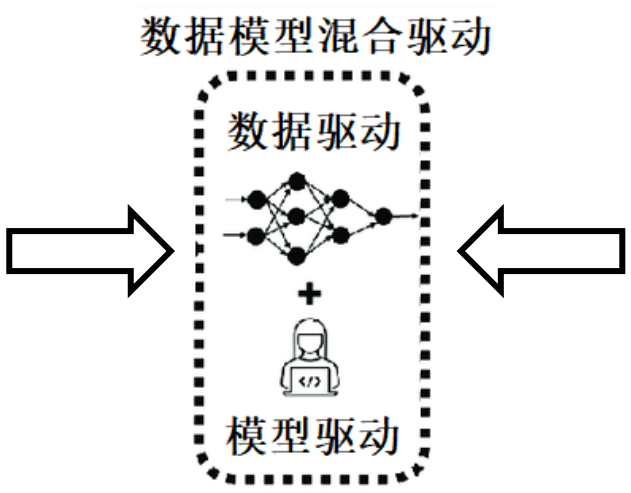
1）**边界难界定** 2）**异常会伪装** 3）**场景影响大** 4）**迁移难兼容** 5）**数据难获取**

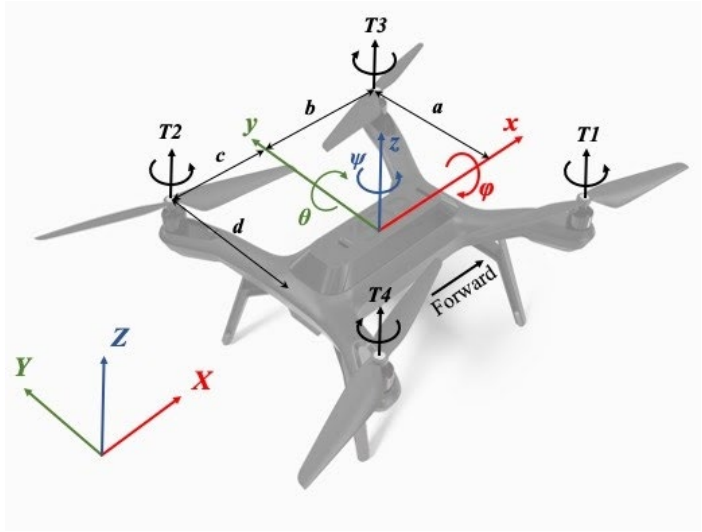
6.7.2 攻击检测

攻击检测技术常用方法

- **模型驱动：** 可解释性强，模型稳定、可靠性高；会随着维度上升而在精度和效率上难以两全。
- **数据驱动：** 借助AI，适用于复杂大型系统；对数据需求大，模型不明确，可解释性低。
- **数据模型混合驱动：** 将模型驱动和数据驱动的优势相结合，尽可能的提高异常检测的性能，避免两种方法的缺点。



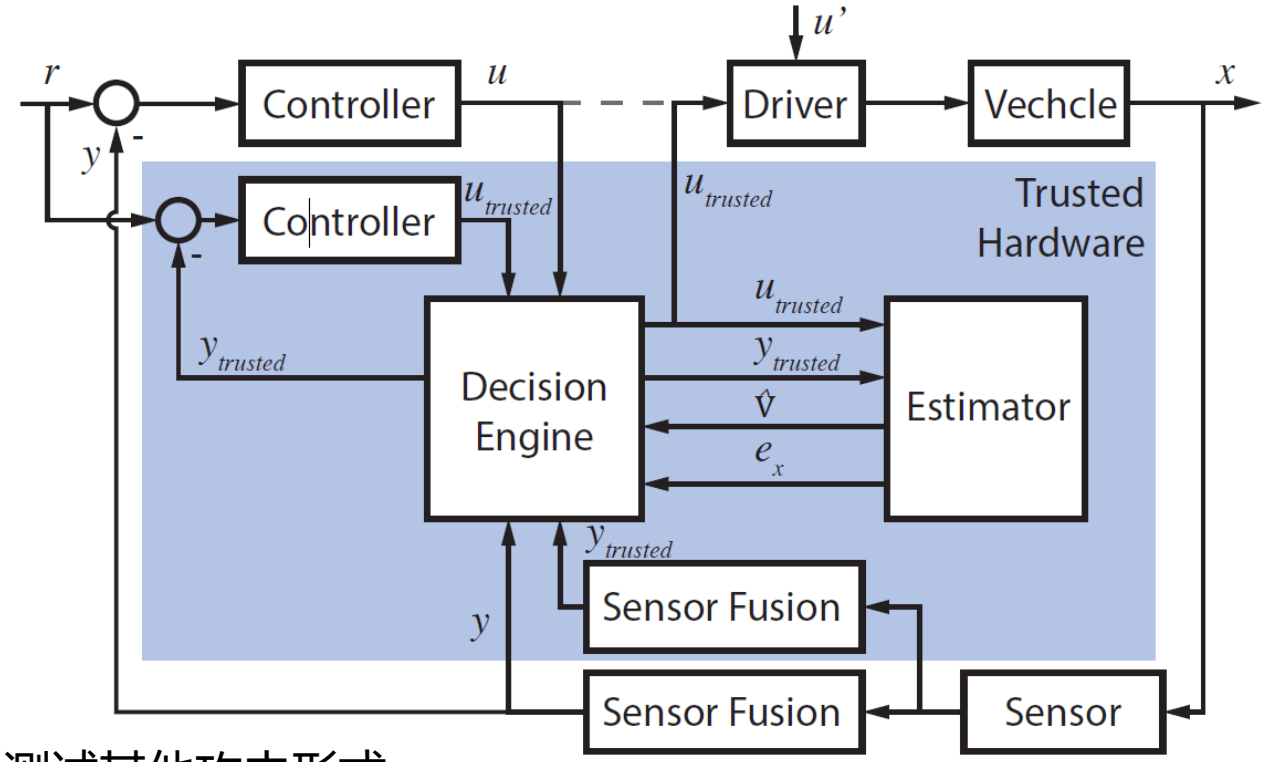




6.7.2 攻击检测

BlueBox

- 其在实际的RV上加装了由可信硬件组成的计算和存储单元，对RV的实际状态信息做出评估，利用SVSF的方法实现异常检测和诊断。
- 可信硬件直接获取飞控板处理后的结果消息，可以有效的避免软件挟持、恶意代码攻击等攻击。



缺陷:

- 文章仅测试了嵌入恶意代码的攻击形式，并未测试其他攻击形式；
- 需要外加硬件设备，增加了应用成本；
- 检测能力通过自身外置硬件上运行冗余传感器融合和控制算法进行支持。

Fei, Fan, et al, "Cross-layer retrofitting of UAVs against cyber-physical attacks." ,ICRA'18.

6.7.2 攻击检测

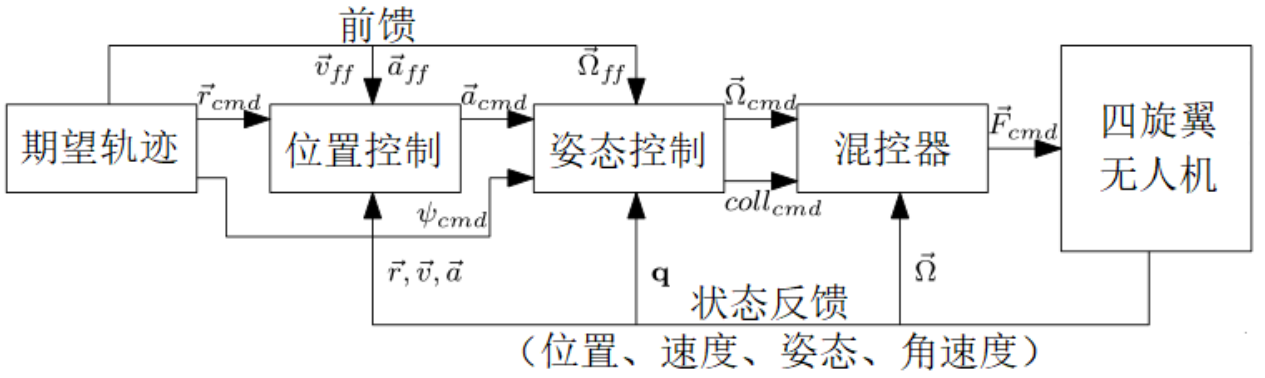
控制不变量用于无人机防护

无人机的控制不变量可由系统的状态空间模型表示：

$$\begin{cases} \dot{s} = A(s) + B(s)u(t) + d \\ y = C(s) + v \end{cases}$$

- $u(t)$（即目标状态）是系统输入，$y(t)$是系统输出。状态空间根据当前状态和控制信号确定系统的下一个状态和输出。RV的A、B、C、D矩阵通过**系统辨识**的方法获得。
- 针对四旋翼无人机而言，控制不变量主要由两个方面决定：**无人机的动力学模型**和**底层的控制算法逻辑**。

$$\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \\ \ddot{\phi} \\ \ddot{\theta} \\ \ddot{\varphi} \end{bmatrix} = \begin{bmatrix} \frac{1}{m}\left(-\left[\cos(\phi)\cos(\varphi)\sin(\theta)+\sin(\phi)\sin(\varphi)\right]\sum_{i=1}^{4} k\omega_i^2\right) \\ \frac{1}{m}\left(-\left[\cos(\phi)\sin(\varphi)\sin(\theta)-\sin(\phi)\cos(\varphi)\right]\sum_{i=1}^{4} k\omega_i^2\right) \\ \frac{1}{m}\left(-\left[\cos(\phi)\cos(\theta)\right]\sum_{i=1}^{4} k\omega_i^2\right)+g \\ \frac{1}{J_x} L\rho\left(\omega_4^2-\omega_2^2\right) \\ \frac{1}{J_y} L\rho\left(\omega_1^2-\omega_3^2\right) \\ \frac{1}{J_z}\gamma\left(\omega_1^2-\omega_2^2+\omega_3^2-\omega_4^2\right) \end{bmatrix}$$

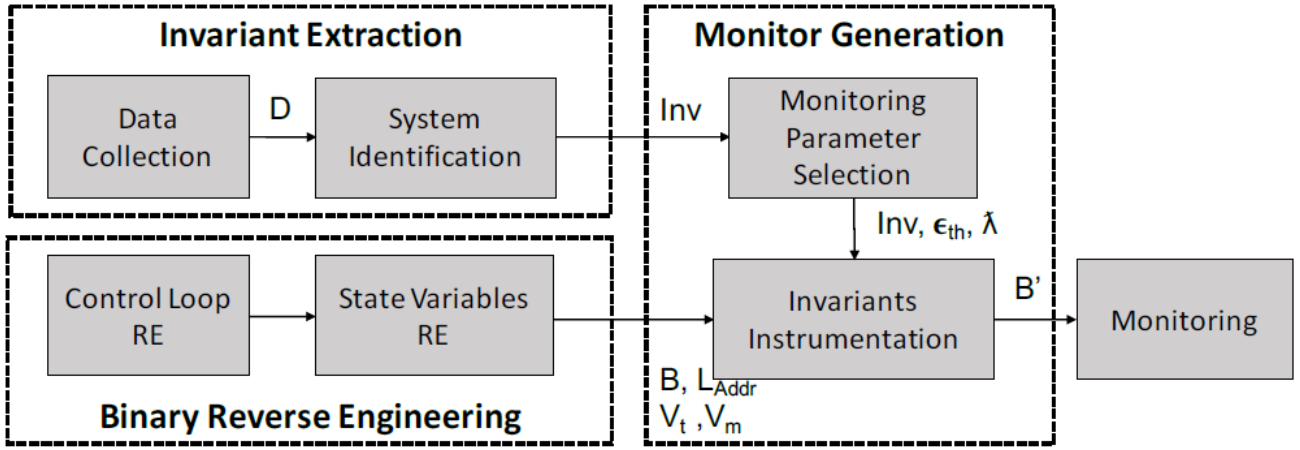


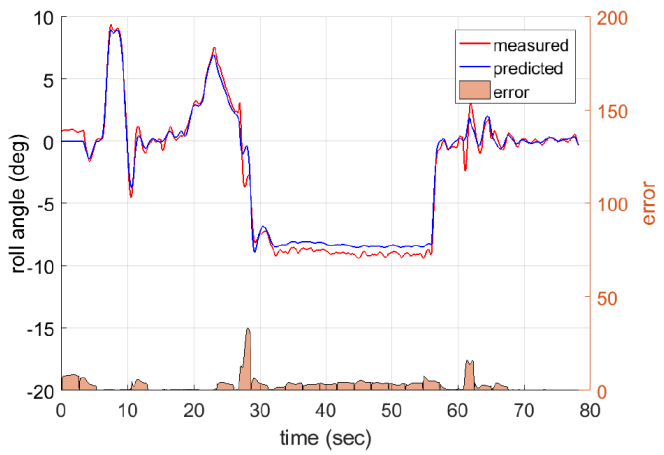
$$u(t) = K_p e(t) + K_i \int_0^{\tau} e(\tau)d\tau + K_d \frac{de(t)}{dt}$$

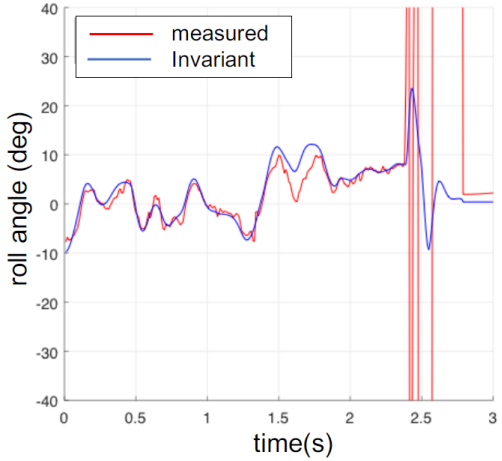
6.7.2 攻击检测

控制不变量用于无人机防护

- 低速飞行情况下，**线性模型**能很好近似实际飞行情况，便于状态空间的建立。
- 通过少量无攻击情况的飞行数据，进行系统辨识，**获得状态空间模型的系数。**
- 通过对实际飞行代码的**二进制逆向工作**，获得实际控制回路逻辑和状态变量数据。
- 采用**定时限的检测窗口，**减小检测开销。
- 环境干扰小的场景中，软件在环的模型检测能较好跟踪飞行姿态和轨迹，完成攻击监测。





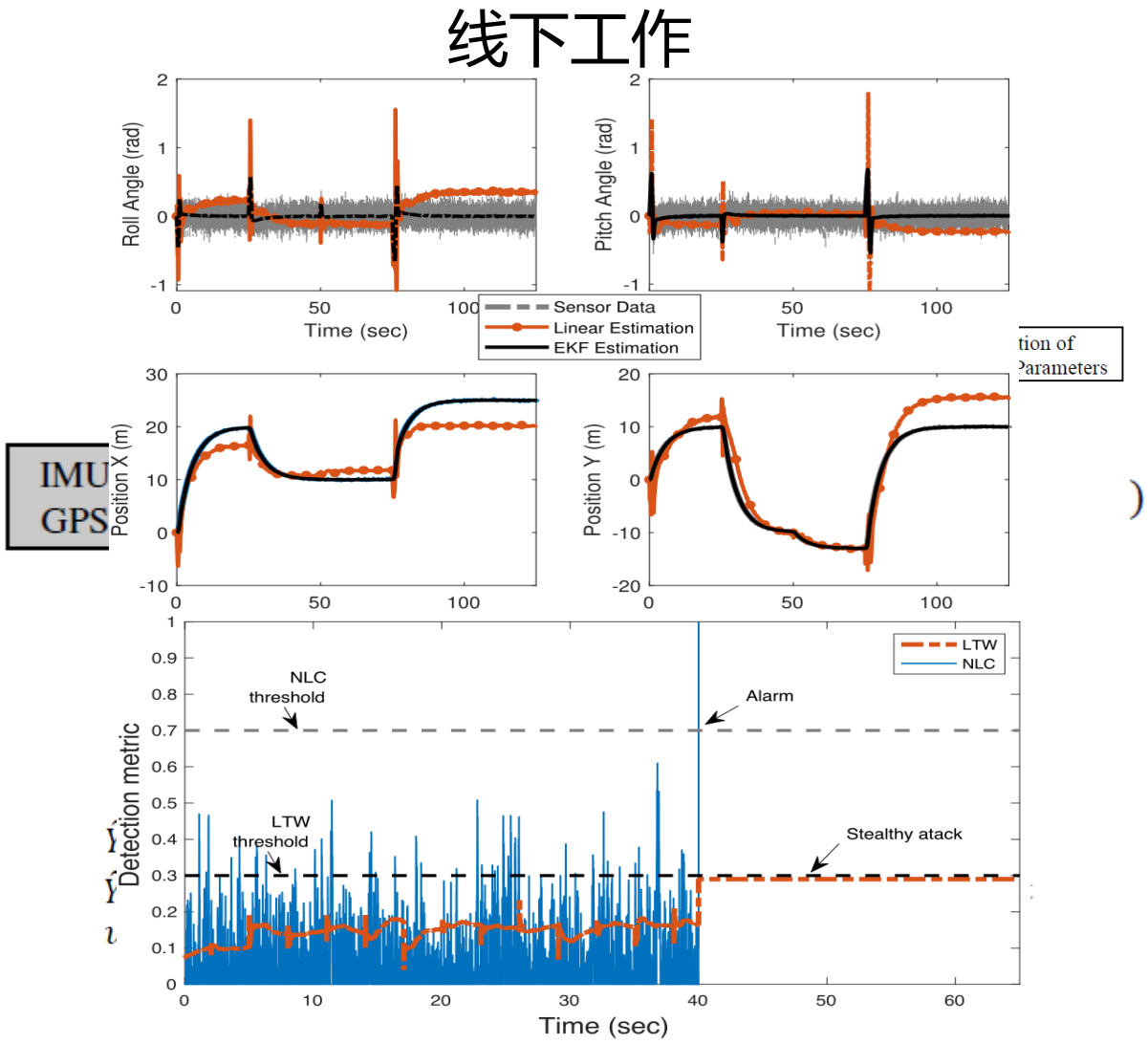


Choi et. al., Detecting Attacks against Robotic Vehicles: A Control Invariant Approach, ACM CCS'18

6.7.2 攻击检测

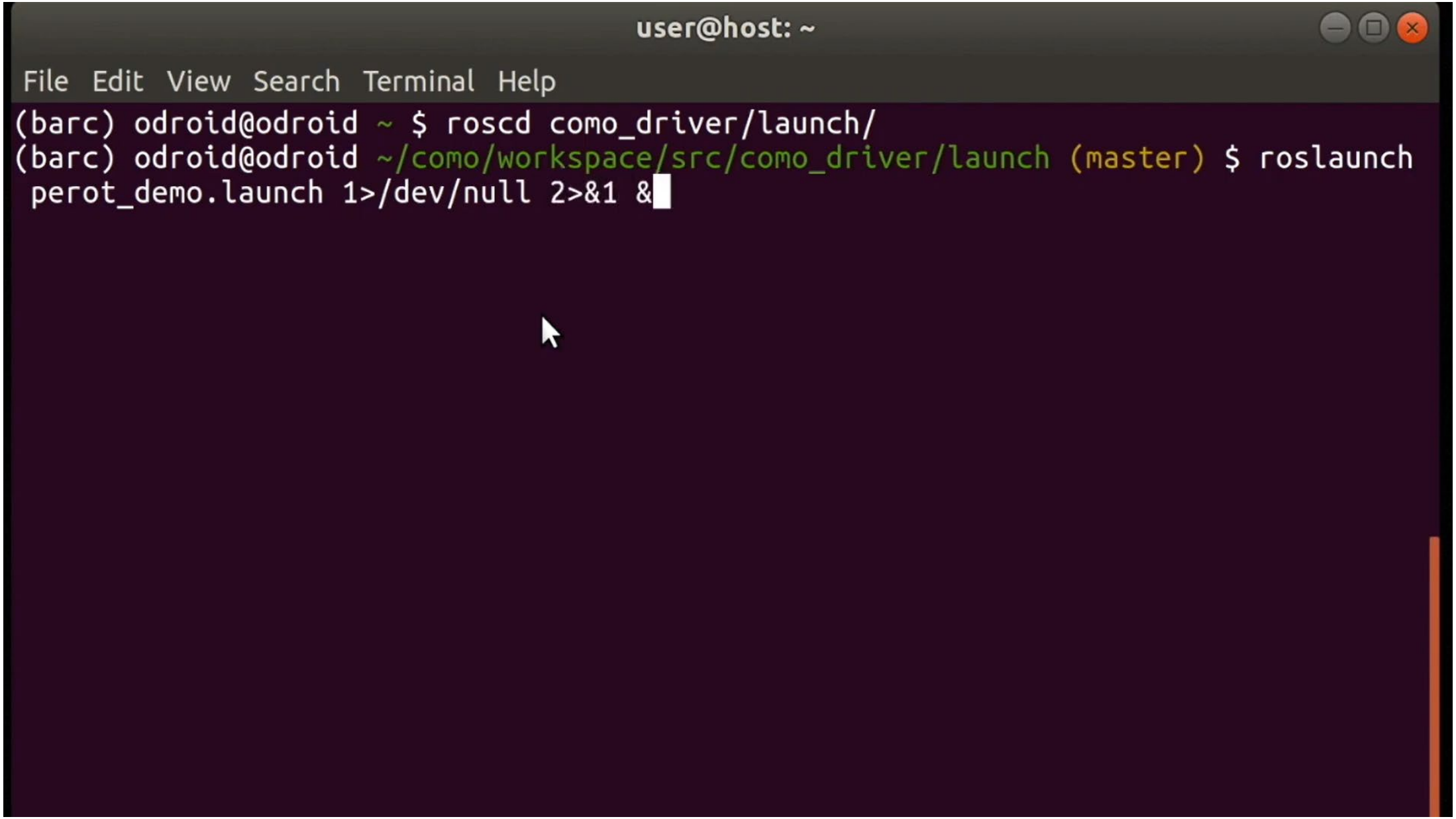
控制不变量的提升和扩展

- **非线性模型**，对于实际的输入输出拟合更为准确，提高模型的精度。
- 线下工作：系统辨识**获取飞行器物理常量。**
- 线上工作：**扩展卡尔曼滤波**完成数据融合滤波，避免网络层的攻击，**CUSUM误差累计和算法**完成异常监测工作。
- **非定时限的检测窗口**更好的跟踪残差历史变化。
- 对于微小渐进偏差的**隐式攻击检测。**
- NLC有比LTW更好的模型跟踪和检测性能，但也带来了更大的功耗开销。



Quinonezet. al., SAVIOR: Securing Autonomous Vehicles with Robust Physical Invariants, UsenixSecurity'20

6.7.2 攻击检测


user@host: ~
File Edit View Search Terminal Help
(barc) odroid@odroid ~ $ roscd como_driver/launch/
(barc) odroid@odroid ~/como/workspace/src/como_driver/launch (master) $ roslaunch
perot_demo.launch 1>/dev/null 2>&1 &

6.7.2 攻击检测

Attack Detection Demo
in Real-world Scenarios

6.7.3 实时防御

- **智能无人系统实时防御技术：**

智能无人系统实时防御技术是为了在系统运行时及时识别、防范和响应各种潜在的安全威胁和攻击而采取的一系列措施。这些技术旨在保障系统在运行时持续处于安全状态，有效应对实时的安全风险，防止潜在的威胁对系统造成损害。

- **实时防御的定义：**

智能无人系统实时防御技术是一系列用于检测、响应和抵御潜在的安全威胁和攻击的措施。这些技术包括实时监测系统活动、识别异常行为、自动化漏洞修复、云安全服务、行为分析、机器学习等，以确保系统在运行时能够快速、有效地应对各种安全挑战。

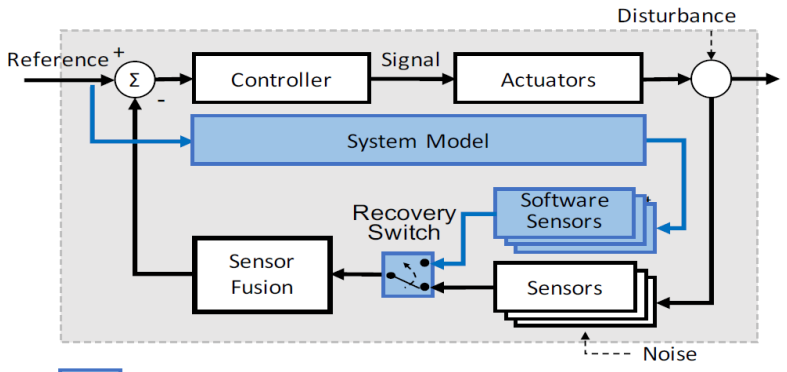
6.7.3 实时防御

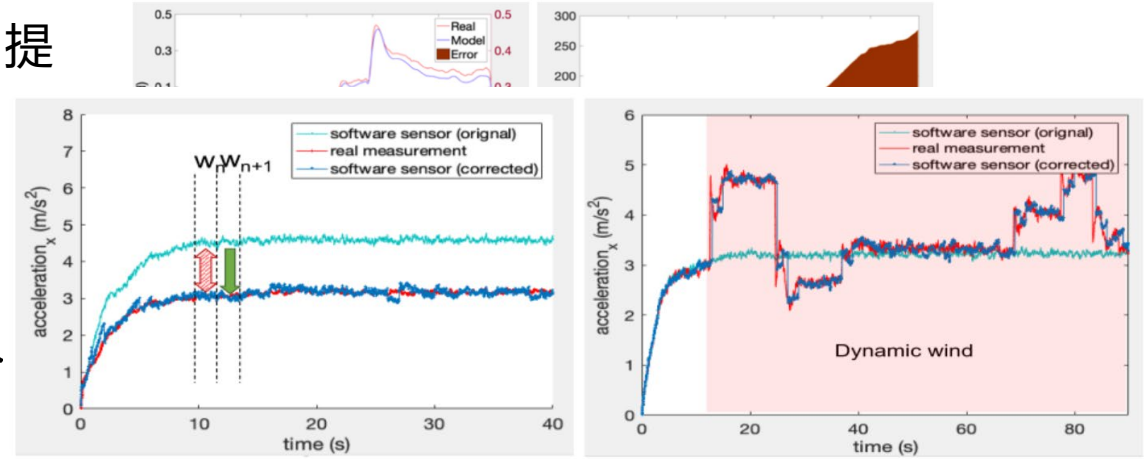
Software-based Realtime Recovery （软件传感器）

- 文章提出了一种以不变量为基础的修复传感器攻击的方法——“软件传感器”。
- 当物理传感器受到攻击时，相应的软件传感器可以单独隔离和替换受攻击的实际传感器。
- 通过集成纠错技术来补偿误差来容忍模型的不准确性，提高软件传感器的精度。

缺陷：

- 恢复过程中的漂移是不可避免的，软件传感器无法永久或长期替代物理传感器。
- APT攻击如果能逃过不变量的检测，那么就无从防御。





(a) Constant wind (b) Dynamic wind

(b) Roll prediction and error correction with synchronization

Choi, et al. "Software-based realtime recovery from sensor attacks on robotic vehicles." , RAID'20.

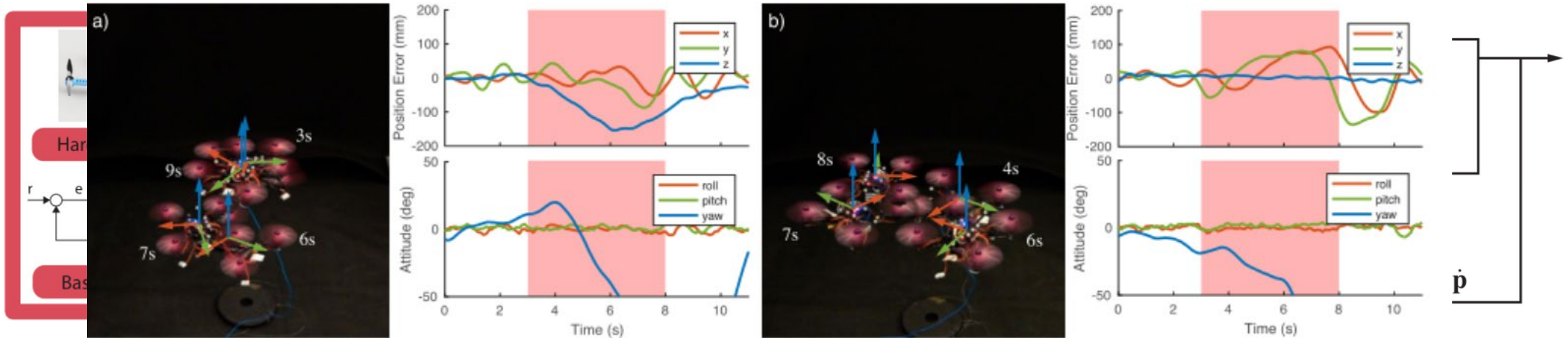
6.7.3 实时防御

Learn-to-Recover

- 文章提出了一种基于强化学习的通用容错控制策略，可以使RV从传感器和执行器故障中恢复。

方案:

- 用不变量的方法通过对源码或二进制逆向工作完成对RV模型的识别，基础控制器选取PID；
- 在高保真仿真器中依托标准RL模型进行训练；
- 将容错控制模型在实机上完成验证。



Fei, et al. "Learn-to-recover: Retrofitting uavs with reinforcement learning-assisted flight control under cyber-physical attacks." ,ICRA'20.

6.7.4 溯源定位

智能无人系统事故溯源技术：

智能无人系统事故溯源技术是一种用于追溯和分析系统发生事故或安全事件的过程。该技术旨在帮助确定事故的起因、传播路径以及相关责任方，以便更好地理解和解决系统中的问题，并采取相应的纠正措施，提高系统的安全性和可靠性。。

事故溯源的定义：

智能无人系统事故溯源技术是通过系统日志、事件记录、监控数据等手段，对系统中发生的事故或安全事件进行追踪和分析的方法。它的目标是确定事故的起因、识别相关活动、追踪事故的传播路径，以及了解事故对系统的影响，从而为事故的纠正和未来的预防提供有力的支持。

6.7.4 溯源定位

RVPLAYER（攻击溯源和取证框架）

- 构建了一个主体RV的动力学模型，并在运行过程中作为真实系统的影子系统运行。（**构建不变量检测模型**）
- 当模型能正确预测真实系统的行为时，RV被认为没有大量的环境干扰，并使用低的记录频率。否则，就使用与异常水平成正比的高频率。（**自适应的检测频率，飞控资源的优化**）
- 与传统的CPS重放技术将来自物理世界的外部输入作为一个整体处理不同，我们的技术**通过电机推力将其与环境干扰和影响解耦**，从而使前者可以独立于后者进行保存和重放，实现物理环境的再现。**（记录和解耦环境干扰）**
- 在重放过程中，应用所记录的干扰来**重现环境条件**。其可在重放期间有选择地启用/禁用某些信息，进行假设推理。**（可针对性的进行攻击分析和测试，假设推理）**
- 文章重放策略可以精确地找出研究的所有攻击的原因，没有假警报，并确定攻击的起始时间。相比之下，适应于异常检测方法的取证方法只能确定非常有限的攻击的根本原因。（**开销小，且更精准）**

Choi, et al. "RVPLAYER: Robotic Vehicle Forensics by Replay with What-if Reasoning." ,NDSS'22.

小结

- **概述**

XXXXXXXXXXXXXXXX

- **感知安全**

XXXXXXXXXXXXXXXX

- **算法安全**

XXXXXXXXXXXXXXXX

- **通信安全**

XXXXXXXXXXXXXXXX

- **控制安全**

XXXXXXXXXXXXXXXX

- **智能无人系统的安全防护**

XXXXXXXXXXXXXXXX