软件学报 ISSN 1000-9825, CODEN RUXUEW
2025,36(4):1859−1880 [doi: 10.13328/j.cnki.jos.007272] [CSTR: 32375.14.jos.007272]

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

智能网联汽车自动驾驶安全: 威胁、攻击与防护*

郗来乐[1,2], 林声浩[1,2], 王 震[1,2], 谢天鸽[1], 孙玉砚[1,2], 朱红松[1,2], 孙利民[1,2]

[1](物联网信息安全技术北京市重点实验室 (中国科学院 信息工程研究所), 北京 100085)

[2](中国科学院大学 网络空间安全学院, 北京 100049)

通信作者: 孙玉砚, E-mail: sunyuyan@iie.ac.cn

摘 要: 智能网联汽车在国家发展战略中占有重要地位, 是关系汽车产业革新、大国核心竞争力的关键技术, 自动驾驶是智能网联汽车发展的最终目标, 智能网联汽车自动驾驶 (以下称“自动驾驶汽车”) 的安全问题直接影响人民生命财产安全、国家公共安全, 但目前还缺少对其的系统性研究. 深度剖析自动驾驶面临的安全威胁能对其安全防护和保障提供指导, 促进其大规模应用. 通过整理学术界与工业界对自动驾驶安全的相关研究工作, 分析和总结自动驾驶所面临的安全问题. 首先介绍自动驾驶汽车架构、安全的特殊性, 其次从模型视角出发, 全过程地梳理自动驾驶的物理域输入、信息域输入和驾驶模型这 3 个方面可能存在的 9 个攻击作用点及其攻击方式与安全防护手段, 最后通过对近 7 年相关研究论文数据的统计分析, 总结自动驾驶安全研究的现状, 讨论未来的研究方向.

关键词: 智能网联汽车; 自动驾驶; 安全威胁; 攻击方式; 安全防护

中图法分类号: TP393

中文引用格式: 郗来乐, 林声浩, 王震, 谢天鸽, 孙玉砚, 朱红松, 孙利民. 智能网联汽车自动驾驶安全: 威胁、攻击与防护. 软件学报, 2025, 36(4): 1859–1880. http://www.jos.org.cn/1000-9825/7272.htm

英文引用格式: Xi LL, Lin SH, Wang Z, Xie TG, Sun YY, Zhu HS, Sun LM. Autonomous Driving Security of Intelligent Connected Vehicles: Threats, Attacks, and Defenses. Ruan Jian Xue Bao/Journal of Software, 2025, 36(4): 1859–1880 (in Chinese). http://www.jos.org.cn/1000-9825/7272.htm

Autonomous Driving Security of Intelligent Connected Vehicles: Threats, Attacks, and Defenses

XI Lai-Le[1,2], LIN Sheng-Hao[1,2], WANG Zhen[1,2], XIE Tian-Ge[1], SUN Yu-Yan[1,2], ZHU Hong-Song[1,2], SUN Li-Min[1,2]

[1](Beijing Key Laboratory of IoT Information Security Technology (Institute of Information Engineering, Chinese Academy of Sciences), Beijing 100085, China)

[2](School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Intelligent connected vehicles (ICVs) hold a significant strategic position within the national developmental framework, epitomizing a critical technological facet underpinning automotive industry innovations and serving as a nucleus of core national competitiveness. The culmination of ICV development resides in the realization of autonomous driving capabilities, herein termed “autonomous vehicles”. Security ramifications intrinsic to autonomous vehicles bear direct implications for public security, individual safety, and property integrity. However, a comprehensive, methodologically rigorous investigation of these security dimensions remains conspicuously absent. A comprehensive examination of the security threats germane to autonomous vehicles, thus, serves as a compass guiding security fortifications and engendering widespread adoption. By collating pertinent research endeavors from both academia and industry, this study undertakes a methodical and comprehensive analysis of the security issues intrinsic to autonomous driving. Inceptive discourse elaborates on the architectural contours of autonomous vehicles, interlaced with the nuances of their security considerations. Subsequently, embracing a model-centric vantage point, the analysis meticulously delineates nine prospective attack vectors across the tripartite domains of physical inputs, informational inputs, and the driving model itself. Each vector is expounded alongside its associated attack modalities and corresponding security mitigations. Finally, through quantitative analysis of research literature encompassing the last

* 基金项目: 国家自然科学基金 (61931019)
收稿时间: 2024-01-05; 修改时间: 2024-06-03; 采用时间: 2024-08-06; jos 在线出版时间: 2024-12-18
CNKI 网络首发时间: 2024-12-19

septennium, the prevailing terrain of autonomous vehicle security scholarship is scrutinized, thereby crystallizing latent trajectories for future research endeavors.

Key words: intelligent connected vehicle (ICV); autonomous driving; security threat; attack method; security protection

自 1984 年美国国防高级研究计划署 (defense advanced research projects agency, DARPA) 提出自主陆地车辆 (autonomous land vehicle, ALV) 计划以来, 国内外学术界与工业界对自动驾驶技术的研究热忱一直居高不下. 随着传感器技术、物联网技术、通信技术和人工智能技术的发展, 自动驾驶从概念逐步落地成为现实, 百度、特斯拉、谷歌等国内外高新科技公司纷纷进入自动驾驶领域, 各种基础设施飞速建设, 相关标准纷纷出台, 社会各层面已经准备迎接自动驾驶时代的到来.

根据《汽车驾驶自动化分级》标准[1], 自动驾驶汽车是指在任何可行驶条件下能够持续地执行全部动态驾驶任务并自动执行最小风险策略的汽车. 自动驾驶汽车相比于传统汽车出现了颠覆性变化, 可以在没有任何人类主动操作的情况下, 自主安全地完成驾驶任务. 在技术足够成熟的前提下, 自动驾驶的综合安全性会比人类驾驶高一个量级. 根据美国国家交通公路安全管理局 (national highway traffic safety administration, NHTSA) 公布数据显示, 94% 的交通事故是人为原因导致的, 不安全因素主要包括注意力不集中、决策误判、疲劳/醉酒驾驶等. 而自动驾驶可以获取比人更宽阔的视野, 在数以亿计的数据中学习经验, 以微米的粒度控制机械, 不会疲劳驾驶、酒驾和情绪驾驶. 同时自动驾驶汽车及其基础设施的应用, 也为智能交通管理和智慧城市建设提供了新机遇, 对拥堵治理、城市污染治理、交通安全等关乎人民生命财产安全的关键领域具有重大意义.

然而, 自动驾驶汽车攻击事件也屡屡发生. 根据 2022 年 NHTSA 披露数据显示, 过去一年 L2 级自动驾驶汽车报告了 392 起事故, L3–L5 级自动驾驶汽车报告 130 起事故, 这表明安全已经成为自动驾驶汽车大规模普及的首要条件. 自动驾驶汽车安全面临的挑战是多层次的, 既有传统汽车设计遗留的风险点, 如 2016 年 Charlie Miller 入侵 Jeep 自由光事件; 也有机器学习或深度学习应用导致的新威胁, 如 2016 年、2019 年和 2020 年多次发生的特斯拉撞击白色物体事件; 此外汽车联网在给用户带来便捷体验的同时, 也使自动驾驶汽车面临更严峻的安全挑战.

因此, 安全是实现自动驾驶大规模应用的前提, 驾驶模型是自动驾驶安全的关键, 系统科学地分析其面临的安全威胁对于最终实现完全自动驾驶并发挥其巨大技术效益具有重要意义. 本文综述了近 7 年学术界、工业界对自动驾驶的主要安全研究工作, 介绍了自动驾驶汽车的架构、安全特性, 围绕自动驾驶汽车基本工作流程深入分析了模型视角下自动驾驶内外全过程的威胁并提出相应对策, 最后通过对最新论文的统计分析, 总结了当前研究现状, 并展望未来的研究趋势.

本文以模型为视角关注智能网联汽车自动驾驶内外全过程安全, 与现有工作相比有很大不同. 现有工作可以分为两类: 一是以智能网联汽车整体作为研究对象, 关注汽车在组件层面的安全性[2–5], 本文与之相比不仅改进了分类方法以更深入地讨论自动驾驶的安全问题, 并且重点讨论了近年出现的新型自动驾驶安全威胁; 二是关注智能网联汽车某一领域的安全性, 如关注 CAN 总线的安全性[6], 关注基于机器学习的车联网安全[7], 关注汽车中存在的数据中毒攻击[8]、对抗样本攻击[9]以及关注自动驾驶测试[10], 本文与之相比提供了自动驾驶内外全过程安全的整体讨论. 本文旨在为智能网联汽车自动驾驶安全领域的新研究人员提供一些基本指导. 同时本文尝试总结和分析当前的工作, 并关注了端到端自动驾驶发展趋势下的新问题, 以提供以前论文未涉及的智能网联汽车自动驾驶安全的细节.

1 背景介绍

1.1 自动驾驶汽车典型架构

从自动驾驶汽车业务分层看, 一个典型的自动驾驶汽车架构分为环境感知、驾驶模型、车载部件和 V2X 通信这 4 个模块, 如图 1 所示.

环境感知模块利用摄像头、激光雷达、超声波雷达、毫米波雷达、加速度计、GPS 等传感器, 实时获取道路宽度、车道线、路标等道路信息, 检测车辆、行人、自行车等障碍物信息, 实现物理环境语义的数字化表示, 是自动驾驶汽车运行的关键基础.

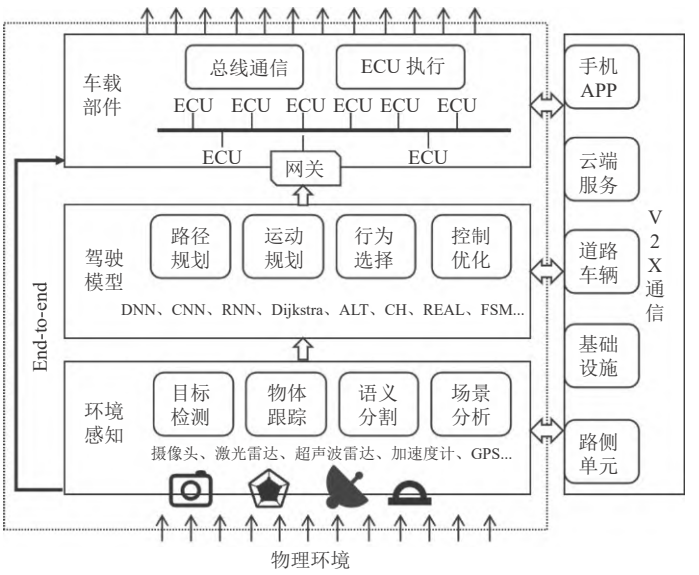

图 1 自动驾驶汽车典型架构

驾驶模型是自动驾驶汽车的核心控制模块, 通过深度学习模型 (CNN、RNN 等) 和经典算法 (Dijkstra、ALT、FSM 等), 实现路径规划、运动规划、行为选择和控制优化等任务, 根据感知语义信息和车辆状态做出超车、避障等驾驶决策, 并转化为车辆控制指令, 以控制车辆的加速、转向、制动等操作, 从而实现最优化的车辆控制, 随着端到端学习的发展, 未来驾驶模型将融合环境感知模块, 承载更多更复杂的任务.

车载部件包括油门、制动等执行器和车载信息娱乐系统等, 通过车内通信协议 (如 CAN 总线、车载以太网等) 接收驾驶模型生成的控制指令, 并将其传输到对应电子控制单元 (electronic control unit, ECU) 执行, 实际控制车辆运动状态, 确保车辆按照驾驶模型规划的路径行驶, 保证行驶的安全性和稳定性. 从网络结构上看, 自动驾驶汽车是一个异构分布式实时系统, 不同车载网络将 ECU 连接起来, 实现决策指令的下达和执行. 总线作为汽车内部网络的构建核心, 是决策指令的传输通道, 其设计与实现与汽车安全高度相关. 自动驾驶汽车内部普遍采用域控制技术, ECU 作为控制系统的最末端节点, 直接操纵物理器件完成控制指令的执行.

V2X (vehicle-to-everything) 通信模块实现车辆与道路基础设施、其他车辆和云端等的实时通信, 是自动驾驶汽车获取高清地图、道路信息的重要途径, 辅助驾驶模型进行路径规划, 从而完成行驶任务. 但 V2X 也使得自动驾驶汽车及其连接的对象在广义上成为巨大的系统, 从而加剧了自动驾驶汽车安全分析的复杂性.

1.2 自动驾驶汽车安全特殊性

自动驾驶汽车是一个多层级网络与终端融合的新型信息物理系统, 相比于传统汽车、计算机、工业控制系统等, 自动驾驶汽车具有高度智能化、泛网络连接、高实时要求、强安全保障等特点, 因此其安全也呈现出新特性.

1) 物信跨域新威胁

传统互联网安全普遍研究信息域内的攻击, 随着物联网技术的发展, 新型传感器和执行器实现了信息域与物理域的连接, 使得安全研究发生了颠覆性的变化, 物理安全得到了更多重视. 自动驾驶汽车天然地连接了物理域和信息域, 又与人们生命财产安全、城市交通安全高度相关, 这导致了许多复杂、隐蔽的新型攻击手段, 它们可以直接作用到物理世界, 造成更大破坏.

2) 功能安全与信息安全融合

功能安全是自动驾驶汽车应用的必要前提, 信息安全是自动驾驶汽车接续发展的内在要求, 随着智能化、网联化程度加深, 自动驾驶汽车逐步成为功能安全与信息安全融合的新型移动终端. 其功能安全与信息安全相互约束、相互作用, 共同使用汽车软硬件、通信等资源, 共同决定了汽车及其驾乘人员的安全. 以驾驶模型为核心的自动驾驶汽车所面临的安全问题往往表现为二者融合的结果, 因此自动驾驶汽车从设计到实现应该更关注功能安全

与信息安全融合的新型安全需求.

3) 丰富的攻击入口

自动驾驶汽车配备大量传感器 (摄像头、激光雷达、GPS 等)、提供诸多车载接口 (OBD-II、USB、IVI、充电接口等)、使用多种无线通信协议 (蓝牙、4G/5G、WiFi 等), 潜在攻击者往往通过这些丰富的入口对自动驾驶汽车实施各种攻击. 如文献 [11] 通过欺骗激光雷达可以实现控制自动驾驶汽车撞击其他车辆或障碍物, 造成严重交通事故; 文献 [12] 通过 OBD 接口, 破坏汽车总线通信传输, 瘫痪整个汽车电子控制系统.

4) 数据泄露风险剧增

自动驾驶汽车依靠强大的环境感知来理解道路状况, 这在保障汽车安全智能行驶的同时, 也使得用户个人数据和城市公共数据面临严重泄露风险. 一方面, 自动驾驶汽车采集道路交通数据、外部环境地理数据、驾乘人员行为数据、车载音视频数据和汽车控制数据等多种敏感数据; 另一方面, 自动驾驶汽车采集的车辆状态数据超过 70 维, 用户个人数据超过 20 维, 环境数据超过 30 维. 如此大规模高维敏感数据通过 V2X 通信与云端和其他车辆等共享或应用, 使得数据去向难以确定, 进一步加剧了数据泄露的风险.

2 自动驾驶汽车威胁总览: 模型视角

通常来说, 威胁是指可能造成系统信息丢失或功能失效等情况发生的事件, 可以分为蓄意攻击和无意事件[13,14]. 自动驾驶汽车是一个高度重视功能安全的强实时控制系统, 设备故障等无意事件对自动驾驶汽车安全威胁的影响已经越来越小. 与此相反, 随着自动驾驶汽车连通性、智能性与日俱增, 越来越多实体与汽车进行交互提供了更加丰富的功能, 同时也引入了更多的脆弱性, 加剧其受到蓄意攻击的安全威胁. 因此, 本文更关注自动驾驶汽车蓄意攻击相关的安全威胁.

基于自动驾驶汽车基本工作流程, 本文以驾驶模型为视角, 从物理域输入、信息域输入 2 个模型输入角度分析了外界可能引入的安全威胁, 定义了物理信号、传感元件、信号输出、感知语义、融合算法和通信链路等 6 个攻击作用点 (A–F); 从模型训练、模型执行、模型更新 3 个阶段分析了驾驶模型运行各阶段可能存在的安全威胁, 定义了训练样本、执行环境、更新信息 3 个攻击作用点 (G–I), 系统地梳理了模型视角下自动驾驶内外面临的全过程威胁, 如图 2 所示.

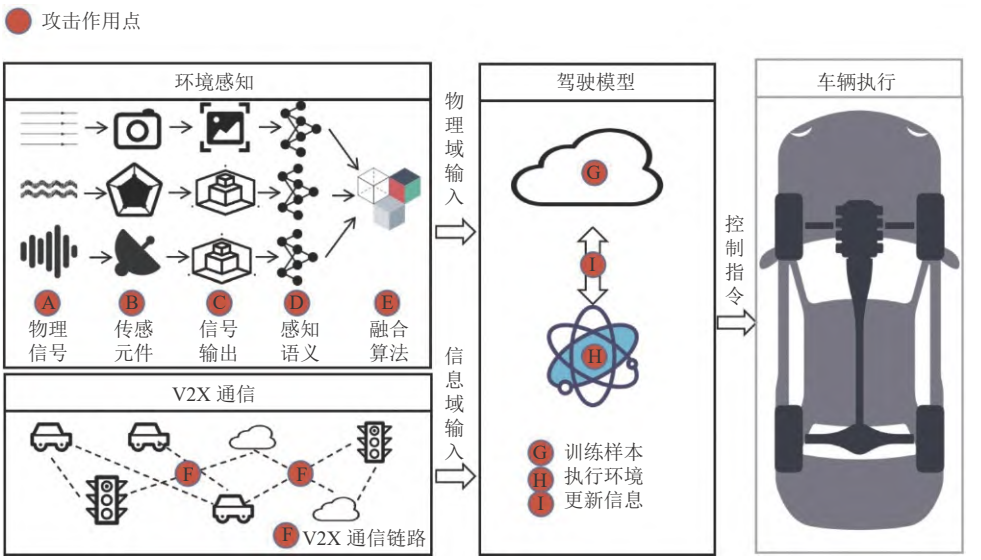

图 2 模型视角下自动驾驶汽车全过程的安全威胁

模型的物理域输入是指通过环境感知模块, 采集汽车运行所需的红绿灯、障碍物、车道线等物理环境信息. 模型的物理域输入主要面临欺骗和干扰的威胁, 是自动驾驶汽车物信跨域新威胁这一特性的集中体现.

模型的信息域输入是指通过 V2X 通信, 采集交通流量、高清地图等周围车辆、路测基础设施和云端的数据, 是自动驾驶汽车网联协同的重要手段. 本文主要关注 V2X 通信链路安全对自动驾驶汽车运行产生的影响.

驾驶模型是自动驾驶汽车的实际控制核心, 模型安全直接影响到自动驾驶汽车的安全. 本文从模型训练、执行、更新等全生命周期考察驾驶模型运行面临的问题, 不仅关注数据中毒、篡改等安全问题, 还关注了模型数据泄露问题.

需要注意的是, 基于现有汽车架构, 模型视角下的自动驾驶汽车车载部件 (包括车载总线、电子控制单元等) 仅执行驾驶模型生成的控制指令, 并不直接影响自动驾驶模型的安全性. 因此自动驾驶汽车车载部件的安全问题固然重要, 但不在本文讨论范围.

3 环境感知安全分析: 威胁、攻击与防护

感知模块是自动驾驶的基础, 直接决定了任务的完成度. 一个典型的感知模块往往包括摄像头、激光雷达、惯性导航单元 (inertial measurement unit, IMU) 和 GPS 等多传感器以及对应的数据处理方法 (数据清洗、特征提取、深度学习模型等) 与多传感器融合算法 (multi-sensor fusion, MSF) 组成. 环境数据被不同传感器采集后, 在数据处理阶段通过数据清洗保留强相关数据, 经过特征提取和目标检测之后, 形成用以描述采集对象的语义信息, 之后送入 MSF 算法中, 将各传感器关于目标的语义信息进行关联, 得到该目标的一致性解释与描述. 常见的传感器及其特性和作用如表 1 所示.

表 1 自动驾驶汽车中常见传感器的特性和作用

传感器种类	信号形式	应用场景
摄像头	自然光	识别交通标志、道路标线、行人等
激光雷达	激光束	高精度的障碍物检测
超声波雷达	超声波	检测汽车周围的近距离障碍物
毫米波雷达	毫米波	低能见度情况下检测物体的位置和速度
GPS	电磁波	提供定位信息和环境地理信息
惯性传感器	机械力	检测汽车的加速度和角速度, 提供运动状态和姿态信息

种类繁多、功能各异的诸多车载传感器显著地增加了自动驾驶的攻击面, 与人控汽车相比, 自动驾驶汽车传感器采集信息一旦出错或被恶意控制, 将直接影响汽车控制及人员安全. 本文根据传感器工作原理以及自动驾驶汽车感知流程, 梳理了 5 种攻击作用点 (A–E), 分别对应信号干扰、元件干扰、信号欺骗、语义欺骗和联合攻击这 5 类攻击.

3.1 信号干扰

干扰在本节是指以干扰车辆感知模块为直接目的, 并由攻击者操纵的恶意干扰, 而不包括由于复杂电磁环境引发的无意干扰. 根据对象来分, 干扰可以分为信号干扰和元件干扰.

信号干扰主要指增加频率噪声基底以降低目标物理信号 (攻击作用点 A) 的信噪比. 信噪比越大, 说明信号中含有更多有效信号, 则传感器接收信号的质量越高, 反之则传感器越难正确接收信号. 信噪比在一定阈值之下, 会使目标传感器被致盲. 信号干扰主要存在于以电磁波作为信号载体的 GPS、毫米波雷达等传感器中[15].

GPS 信号是实现自动驾驶汽车定位、导航和授时功能的重要基础, 但是 GPS 无法在存在干扰的情况下保持准确、连续和可靠的信号, 极易导致汽车失控、时间同步丢失和人员伤亡, 严重危及汽车及交通系统安全. GPS 干扰的本质是干扰信号的功率抑制了导航信号的功率, 使其无法准确定位[15]. 文献 [16−18] 评估了全球导航卫星系统 (global navigation satellite system, GNSS) 干扰器对自动驾驶的威胁程度, 着重强调了抗干扰能力对汽车安全运行的重要性.

毫米波雷达主要用于在恶劣天气或低能见度情况下检测物体的位置和速度. 然而由于雷达工作频段大多相同, 导致多部雷达在同一时间以相同的频率发射可能会发生相互干扰[19]. 干扰能够使本底噪声水平升高从而降低目标的检测概率, 可能导致雷达截面积 (radar cross section, RCS) 较小的目标消失在雷达视野中[20,21]. 文献 [22] 演

示了通过发送相同的波形信号对特斯拉 Model S 毫米波雷达的干扰攻击, 可以致盲受害车辆.

由于目前毫米波雷达、激光雷达等传感器普及率还不高, 在传统安全观念中, 较欺骗而言, 干扰的威胁程度较低. 然而, 随着自动驾驶技术逐步落地, 未来自动驾驶汽车必将装备大量车载传感器, 76–81 GHz 频谱将被大量占用[23], 因此干扰问题需要得到足够重视. 同时研究者在对毫米波进行干扰攻击时发现相比于传统汽车, 毫米波雷达干扰攻击对自动驾驶汽车威胁更大[24].

3.2 元件干扰

元件干扰是指通过声光电磁热等媒介, 利用电磁感应、瞬态变化、共振等物理现象带来的带外脆弱性, 通过传感器的非预期设计功能, 干扰物理传感元件 (攻击作用点 B) 转换测量过程, 使得自动驾驶汽车感知能力下降、错误甚至完全丧失.

传感器在工作时通常由敏感元件、转换元件、变换电路和辅助电源这 4 部分组成, 敏感元件是指传感器中能直接感受或响应被测物理量的部分; 转换元件是指传感器中能将敏感元件感受或响应的被测物理量转换成适合于传输或测量的电信号部分; 变化电路用于进行信号转换、放大、运算与调制, 以便于使传感器输出数据显示和参与控制. 本文所指传感元件主要是指敏感元件、转换元件和变换电路. 通过带外方式对物理元件施加影响, 可以增大、减小或破坏敏感元件的感受能力[25,26]、转换元件的转换能力或变换电路的调理能力[27,28], 使之偏离正常值 (未施加影响的测量值), 当这种改变的能力足够大时, 就可以控制传感器的工作, 实现控制车辆行驶的效果. 表 2 列举了现有元件干扰的研究工作.

表 2 元件干扰相关研究工作

攻击目标	攻击原理	攻击后果	文献
陀螺仪	调节扬声器声音频率, 使其与陀螺仪发生共振	干扰陀螺仪测量角度变化量	[27]
加速度计	利用声波会对其传播路径上物理对象施力的原理, 引起加速度计电容变化	控制加速度计测量值	[28]
摄像头	恶意增加摄像头自动曝光	致盲摄像头	[25]
摄像头	激光照射下, 由于温度场引起的热应力, 摄像头表面温度会上升	损坏摄像头	[26]

3.3 信号欺骗

信号欺骗是指攻击者通过变造传感信号 (如 GPS 信号、雷达信号等), 将其直接注入对应感应装置, 欺骗传感器的直接输出 (攻击作用点 C). 这种攻击无需考虑内部数据处理过程. 信号欺骗广泛存在于 GPS、毫米波雷达、超声波雷达等无需复杂数据处理的传感器中.

早期的欺骗技术较为简单, 如直接增加 GPS 强度[29,30], 重放激光、音波、毫米波信号[25]等. 由于存在断续、突变等原因, 这些攻击极易被察觉[31], 于是攻击者设计了两阶段的欺骗技术, 首先进行信号同步, 如对 GPS 进行原始信号与欺骗信号的平滑同步[32]、跟踪激光发射角度和强度[33]; 随后通过改变信号的到达时间或频率[25,32]来操纵传感器, 实现隐藏或制造障碍物的效果. 如文献 [34−36] 指出当两个或更多雷达之间的定时、波形和频率相匹配且回波信号功率超过一定阈值的时候, 会产生虚假目标从而造成雷达的误检. 文献 [37] 研究了汽车雷达的线性调频序列 (chirp sequence, CS) 与各种波形 (如连续波、FMCW 和 CS) 相互作用的效果, 推导了虚假目标出现的概率.

总的来说, 由于缺少人的参与, 无法通过人工验证信号真实性, 信号欺骗攻击在自动驾驶场景下具有更大的威胁[15,19]. 到目前为止, 信号欺骗技术已经较为成熟, 研究者主要关注欺骗实现的成本与效率[16,38].

3.4 语义欺骗

语义欺骗是自动驾驶汽车特有的攻击方式, 是信号欺骗的特殊形式. 通俗地讲, 语义欺骗主要是利用对抗性输入, 按照一定规则向传感器输入刻意制造并富含语义的感知信号, 旨在欺骗感知模块的目标检测模型, 产生如虚假障碍物信息等具有欺骗效果的语义信息. 与信号欺骗相比, 语义欺骗更具有威胁性, 特别是基于雷达和摄像头构建的感知系统. 一般来说, 在自动驾驶背景下, 对雷达、摄像头的信号欺骗已经很难经过预处理阶段成功转化成语义[11], 而这两个传感器极大程度上决定了自动驾驶感知的精确性, 因此对感知语义的攻击将对自动驾驶汽车及其人员造成致命的威胁, 而且它的攻击实现更为隐蔽, 通常运行中的自动驾驶汽车难以检测抵御此类攻击.

语义欺骗主要发生在作用点 D, 攻击需要考虑内部数据处理过程, 是信号级欺骗的进一步发展, 可以分为攻击向量构造阶段和攻击向量传递阶段, 如图 3 所示.

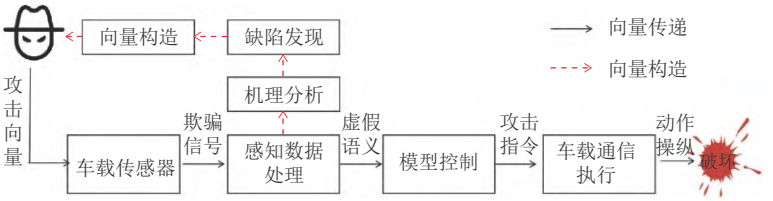

图 3 语义欺骗攻击向量构造

攻击向量构造阶段可以分为机理分析、缺陷发现和向量构造 3 个步骤. 机理分析的主要对象是自动驾驶汽车内部感知数据处理保护机制, 目的是找到能够通过处理保护机制的数据模式; 缺陷发现通常利用对抗性学习对找到的数据模式进行特征分析, 找到影响数据筛选和处理的关键特征; 向量构造则利用找到的关键特征, 构造能够产生攻击语义的异常向量. 如文献 [11] 发现自动驾驶汽车激光雷达三维点云中欺骗点的数量和位置对攻击成功具有决定意义, 从而将攻击向量构造问题转化为寻找最佳点云集的优化问题; 文献 [39] 通过识别障碍物表面的关键点位, 在这些位置放置少量具有反射表面的物体, 使得自动驾驶汽车激光雷达无法检测到障碍物; 文献 [40] 通过激光干扰或物理遮挡, 移除激光雷达点云中的关键点, 使自动驾驶系统无法检测到特定的障碍物; 文献 [41] 通过增加对抗性的检测框来改变摄像头的环境感知语义; 文献 [42] 通过使用特定形状的物体来反射毫米波雷达信号, 从而误导 DNN 模型的检测结果.

攻击向量传递阶段可以分为攻击向量注入、欺骗信号输出、虚假语义产生、攻击指令生成、动作操纵执行这 5 个步骤, 其中攻击向量的注入是关键步骤, 而其他 4 步骤都是在攻击向量被传感器接收后, 伴随自动驾驶工作流程自动完成的. 攻击向量注入的研究对象是传感器接收装置、信号传播介质以及信号发射位置, 如文献 [11] 研究激光发射的角度、距离和倾斜度对于激光雷达欺骗的作用, 给出了最佳的欺骗角度、距离的组合.

3.5 联合攻击

信号干扰、元件干扰、信号欺骗和语义欺骗都是针对单传感器的攻击手段, 随着自动驾驶感知技术的进一步发展, MSF 成为校正单传感器测量结果、防御单传感器攻击的重要方法.

针对 MSF 算法的攻击已经成为当前的研究热点. 自动驾驶汽车通常配备多个不同传感器, 通常采用多传感器融合感知的架构, 单一的欺骗攻击手法已经不足以对自动驾驶汽车感知产生严重安全威胁, 攻击者更多地通过细微的观察和精妙的构造, 研究实现对多传感器的联合攻击. 联合攻击是指针对性的破坏多传感器融合感知算法 (攻击作用点 E) 的攻击手段, 一般表现为攻击者同时攻击多个传感器.

对于不同的融合算法实现和工作原理, 联合攻击具体形式也分为 3 类: 共同欺骗、机制破坏和误差利用. 共同欺骗是指同时欺骗所有传感器从而达到欺骗感知模块的作用, 难点在于攻击向量的选取, 以产生多欺骗效果, 如文献 [43] 通过优化方法攻击了基于激光雷达和摄像机的 MSF 框架, 设计了 3D 打印物体, 既可以改变形状 (针对激光雷达), 也可以改变纹理 (针对摄像机), 从而可以同时作为激光雷达和摄像机的攻击向量; 机制破坏是指压制或干扰其他传感器正常工作, 只保留少数甚至一个传感器进行感知, 从而使得多传感器感知退化成单传感器感知, 难点在于传感器的压制或干扰技术的设计, 如文献 [44] 研究了基于激光雷达、GPS、IMU 的 MSF 框架, 发现在持续的 GPS 欺骗下, MSF 会出现不自信的时期, 在此期间攻击者控制的 GPS 欺骗信号逐渐成为 MSF 的主要输入从而排斥激光雷达、IMU 等输入, 可以诱导自动驾驶汽车偏离正常车道; 误差利用是指基于传感器本身测量的不精确性并通过干扰其修正或补偿机制, 从而不断扩大由不精确性引发的测量误差的攻击方法, 如文献 [45] 研究了误差对 IMU 的威胁性, 可以根据当前的交通流、交通灯以及意外拥堵情况, 通过对误差的不断累积, 使得汽车在不被监控系统发现的前提下偏离既定目的地超过 10 km, 这对于罪犯管控、贵重物品运送等具有严重威胁.

3.6 防护措施

在驾驶模型为中心的视角下, 自动驾驶汽车在环境感知模块的核心目标是保证来自物理域的输入正确及时的

反映当前物理环境信息. 从保证物理域输入的正确性来看, 保证传感器感知的正确性是关键要素, 当前主要有两类防护措施: 传感器安全加固方法和多传感器融合方法.

1) 安全加固方法: 自动驾驶汽车感知层存在攻击的根本原因之一是各类车载传感器精度不够, 缺少适当的安全防护机制, 使得汽车在遭受攻击时难以判断感知信息的真假, 因此研究高精度传感器、增强传感器保护机制是解决感知安全威胁的必由之路.

增加认证、加密机制是增强传感器保护的重要方式, 可以有效抵抗欺骗攻击、干扰攻击. 由于车载传感器与其他类型的设备存在显著的差异, 实现认证机制需要特殊的认证因子. 如文献 [46] 提出基于光电晶体管阵列的传感器内加密技术, 实现高度可信的图像认证以避免被动攻击和未授权的版本. 文献 [47] 利用侧信道信息来认证消息, 测信道信息来自于汽车中的一个加密设备. 该设备正在使用 AES (高级加密标准) 对一个密码密钥进行大量计算, 并读取这些计算过程中的电磁辐射. 然后, 这些信息被用来调制和解调激光器的振幅. 然后它只会接受这种调制的回声. 为了更安全地进行认证, 可以搭载基于安全芯片的可信执行环境, 并在此基础上实现认证机制.

攻击检测是增强传感器保护的另一重要方法, 主要分为特征检测和异常检测两类. 特征检测技术通过生成攻击特征库, 从而识别已知攻击. 异常检测技术与特征检测相反, 通过机器学习的方法生成正常行为模式, 将与正常行为模式差距超过一定阈值的行为判定为入侵攻击, 从而有效识别未知攻击. 车载传感器面临的攻击事件有限, 一般可采用特征检测技术[48–51], 一旦出现新的攻击形式, 可添加新的攻击特征. 由于汽车运行场景特殊, 因此检测技术的设计具有特殊性, 如文献 [33] 利用饱和度来检测恶意激光信号, 但是这种方法在拥挤的道路上会大量误报, 可能对具有内置故障安全模式的汽车造成危险. 同时自动驾驶运行的上下文信息也可以用以实现攻击检测, 如文献 [52] 利用车辆轨迹的时空一致性以及上下文数据实现了对感知攻击的交叉检测.

此外, 一些高新材料的使用也可以增强传感器的安全性. 文献 [25] 提出使用可拆卸的近红外切割滤光或光致变色透镜片可以防御致盲攻击. 这些透镜可以过滤恶意强光, 但会影响汽车的红外夜视功能, 可以搭配智能算法来实现仅攻击时使用.

2) 多传感器融合方法: 自动驾驶汽车感知层存在攻击的原因之二是单一传感器不可避免地存在感知短板, 易被攻击者利用使得传感器失效, 因此利用多传感器互补的防御方法成为当前自动驾驶汽车的主流方案. 如文献 [53] 首次系统地介绍了 MSF 基本思想和应用, 文献 [54] 综述了 MSF 最新的技术发展. 文献 [55] 利用计算机视觉和机器学习来分析相机和激光雷达数据上的特征, 通过不能同时映射到两者上的特征来检测恶意攻击. 文献 [56] 提出了一种利用总包机制检测 GPS 攻击的方案.

此外, 冗余也是重要的防护手段, 可以通过交叉验证恶意点来更好地防止欺骗. 文献 [25,26] 证明具有重叠视角的摄像头至少会使致盲攻击更难执行. 冗余在激光雷达中也有积极作用[33], 使用冗余的激光雷达传感器, 受害者车可以在受到攻击时放弃来自被攻击传感器的输入. 冗余通常会大幅增加汽车成本, 为此文献 [25] 提出通过 V2V 技术, 受害者车辆可以与相邻数据进行交叉验证, 观察不一致性.

4 V2X 通信安全分析: 威胁、攻击与防护

未来自动驾驶汽车将越来越多地依赖于高质量无线通信实现与外部环境连接, 获取信息域的各类数据, 辅助驾驶模型完成行驶任务. 自动驾驶汽车相比于传统汽车的另一特点是内外交互频繁, 由此引入诸多新威胁. 因此外部通信链路的安全性也将影响自动驾驶汽车的安全.

自动驾驶汽车与外界通信主要通过 V2X (vehicle-to-everything) 技术, 如图 4 所示, 主要包括 V2N (vehicle-to-network)、V2V (vehicle-to-vehicle)、V2I (vehicle-to-infrastructure)、V2R (vehicle-to-roadside unit) 这 4 类.

1) V2N: 车辆与云端通信, 基于蜂窝网络 (3G、4G、5G)、路边 WiFi 以及卫星等技术将车辆连接到提供交通更新服务的云端;

2) V2V: 车辆间通信, 实现道路中车辆间信息和数据的共享;

3) V2I: 车辆与基础设施通信, 实现车辆与交通信号灯或铁路交叉口等基础设施的连接, 获取路况信息;

4) V2R: 车辆与路侧单元的通信是 V2I 通信的特殊情况, 将车辆与路侧单元连接起来, 实现车辆速度测量等.

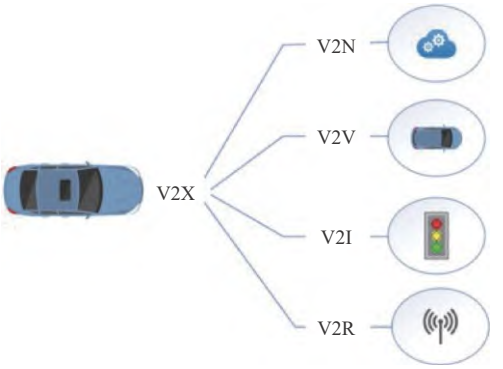

图 4 V2X 主要通信方式

V2X 通信是一种 ad-hoc 网络, 具有网络规模大、无固定的安全基础设施、开放的通信介质和节点分布不均匀等特点, 导致在车辆快速移动中 V2X 通信具有快速变化的动态拓扑结构、频繁的连接中断等脆弱性, 使得 V2X 在通信链路 (攻击作用点 F) 上容易遭受诸多攻击. 随着移动通信对低时延、高可靠、高带宽、大容量的性能要求, V2X 必须关注安全问题. 因为不准确的传入数据或恶意欺骗数据可能会导致通信拥塞、能源消耗, 甚至危及人们的生命. 本节梳理了针对 V2X 通信的 12 类典型攻击方式, 按照其破坏的安全属性分为破坏可用性、破坏完整性和破坏机密性这 3 类, 其攻击原理、攻击后果和相应的防御方案如表 3 所示.

表 3 V2X 通信安全相关研究

攻击目标	攻击方式	攻击原理	攻击后果	防御技术
破坏可用性	拒绝服务攻击[57]	将大量的非法申请封装成包传送给指定的目标主机	耗尽自动驾驶汽车通信资源, 致使通信高时延甚至不可用, 无法连接云端	加密认证[58]、IP阻塞[59]、蚁群优化[60]
	女巫攻击[61]	创建多个虚拟节点, 以多重身份和不同位置出现, 向网络中注入虚假信息	显著降低智能交通系统的工作质量, 造成交通拥堵、灾难性道路交通事故等	自编码器[61]、RNN[61]
	干扰攻击[62]	干扰机间歇性地发射信号以中断V2X通信信道	迫使自动驾驶汽车与外界失去联系	Deep Q-Network[62]
	交火攻击[63]	僵尸机向关键链路发送大量低强度流量	使得特定区域的车载V2X网络被“隔离”	ANN[63]、CNN [63]、LSTM[63]
	黑洞攻击[64]	周围节点误以为攻击者所在节点属于最短路径	自动驾驶汽车无法接受关键信息	FFNN[64]
	虫洞攻击[65]	使虫洞隧道附近节点的邻居列表混乱, 路由发现协议失效	导致自动驾驶汽车丢失特定的数据包, 也会阻止车辆发现合法通信节点	KNN[65]、SVM[65]
	Platoon攻击[66]	修改自动驾驶汽车队列的局部控制指令	破坏整个Platoon的稳定, 导致交通拥堵等问题	CNN[66]
	虚假数据注入攻击[67]	通过捕获开放信道的频段, 向V2X网络中发送虚假数据	造成车载应用的失效, 危及乘客和行人的安全	RNN[67]
破坏完整性	空中下载攻击[68]	劫持自动驾驶车辆之间的传输通道	误导乘客使用恶意软件, 造成车辆系统植入后门	密钥管理[68]
	中间人攻击[69]	窃听、延迟、丢弃或篡改通信链路中的真实信息	影响自动驾驶汽车路况信息的传递	单向哈希函数[69]
	灰洞攻击[70]	拦截并篡改合作意识信息或基本安全信息等	导致车辆获取错误的交通信息	FFNN[70]、SVM[70]
破坏机密性	欺骗攻击[71]	伪装成一个合法的网络用户	窃取车主身份、地址或汽车协议地址、服务器等信息	基于强化学习的身份验证[71]

4.1 破坏 V2X 通信可用性

在驾驶模型做出关键决策 (如制定或改变行驶路径) 时, 往往持续依赖 V2X 传来的数据, 通信链路的可用性是

最基本的要求, 任何预期之外的通信链路时延或停止服务都会对驾驶决策造成影响. 破坏可用性的关键在于使目标车辆或主机及其通信链路拒绝服务或无法正常服务, 本节总结了 3 类攻击模式.

1) 发送“垃圾数据”, 恶意消耗目标资源

如果自动驾驶参与方耗费过量的资源处理接收的“垃圾数据”, 将使得正常信息得不到及时处理, 甚至导致自动驾驶汽车与外界失去联系, 如拒绝服务攻击[57]、女巫攻击[61]、干扰攻击[62]、交火攻击[63]等.

2) 利用路由协议缺陷, 阻塞目标通信链路

V2X 通信是具有快速变化的动态拓扑结构的自组织网络, 必须依赖路由发现协议完成消息的发送, 但是由于这一过程可以被欺骗, 使得恶意攻击者可以通过篡改邻居列表等方式吸收受害车辆的流量, 使其无法接受关键信息. 如黑洞攻击[64]、虫洞攻击[65].

3) 篡改局部控制指令, 引发 Platoon 控制异常

Platoon 控制是车联网中的重要概念. “Platoon”是指在同一车道上距离很近、车速相同一组车, 将之作为一个整体参与通信可以节省能源、提高道路容量、有效地管理交通[72,73]. 攻击者为破坏 Platoon 的稳定性采取的任何类型行动都被称为 Platoon 攻击. 如攻击者通过修改局部控制指令来破坏整个 Platoon 的稳定[66], 导致车祸、交通阻塞等.

4.2 破坏 V2X 通信完整性

破坏完整性的关键在于破坏 V2X 通信传输的一致性, 包括数据体的一致性以及上下文的一致性. 完整性被破坏, 一般会使得自动驾驶汽车获取错误的交通信息、车主下载恶意软件造成汽车系统植入后门, 危及乘客或行人安全.

破坏 V2X 通信完整性一般要先劫持通信链路, 然后通过篡改实现对链路中消息的完整性破坏. 劫持通信链路一般通过中间人攻击[69]实现, 在通信节点间插入恶意节点充当第 3 人, 控制车辆间通信过程. 成功劫持链路后, 攻击者就可以破坏通信链路的数据一致性, 如虚假数据注入攻击[67]、空中下载攻击[68]和灰洞攻击[70].

4.3 破坏 V2X 通信机密性

破坏机密性的目的在于非法获取自动驾驶汽车在通信过程中传输的敏感信息. 由于敏感信息不会直接造成自动驾驶汽车的运行安全问题, 因此并未得到安全研究者足够重视. 但是获取自动驾驶汽车的敏感信息有时会对攻击实现提供重要的辅助作用, 如获取汽车协议地址、服务器等信息, 可以有针对性地实现对通信链路劫持. 文献 [71] 通过伪装成一个合法的网络用户, 获得对车辆及驾驶员信息的访问, 窃取身份、物理地址、域名服务器 (DNS) 信息、互联网协议 (IP) 地址等信息.

4.4 防护措施

随着自动驾驶汽车网联化进一步深化, 未来 V2X 通信将成为汽车从网络环境中获取信息的重要途径, 如通过与其他车辆交互获取道路路况、智能寻找停车位等. V2X 通信的安全性将决定着自动驾驶汽车集群的安全性与可靠性, 如果攻击者能够成功攻击车群内的一辆车, 以此为跳板, 攻击者可能对集群里其他车辆实现间接攻击, 从而导致恶性大规模连环交通事故的发生. 因此保障 V2X 通信安全对于自动驾驶至关重要.

V2X 通信需要满足实时性、动态性等特点, 传统的安全保护措施例如基于密码学的加密认证等方法具有较大的局限性, 增加了处理时间和网络开销, 如基于 PKI 的身份认证方案[74]、动态密钥管理方案[75]等. 因此现有基于密码学的工作重点关注方案在时间和空间上的高效性, 如文献 [76] 提出了一种基于椭圆曲线加密的安全管理方案, 通过密钥交换、数字签名和加密来确保车辆身份和信息交换的真实性, 提供完全前向保密性和组前后保密性, 抵御重放攻击和中间人攻击.

由于检测方法对计算资源要求低, 不对协议进行改变, 因此目前基于机器学习或深度学习的攻击检测方法已经成为主流, 如表 3 所示. 根据检测领域的不同, 主要有基于车辆的时间、位置、速度、加速度和加速度变化量等数据特征, 利用 SVM 算法检测女巫攻击[77]; 基于车辆速度、相对位置等数据特征, 利用 CNN 检测针对 Platoon 稳定性的攻击[66]; 基于数据包 ID、IP 来源、有效载荷等 15 种流量特征, 分别利用 FFNN 和 SVM 算法检测灰洞攻击[70]; 基于 SUMO 交通模拟器生成目标 IP、传输和接收的字节数、丢弃的字节数等行驶轨迹流量特征, 使用

KNN 和 SVM 算法检测虫洞攻击[65].

此外, 为了应对 Platoon 控制攻击, 文献 [78] 提出了一种 Platoon 弹性控制器的设计方法, 基于 DoS 攻击对系统传输延迟和服务时间的影响考虑, 通过对闭环系统的动态进行多面体过近似来合成鲁棒控制器, 进而推导出满足 L2 字符串稳定性的条件以优化控制器参数, 从而保证 Platoon 在各种攻击模式的稳定性和性能.

5 驾驶模型安全分析: 威胁、攻击与防护

在端到端自动驾驶迅速发展的情况下, 以深度神经网络 (deep neural network, DNN) 为代表的各类模型是实现自动驾驶汽车各项功能的控制核心, 广泛应用于自动驾驶汽车的图像识别、轨迹追踪、路径规划、行为预测等领域[79], 一般分为前馈神经网络 (feed-forward neural network, FFNN)、卷积神经网络 (convolutional neural network, CNN)、循环神经网络 (recurrent neural network, RNN) 及其变体. 它们的具体特点与在自动驾驶汽车中的应用如表 4 所示.

表 4 神经网络在自动驾驶汽车中的应用

名称	特点	解决任务
FFNN	单向信号传播	车辆速度预测、路径规划、控制
CNN	卷积计算的前馈神经网络	车道线识别、目标检测、人体姿态估计[80]
RNN	输入是序列数据, 可以表征记忆	命令语音识别、视频处理, 车辆状态精准预测与控制[81]

出于实现成本、运行效率和数据表征等多方面考虑, 汽车厂商在自动驾驶汽车中所使用模型的总体架构往往是表 4 中多种神经网络的叠加. 如特斯拉使用的 HydraNet 架构以及 Google 的自动驾驶汽车 Waymo 使用的主动学习框架[82]. 此外, 端到端学习技术的发展克服了传统的自动驾驶模型需要人为设计特征提取器的弊病, 能够直接从原始输入数据中学习数据特征、理解复杂时空条件, 既能节约人力与时间成本, 又能更好地适应多变的交通环境[83], 因此未来自动驾驶的模型架构更有可能使用端到端学习. 综上所述, 本节对自动驾驶汽车中模型安全的分析, 将重点关注汽车运行时共性安全问题, 不再区分感知、规划和控制等不同模块.

以数据驱动的视角来看, 自动驾驶汽车驾驶模型往往具有训练、执行、更新 3 个阶段, 如图 5 所示. 云端使用训练集和车辆运行中采集作为样本进行训练得到神经网络模型, 并将模型部署到车辆中, 使车辆辆能够在实际驾驶中执行障碍物识别, 物体分割, 车道检测, 轨迹预测, 行为决策等任务, 同时在运行中持续采集相关数据, 进行新一轮训练. 本节梳理了 3 个阶段所面临的安全威胁和攻击方式, 如表 5 所示.

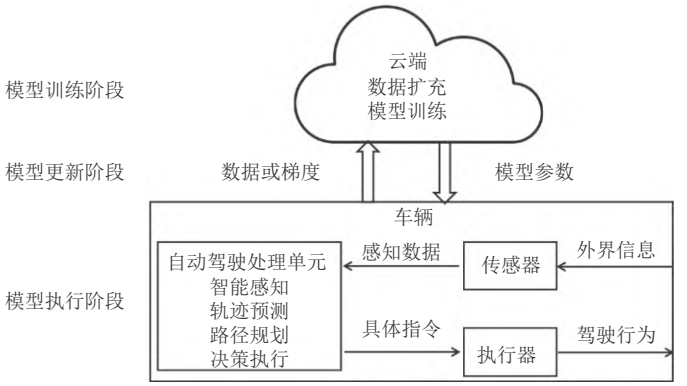

图 5 驾驶模型运行各阶段攻击面

5.1 驾驶模型训练面临威胁: 训练样本

驾驶模型训练阶段是将样本数据输入模型以调整其参数, 从而使神经网络学习特征与结论的潜在映射关系, 增强其判断准确性的过程. 模型训练阶段的攻击主要发生在训练样本 (攻击作用点 G) 上. 自动驾驶汽车在模型训

练阶段中使用的训练样本有两种主要来源.

1) 公开数据集, 如谷歌公司的 Waymo 开放数据集, 其拥有在超过 25 个城市的公共道路上行驶超过 2 000 万英里的真实数据, 以及超过百亿英里的模拟行驶数据.

2) 商业数据集, 如各厂商在车辆测试过程中采集的数据, 这种数据采集方法的优点是可以自定义数据特征, 情景与采样方式, 例如特斯拉提出的影子驾驶模式.

表 5 自动驾驶模型各阶段面临的安全威胁

阶段	安全威胁	攻击方式	文献
模型训练阶段	恶意样本	数据中毒攻击	[8,84–87]
模型执行阶段	模型误差	对抗性攻击	[9,88,89]
	模型篡改	模型扰动攻击	[90–92]
		模型扩展攻击	[93–95]
	模型滥用	模型提取攻击	[96,97]
		成员推理攻击	[98]
		模型逆向攻击	[99]
模型更新阶段	梯度或数据泄露	梯度泄露攻击	[100]

攻击者可以通过向这些来自车辆或云端的训练样本中注入少量精心设计的恶意数据, 达到污染模型的目的, 使得模型在一定条件下对特定输入给出错误输出, 这种攻击叫作数据中毒攻击. 在自动驾驶汽车中, 攻击者可以通过将恶意构造的有害样例上传到公开数据集中[8], 或在具有影子驾驶模式的车辆的运行中构造特定情景, 来实现对自动驾驶汽车模型的数据中毒攻击, 以降低模型的精度或泛化性, 使得汽车产生事故的概率增加. 根据攻击者是否对恶意样本标签进行篡改, 数据中毒攻击通常分为脏标签攻击与干净标签攻击两种.

脏标签攻击通过篡改特定数据标签, 使得特定类型的样本分类结果与实际情形完全相反, 从而降低模型的精度. 由于数据标签异常的样本容易被人类察觉, 因此现有的脏标签攻击工作主要集中在利用神经网络生成尽可能难以识别的脏标签样本, 以及增强样本的可迁移性上. 如文献 [84] 基于后向梯度的优化算法来生成中毒样本; 文献 [85] 提出一种靶向中毒攻击, 使得错误标签样本尽可能接近原始样本, 以达到增强隐蔽性的目的.

干净标签攻击通过将标签正确但特征异常的样本添加进数据集, 导致训练的模型可能对目标样本错误分类, 使得模型精度下降. 由于添加的样本不存在标签错误, 并且模型的精度下降只表现在特定类型的样本, 因此干净标签攻击难以通过人工筛查样本的方式识别, 如文献 [86] 利用特征碰撞来实现数据中毒攻击, 使得样本虽然在像素空间上接近原始样本, 但在特征空间上接近目标样本; 在此基础上, 文献 [87] 提出一种优化中毒样本构造的方法, 通过在特征空间内形成一个凸多面体包裹目标样本, 以增强攻击的不可感知性与攻击的可转移性.

总的来说, 数据中毒攻击可以从源头对模型进行破坏, 攻击面广, 危害性大. 攻击者通过大规模散发有毒样本, 使得所有自动驾驶汽车在训练车载模型时均存在使用有毒样本的可能, 从而在车载模型中设置后门, 实现对汽车的远程破坏.

5.2 驾驶模型执行面临威胁: 执行环境

模型执行阶段是将训练好的驾驶模型部署至自动驾驶汽车中后, 利用感知数据实现路径规划、决策控制等任务, 直至下一次模型更新的过程. 模型在执行阶段面临更为复杂的威胁, 这些威胁主要可以分为 3 类.

1) 对抗样本攻击: 由于有限的模型训练样本无法穷尽复杂的真实路况, 模型本身不可避免地存在极少量判断失误的情况, 攻击者就会通过构造一些有针对性的错误数据让车载模型做出误判, 如攻击者可以使用添加扰动[88]或生成对抗网络[9]的方式构造出导致这些失误的样本并进行利用, 以干扰模型的正常执行流程, 从而引发自动驾驶汽车的安全事故. 如文献 [89] 在自动驾驶领域复现了 5 种对抗性攻击方法, 证明目前的生成对抗网络可以应用于自动驾驶模型的回归与分类任务, 但对黑盒模型表现不佳. 对抗样本攻击在语义欺骗 (第 3.4 节) 中应用广泛.

2) 模型篡改攻击: 自动驾驶汽车内部用于模型执行的张量处理单元、图形处理单元、应用特定集成电路和现场可编程门阵列互相协作, 形成了一个新的中央操作单元[101]. 中央操作单元作为模型的执行环境 (攻击作用点 H),

能够高效地完成模型的具体任务. 然而, 与车载操作系统类似, 这样的中央操作单元在模型执行阶段中可能会受到非法访问[5], 进而使得正确部署的模型在实际运行中得到错误的结果. 模型篡改攻击一般表现为模型扰动攻击和模型扩展攻击[93]. 其中模型扰动攻击通过对模型二进制数据进行篡改, 导致模型神经元权重变化, 进而对总体模型产生精度影响[90]. 如文献 [91] 生成一组加权扰动并不断迭代以寻找最佳扰动和相应神经元, 并在提权后将扰动添加到网络中; 文献 [92] 利用恶意软件来修改受害者进程地址空间中的模型神经元权重, 从而造成模型的更改. 模型扩展攻击通过对模型结构进行特定修改, 实现神经网络的后门植入, 从而令受害模型对特定输入输出错误结果[102]. 如文献 [94] 将精心制作的木马网络与受害者网络合并, 达到修改最终预测结果的目的; 文献 [95] 将 DNN 模型的二进制文件分解为一个数据流图, 在数据流图中添加木马子图以识别特定样本, 最后将它重新编译为一个带有后门的模型.

3) 模型滥用攻击: 为了保护模型与数据的机密性, 自动驾驶汽车厂商并不对外公布模型实现与训练数据的具体细节, 相关模型 API 也不对外开放. 但如果攻击者在执行环境中获取了任意调用车内模型 API 的权限, 就可以通过特定的模型参数输入与调用序列, 分析出模型本身信息与训练集[103], 从而降低对抗性样本生成等攻击方式的实现难度. 模型滥用可能导致模型提取攻击、成员推理攻击与模型逆向攻击. 模型提取攻击指通过循环发送数据并查看对应的响应结果, 来推测机器学习模型的参数或功能, 复制出一个功能相似的机器学习模型[96]. 如文献 [97] 使用对抗训练训练一个逼近受害者模型的模型来近似获取模型信息. 成员推理攻击指通过分析黑盒模型对给定数据样本的输出, 以判断该样本是否在模型的训练集中. 如文献 [98] 在模型提取攻击成功的前提下, 将提取的模型使用不同的数据进行训练, 并将样本分别输入原始模型与提取出的模型, 通过分析置信度的方式来判定样本是否处于原始模型的训练集中. 模型逆向攻击指在得到模型初步信息与调用权限的前提下, 通过逆向分析模型得到训练数据集的统计特征. 如文献 [99] 在仅拥有数据标签的情况下, 通过最大化模型在目标预测值上的置信度获得模型训练数据的平均值.

就总体而言, 自动驾驶在模型执行阶段可能受到的攻击种类繁多, 角度多样. 目前在自动驾驶汽车上已经实现的攻击主要集中在对抗样本攻击上, 而模型篡改攻击与模型滥用攻击的具体应用较少. 但相对于对抗样本攻击, 后两者可以实现对汽车的完全控制与数据窃取, 亟待学者的进一步研究.

5.3 驾驶模型更新面临威胁: 更新信息

模型更新阶段是自动驾驶汽车与云端进行行驶记录或模型相关参数的交互, 对模型进行更新优化的过程. 传统汽车更新仅需要以远程升级 (over-the-air, OTA) 的方式下载对应升级包, 只要保证 OTA 通道的安全性, 就可以保证更新的安全性.

自动驾驶汽车模型更新过程有其特殊性, 需要向云端提供运行数据, 以实现车载模型再训练的精确性和泛化性, OTA 方式并不能满足这一要求. 为了适应这一特点, 特斯拉等厂商将联邦学习应用在自动驾驶汽车模型更新过程. 联邦学习将模型训练过程分散到多个本地设备上进行, 每个设备只训练自己本地的数据, 然后将模型参数聚合起来形成全局模型[104].

联邦学习一定程度上避免了中央服务器集中处理大量敏感数据所带来的数据直接泄露的风险, 但是由于汽车与云端不定期进行信息交互, 攻击者可以通过获取模型再训练时车云之间交互的更新信息 (攻击作用点 I), 窃取目标汽车或其他自动驾驶汽车的模型数据, 从而给自动驾驶汽车引入新的攻击威胁.

首先自动驾驶汽车向云端发送模型梯度或其他数据, 云端向自动驾驶汽车返回再训练后更新的模型参数, 这一过程有可能被攻击者劫持或监听, 导致模型梯度的泄露, 从而攻击者可以推理出车辆运行数据, 如行驶图像, 车辆位置等. 此外, 由于联邦学习通过梯度共享机制实现联合建模, 攻击者可以学习梯度差异来窃取其他车辆模型数据, 如文献 [100] 在整个模型更新梯度的过程中通过不断减少梯度差异, 反推更新输入样本和标签信息, 从而实现对其他自动驾驶汽车模型数据窃取.

随着未来自动驾驶汽车的普及以及机构与个人对于敏感数据的愈发重视, 联邦学习将成为自动驾驶汽车模型再训练的首选技术, 而其引入的模型更新威胁也将得到研究者更多的关注.

5.4 防护措施

驾驶模型安全防护的重点在于保障模型的正确性, 从而根据环境条件正确进行安全驾驶决策, 此外模型的完整性和数据也需要得到关注. 本节从驾驶模型的全阶段讨论其防护措施.

在模型训练阶段, 主要可以通过数据清洗与增强模型鲁棒性来提高模型训练阶段的安全性. 如文献 [105] 通过 KNN 算法对训练样本做异常检测, 去除过度偏离均值的样本, 使得模型训练不会过拟合, 达到提升模型正确性的目的. 文献 [106] 通过训练多个模型的方式提升模型的鲁棒性, 利用聚类后的不同样本集训练不同的模型, 在分类时使用投票的方式决定最终结果, 从而使得最终模型能够在各类情况下具有良好的适用性.

在模型执行阶段, 除了主动增强模型的鲁棒性与正确性外, 模型执行阶段应该着重部署内置防御机制, 如设计安全车载防火墙[107]、异常行为检测系统[108]等安全防护措施, 以防止攻击者对模型进行改动与窃听. 同时应严格控制车辆各 API 端口的权限管理, 避免自动驾驶功能相关的 API 遭受恶意调用, 造成数据泄露等安全问题.

在模型更新阶段, 应该重点保护模型数据. 一是在保证模型正确性的基础上, 对传递的梯度或数据进行随机扰动. 文献 [109] 使用差分隐私的方式, 向梯度参数中添加噪声完成混淆, 从而阻止攻击者获得训练数据. 文献 [110] 让联邦学习参与者通过添加局部随机扰动来将梯度信息隐藏, 进而屏蔽训练数据. 二是通过数据加密的方式使模型本身数据难以被解读, 如文献 [111] 利用同态加密技术, 重写算法中的多种激活函数, 同时利用密码单元实现密文计算, 以达到保护自动驾驶汽车的数据的目的. 三是增强联邦学习的安全性, 利用安全的联邦学习来辅助驾驶模型的训练和分发, 文献 [112] 对基于联邦学习的物联网系统的威胁进行了分析, 设计了基于联邦学习的物联网安全保护框架.

此外, 对自动驾驶模型的仿真测试也常用来保障驾驶模型的安全性. 按照测试场景生成方法的不同, 可以分为两类: 一是基于场景数据库的测试, 通过利用已有场景数据库或基于数据库构造新的场景对自动驾驶模型进行仿真测试, 如文献 [113] 将得到的一系列场景直接进行测试, 为了降低测试开销, 文献 [114] 引入场景评估模型对场景进行预筛选; 二是基于搜索的场景生成式测试, 通过遗传算法等启发式搜索方式在场景空间中搜索可能的致错场景, 并通过迭代其参数增强这些场景的致错能力以实现对自动驾驶模型的仿真测试, 如文献 [115] 基于遗传算法规定了场景变异和交叉的规则, 能够更快发现致错场景. 同时结合深度强化学习可以提高测试的效率和无偏性, 如文献 [116] 通过删除非安全关键状态并重新连接关键状态来增强马尔可夫决策过程, 从而学习到场景中具有安全关键事件的密集信息, 从而加速测试进程.

6 研究趋势与展望

为了宏观考察本领域研究现状与趋势, 本文以 CCF 推荐中文科技期刊目录与 CCF 推荐国际期刊与会议目录为目标刊物, 从 DBLP 数据库、中国知网数据库收集了自 2018 年到 2024 年 5 月以来 68 万余篇外文论文与 6 万余篇中文论文, 以关键字组合搜索 (关键字搜索表如表 6 所示) 的方式确定自动驾驶汽车安全相关研究论文总计 4218 篇, 其中外文论文 3846 篇 (占比 91.2%), 中文论文 372 篇 (占比 8.8%), 具体情况如图 6 所示.

表 6 自动驾驶汽车安全相关论文检索关键字

领域	关键词 (英文)	关键词 (中文)	论文总量 (英/中)
环境感知	sensor, camera, LiDAR, mm wave, ultrasonic radar, GPS, INS, UMU, GNSS, jamming, blind, damage, spoof, adversarial	传感器, 摄像头, 激光雷达, 毫米波, 超声波雷达, GPS, INS, UMU, GNSS, 干扰, 致盲, 破坏, 欺骗, 对抗	1606/208
V2X通信	in-vehicle networks, in-car wireless networks, vehicular cyber systems, V2X, V2V, V2R, V2I, V2N	车载网络, 车载无线网络, 车载网络系统, V2X, V2V, V2R, V2I, V2N	174/35
驾驶模型	model, network, machine learning, deep learning, federation learning, poison, Trojan, weight, perturbation, adversarial, steal, extraction, inference, inversion	模型, 神经网络, 机器学习, 深度学习, 联邦学习, 中毒, 木马, 权重, 扰动, 对抗, 窃取, 提取, 推理, 逆向	2066/129

从图 6(a) 中可以看出: 自 2018 年以来, 自动驾驶汽车安全相关研究呈现快速增长的态势, 体现了自动驾驶汽车安全日益受到学术界、工业界的重视. 具体到每一部分, 其增长趋势各有特点. 环境感知和 V2X 通信相关安全

研究保持平稳增长, 而驾驶模型的安全相关研究增长迅猛. 2020 年是一个特殊的时间节点, 在此之前, 由于感知是自动驾驶汽车的数据基础, 环境感知安全研究是自动驾驶汽车安全领域研究最多的部分, 而此后驾驶模型安全相关论文数量急剧增加, 逐年远超其他两部分. 出现这一现象可能有两方面的原因, 一是 2019 年 Transformer、BERT、GPT 等预训练模型相继提出, 刺激了深度学习的进一步发展; 二是深度学习模型在自动驾驶汽车中应用的深度和广度进一步提高, 成为自动驾驶汽车的控制核心, 随着端到端自动驾驶的飞速发展, One Model 方案得到越来越多的认可, 因此研究者也愈加关注驾驶模型的安全问题. 此外 V2X 通信安全相关研究相比来看一直较少, 需要得到更多关注.

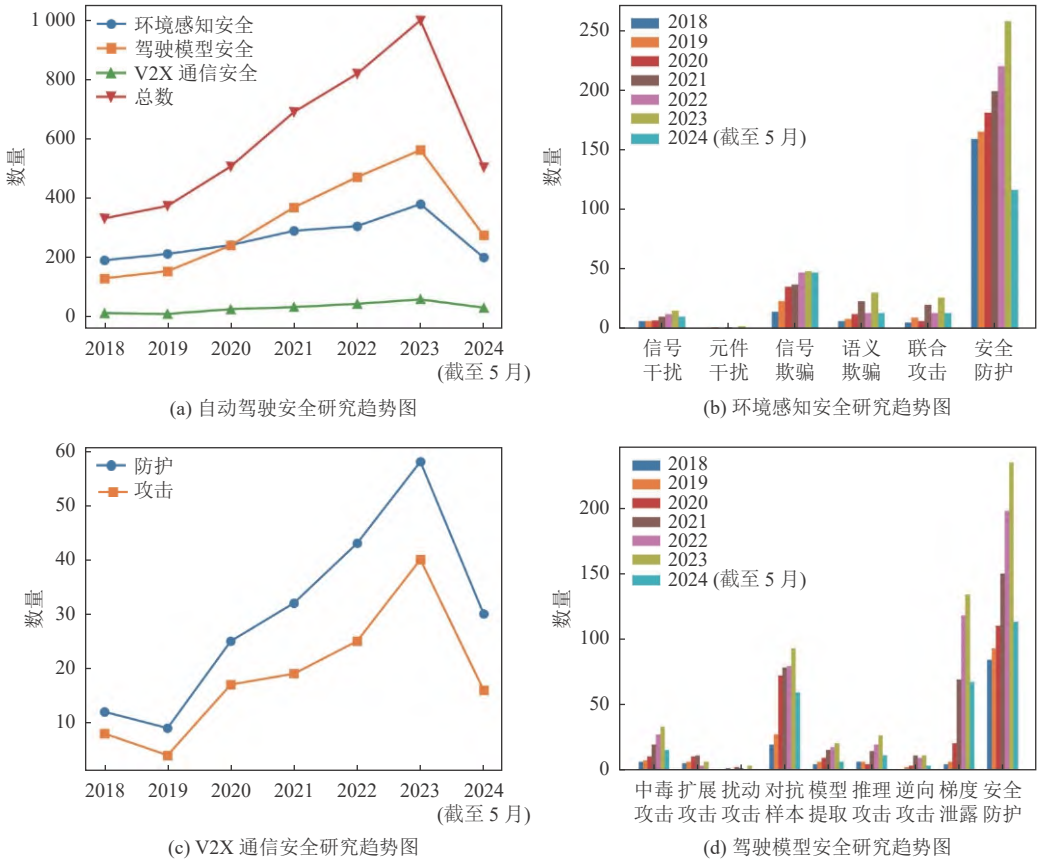

图 6 2018–2024 年自动驾驶汽车安全研究统计分析

从图 6(b) 中可以看出: 首先信号欺骗是环境感知面临最多的威胁, 一方面是由于信号欺骗相关技术比较成熟, 另一方面则是因为信号欺骗实现简单; 其次元件干扰相关攻击极少, 通过对相关论文进行分析, 本文发现元件脆弱性挖掘通常需要更广的知识背景且迁移性较差, 可能导致了相关研究数量不多; 此外, 语义欺骗、联合攻击相关研究方兴未艾, 出现了许多有影响力的成果, 逐渐成为环境感知安全研究的新热点; 最后从安全防护上看, 相关研究数量远大于对攻击的研究, 体现了在传感器大量应用的背景下, 学界和工业界对增强感知的可靠性和安全性的重视程度.

从图 6(c) 中可以看出: 在攻防视角下, 近年来针对 V2X 通信防御的研究热度一直高于对其攻击的研究热度, 且差距逐年增加. 此外无论攻击还是防御, 相关研究均呈现逐年快速增加趋势, 这表明随着自动驾驶汽车网联功能的日益丰富, V2X 通信将进一步发挥更大作用, 引起更多研究者注意, 从而进一步刺激相关安全研究的快速发展.

从图 6(d) 中可以看出: 从 2018 年以来, 对于模型安全的相关研究主要呈现增长态势. 首先模型对抗攻击、数

据中毒攻击、模型逆向攻击作为传统的研究热点, 论文数量逐年稳步增加, 愈加丰富的攻击场景和方案被设计和实现; 其次联邦学习作为深度学习新兴领域, 其在更新阶段引入的安全威胁引发大量关注, 论文数量指数性增长, 引发了各界对联邦学习在自动驾驶汽车中安全应用的担忧, 同时也激发了研究者对模型安全防护研究的热情; 此外模型扰动攻击, 模型扩展攻击的论文较少, 这可能是因为该领域要求研究者具备深度学习、网络安全和软件工程等多维度的知识储备.

综上所述, 自动驾驶汽车攻击与防御处于动态演进的阶段, 相关研究数量大致持平. 未来一段时间, 针对驾驶模型的攻击与防护依旧是热点. 并且随着自动驾驶汽车进一步普及、安全研究向细粒度发展, 感知语义欺骗攻击、联合攻击和 V2X 通信攻击等研究及对应的安全防护研究将引起越来越多研究者关注.

结合前文所述智能网联汽车自动驾驶安全最新研究进展与发展趋势, 为实现更安全的自动驾驶汽车, 本文有 4 个方面的展望.

1) 环境感知方面: 传感器是自动驾驶汽车物理域和信息域的媒介, 使得自动驾驶汽车感知需要尤为注意跨域的安全风险, 同时由于传感器本身元件实现所带来的带外脆弱性, 使得以传感器为核心构建的感知面临多样复杂的威胁, 因此平衡未来的感知能力需求与安全需求, 可从 2 个角度考虑对现有感知模块进行增强: ① 设计精度更高、功能更全面的传感器, 减少车载传感器数量, 从源头阻断传感器误差、噪声等所引发的各类攻击; ② 设计并实现更安全的 MSF 算法, 保障各类传感器以合理权重共同参与环境语义生成.

2) 汽车通信方面: 随着自动驾驶汽车连通性进一步加强, 汽车内外通信协议将面临颠覆性改变: ① 以功能为导向设计实现的车载总线协议 (如 CAN、LIN 等) 将难以符合自动驾驶汽车广泛连接的新特点, 因此需要研究大带宽、抗辐射、高可靠的车载通信协议, 并内置必要的安全策略; ② V2X 通信连接不稳定、网络拓扑变化快速, 一直制约自动驾驶汽车网联功能的安全性、可靠性, 因此需要设计新的传输协议适应车外通信的动态特征, 发展新的转发算法来解决移动通信中的安全问题, 并持续提升异常检测的精度.

3) 驾驶模型方面: 自动驾驶汽车对模型的正确率、鲁棒性和实时性有着极高的要求, 现有的神经网络框架仍然存在正确率不足、鲁棒性差、效率低等问题, 且汽车内部缺乏对内置模型的防篡改和攻击检测措施. 因此, 研究人员需要从以下 2 个方面增强自动驾驶模型的安全性: ① 关注自动驾驶模型的安全性和正确性, 研究自动驾驶汽车端到端安全的仿真测试技术以及基于大模型的安全赋能技术; ② 内置模型运行安全保障机制, 可以抵御窃取、篡改等威胁, 且满足实时性要求.

4) 汽车系统架构方面: 目前自动驾驶汽车依旧沿用传统汽车电子电气架构, ECU 分散独立负责单一功能, 每增加一项新功能都需要一个新的编程 ECU, 既带来巨大成本压力, 又由于更多的软件代码而降低了汽车的安全性和可靠性. 随着自动驾驶汽车功能愈加丰富, 要求部分 ECU 实现从简单控制机械元件运行到完成汽车智能化的功能扩展, 具备图像数据处理、边缘计算、云端通信等能力, 以辅助实现传感器的道路采集、本地 ECU 模型训练等自动驾驶任务. 面向未来, 随着自动驾驶汽车与外界连通性的增强和 ECU 的功能富集, ECU 将成为威胁自动驾驶汽车的新风险点, 因此需要变革现有汽车电子电气框架, 可从以下两点综合考虑: ① 完善基于 ECU 整合的分散式域控制器架构, 以应对大规模异构实时的汽车数据处理, 并支持远程更新; ② 建立纵深防御和重点防御结合的内生安全防护机制, 综合应用入侵检测、认证、加密等主被动安全技术, 分层建立关键 ECU 及其资源的等级保护机制, 保障自动驾驶任务不被恶意破坏或利用.

7 结 论

本文在模型视角下对智能网联汽车自动驾驶安全现有研究进行了系统的分析与归类, 从物信跨域、功能安全与信息安全融合等不同角度, 总结了自动驾驶汽车已经遭受或将要面临的安全威胁和对应策略. 目前对于自动驾驶汽车安全的研究仍不成熟, 存在诸多挑战, 希望本文可以为今后的自动驾驶汽车安全研究提供一个总结性的参考.

References:

[1] Zhang X, Sun H. Analysis of GB/T 40429–2021 “Classification of automotive driving automation”. China Automotive, 2022, (5): 3–5, 7

(in Chinese).

[2] Pham M, Xiong KQ. A survey on security attacks and defense techniques for connected and autonomous vehicles. Computers & Security, 2021, 109: 102269. [doi: 10.1016/j.cose.2021.102269]

[3] Humayed A, Lin JQ, Li FJ, Luo B. Cyber-physical systems security —A survey. IEEE Internet of Things Journal, 2017, 4(6): 1802–1831. [doi: 10.1109/JIOT.2017.2703172]

[4] Hataba M, Sherif A, Mahmoud M, Abdallah M, Alasmary W. Security and privacy issues in autonomous vehicles: A layer-based survey. IEEE Open Journal of the Communications Society, 2022, 3: 811–829. [doi: 10.1109/OJCOMS.2022.3169500]

[5] Ren K, Wang Q, Wang C, Qin Z, Lin XD. The security of autonomous driving: Threats, defenses, and future directions. Proc. of the IEEE, 2020, 108(2): 357–372. [doi: 10.1109/JPROC.2019.2948775]

[6] Jo HJ, Choi W. A survey of attacks on controller area networks and corresponding countermeasures. IEEE Trans. on Intelligent Transportation Systems, 2022, 23(7): 6123–6141. [doi: 10.1109/TITS.2021.3078740]

[7] Talpur A, Gurusamy M. Machine learning for security in vehicular networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 2022, 24(1): 346–379. [doi: 10.1109/COMST.2021.3129079]

[8] Chen YJ, Zhu XT, Gong XL, Yi XJ, Li SY. Data poisoning attacks in Internet-of-Vehicle networks: Taxonomy, state-of-the-art, and future directions. IEEE Trans. on Industrial Informatics, 2023, 19(1): 20–28. [doi: 10.1109/TII.2022.3198481]

[9] Deng Y, Zheng X, Zhang TY, Chen C, Lou GN, Kim M. An analysis of adversarial attacks and defenses on autonomous driving models. In: Proc. of the 2020 IEEE Int'l Conf. on Pervasive Computing and Communications (PerCom). Austin: IEEE, 2020. 1–10. [doi: 10.1109/PerCom45495.2020.9127389]

[10] Tang SC, Zhang ZY, Zhang Y, Zhou JX, Guo Y, Liu S, Guo SJ, Li YF, Ma L, Xue YX, Liu Y. A survey on automated driving system testing: Landscapes and trends. ACM Trans. on Software Engineering and Methodology, 2023, 32(5): 124. [doi: 10.1145/3579642]

[11] Cao YL, Xiao CW, Cyr B, Zhou YM, Park W, Rampazzi S, Chen QA, Fu K, Mao ZM. Adversarial sensor attack on LiDAR-based perception in autonomous driving. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security. London: ACM, 2019. 2267–2281. [doi: 10.1145/3319535.3339815]

[12] Cho KT, Kang GS. Error handling of in-vehicle networks makes them vulnerable. In: Proc. of the 2016 ACM SIGSAC Conf. on Computer and Communications Security. Vienna: ACM, 2016. 1044–1055. [doi: 10.1145/2976749.2978302]

[13] Burns A, McDermid J, Dobson J. On the meaning of safety and security. The Computer Journal, 1992, 35(1): 3–15. [doi: 10.1093/comjnl/35.1.3]

[14] Int'l Electrotechnical Commission. IEC/TS 62351-1 Power systems management and associated information exchange —Data and communications security. Part 1: Communication network and system security—Introduction to security issues. 2007. https://webstore.iec.ch/en/publication/6903

[15] Pirayesh H, Zeng HC. Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey. arXiv:2101.00292, 2021.

[16] He KX, Qin KJ, Wang CY, Fang XY. Research on cyber security test method for GNSS of intelligent connected vehicle. In: Proc. of the 2020 Int'l Conf. on Computer Information and Big Data Applications (CIBDA). Guiyang: IEEE, 2020. 200–203. [doi: 10.1109/CIBDA50819.2020.00052]

[17] Swinney CJ, Woods JC. GPS jamming signal classification with CNN feature extraction in low signal-to-noise environments. Int'l Journal on Cyber Situational Awareness, 2021, 6(1): 1–21. [doi: 10.22619/IJCSA.2021.100135]

[18] Elghamrawy H, Karaim M, Korenberg M, Noureldin A. High-resolution spectral estimation for continuous wave jamming mitigation of GNSS signals in autonomous vehicles. IEEE Trans. on Intelligent Transportation Systems, 2022, 23(7): 7881–7895. [doi: 10.1109/TITS.2021.3074102]

[19] Aydogdu C, Keskin MF, Carvajal GK, Eriksson O, Hellsten H, Herbertsson H, Nilsson E, Rydstrom M, Vanas K, Wymeersch H. Radar interference mitigation for automated driving: Exploring proactive strategies. IEEE Signal Processing Magazine, 2020, 37(4): 72–84. [doi: 10.1109/MSP.2020.2969319]

[20] Kunert M. The EU project MOSARIM: A general overview of project objectives and conducted work. In: Proc. of the 9th European Radar Conf. Amsterdam: IEEE, 2012. 1–5.

[21] Goppelt M, Blöcher HL, Menzel W. Analytical investigation of mutual interference between automotive FMCW radar sensors. In: Proc. of the 2011 German Microwave Conf. Darmstadt: IEEE, 2011. 1–4.

[22] Schmidt LM, Kontes G, Plinge A, Mutschler C. Can you trust your autonomous car? Interpretable and verifiably safe reinforcement learning. In: Proc. of the 2021 IEEE Intelligent Vehicles Symp. (IV). Nagoya: IEEE, 2021. 171–178. [doi: 10.1109/IV48863.2021.9575328]

[23] Bilik I, Longman O, Villeval S, Tabrikian J. The rise of radar for autonomous vehicles: Signal processing solutions and future research directions. IEEE Signal Processing Magazine, 2019, 36(5): 20–31. [doi: 10.1109/MSP.2019.2926573]

[24] Xu WY, Yan C, Jia WB, Ji XY, Liu JH. Analyzing and enhancing the security of ultrasonic sensors for autonomous vehicles. IEEE Internet of Things Journal, 2018, 5(6): 5015–5029. [doi: 10.1109/JIOT.2018.2867917]

[25] Petit J, Stottelaar B, Feiri M, Kargl F. Remote attacks on automated vehicles sensors: Experiments on camera and LiDAR. 2015. https://api.semanticscholar.org/CorpusID:39608826

[26] Yan C, Xu WY, Liu JH. Can you trust autonomous vehicles: Contactless attacks against sensors of self-driving vehicles. 2016. https://cyansec.com/files/articles/16DEFCON-Sensor.pdf

[27] Son Y, Shin H, Kim D, Park Y, Noh J, Choi K, Choi J, Kim Y. Rocking drones with intentional sound noise on gyroscopic sensors. In: Proc. of the 24th USENIX Security Symp. Washington: USENIX Association, 2015.

[28] Trippel T, Weisse O, Xu WY, Honeyman P, Fu K. WALNUT: Waging doubt on the integrity of MEMS accelerometers with acoustic injection attacks. In: Proc. of the 2017 IEEE European Symp. on Security and Privacy (EuroS&P). Paris: IEEE, 2017. 3–18. [doi: 10.1109/EuroSP.2017.42]

[29] Warner JS, Johnston RG. A simple demonstration that the global positioning system (GPS) is vulnerable to spoofing. The Journal of Security Administration, 2002, 25: 19–28.

[30] Volpe JA. Vulnerability assessment of the transportation infrastructure relying on the global positioning system. 2001. https://rosap.ntl.bts.gov/view/dot/8435

[31] Humphreys TE, Ledvina BM, Psiaki ML, O'Hanlon BW, Kintner PM Jr. Assessing the spoofing threat: Development of a portable GPS civilian spoofer. In: Proc. of the 21st Int'l Technical Meeting of the Satellite Division of the Institute of Navigation. Savanna, 2008.

[32] Nighswander T, Ledvina B, Diamond J, Brumley R, Brumley D. GPS software attacks. In: Proc. of the 2012 ACM Conf. on Computer and Communications Security. Raleigh: ACM, 2012. 450–461. [doi: 10.1145/2382196.2382245]

[33] Shin H, Kim D, Kwon Y, Kim Y. Illusion and dazzle: Adversarial optical channel exploits against lidars for automotive applications. In: Proc. of the 19th Int'l Conf. on Cryptographic Hardware and Embedded Systems. Taipei: Springer, 2017. 445–467.

[34] Mehmood S, Malik AN, Qureshi IM, Khan MZU, Zaman F. A novel deceptive jamming approach for hiding actual target and generating false targets. Wireless Communications and Mobile Computing, 2021, 2021: 8844630. [doi: 10.1155/2021/8844630]

[35] Buller W, Wilson B, Garbarino J, Kelly J, Subotic N, Thelen B, Belzowski B. Radar congestion study. Technical Report, DOT HS 812 632, Washington: U.S. Department of Transportation, National Highway Traffic Safety Administration, 2018.

[36] Alland S, Stark W, Ali M, Hegde M. Interference in automotive radar systems: Characteristics, mitigation techniques, and current and future research. IEEE Signal Processing Magazine, 2019, 36(5): 45–59. [doi: 10.1109/MSP.2019.2908214]

[37] Kim G, Mun J, Lee J. A peer-to-peer interference analysis for automotive chirp sequence radars. IEEE Trans. on Vehicular Technology, 2018, 67(9): 8110–8117. [doi: 10.1109/tvt.2018.2848898]

[38] Zeng KX, Liu SN, Shu YC, Wang D, Li HY, Dou YZ, Wang G, Yang YL. All your GPS are belong to us: Towards stealthy manipulation of road navigation systems. In: Proc. of the 27th USENIX Conf. on Security Symp. Baltimore: USENIX Association, 2018. 1527–1544.

[39] Zhu Y, Miao CL, Hajiaghajani F, Huai MD, Su L, Qiao CM. Adversarial attacks against LiDAR semantic segmentation in autonomous driving. In: Proc. of the 19th ACM Conf. on Embedded Networked Sensor Systems. Coimbra: Association for Computing Machinery, 2021. 329–342. [doi: 10.1145/3485730.3485935]

[40] Cao YL, Bhupathiraju SH, Naghavi P, Sugawara T, Mao ZM, Rampazzi S. You can't see me: Physical removal attacks on LiDAR-based autonomous vehicles driving frameworks. In: Proc. of the 32nd USENIX Security Symp. Anaheim: USENIX Association, 2023. 2993–3010.

[41] Ma C, Wang NF, Chen QA, Shen C. SlowTrack: Increasing the latency of camera-based perception in autonomous driving using adversarial examples. In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Vancouver: AAAI, 2024. 4062–4070. [doi: 10.1609/aaai.v38i5.28200]

[42] Zhu Y, Miao CL, Xue HF, Li ZX, Yu YN, Xu WY, Su L, Qiao CM. TileMask: A passive-reflection-based attack against mmWave radar object detection in autonomous driving. In: Proc. of the 2023 ACM SIGSAC Conf. on Computer and Communications Security. Copenhagen: ACM, 2023. 1317–1331. [doi: 10.1145/3576915.3616661]

[43] Cao YL, Wang NF, Xiao CW, Yang DW, Fang J, Yang RG, Chen QA, Liu MY, Li B. Invisible for both Camera and LiDAR: Security of multi-sensor fusion based perception in autonomous driving under physical-world attacks. In: Proc. of the 2021 IEEE Symp. on Security and Privacy (SP). San Francisco: IEEE, 2021. 176–194. [doi: 10.1109/SP40001.2021.00076]

[44] Shen JJ, Won JY, Chen ZY, Chen QA. Drift with devil: Security of multi-sensor fusion based localization in high-level autonomous

driving under GPS spoofing (extended version). In: Proc. of the 29th USENIX Security Symp. USENIX Association, 2020. 931–948.

[45] Narain S, Ranganathan A, Noubir G. Security of GPS/INS based on-road location tracking systems. arXiv:1808.03515, 2018.

[46] Shao BJ, Wan TQ, Liao FY, Kim BJ, Chen JW, Guo JM, Ma SJ, Ahn JH, Chai Y. Highly trustworthy in-sensor cryptography for image encryption and authentication. ACS Nano, 2023, 17(11): 10291–10299. [doi: 10.1021/acsnano.3c00487]

[47] Matsumura R, Sugawara T, Sakiyama K. A secure LiDAR with AES-based side-channel fingerprinting. In: Proc. of the 6th Int'l Symp. on Computing and Networking Workshops (CANDARW). Takayama: IEEE, 2018. 479–482. [doi: 10.1109/CANDARW.2018.00092]

[48] Dang YC, Benzaïd C, Yang B, Taleb T. Deep learning for GPS spoofing detection in cellular-enabled UAV systems. In: Proc. of the 2021 Int'l Conf. on Networking and Network Applications. Lijiang: IEEE, 2022. 501–506. [doi: 10.1109/NaNA53684.2021.00093]

[49] Kapoor P, Vora A, Kang KD. Detecting and mitigating spoofing attack against an automotive radar. In: Proc. of the 88th IEEE Vehicular Technology Conf. (VTC-Fall). Chicago: IEEE, 2018. 1–6. [doi: 10.1109/VTCFall.2018.8690734]

[50] Qayyum A, Usama M, Qadir J, Al-Fuqaha A. Securing connected & autonomous vehicles: Challenges posed by adversarial machine learning and the way forward. IEEE Communications Surveys & Tutorials, 2020, 22(2): 998–1026. [doi: 10.1109/COMST.2020.2975048]

[51] Lee S, Lee DH. From attack to identification: MEMS sensor fingerprinting using acoustic signals. IEEE Internet of Things Journal, 2022, 10(6): 5447–5460.

[52] Man YM, Muller R, Li M, Celik ZB, Gerdes R. That person moves like a car: Misclassification attack detection for autonomous systems using spatiotemporal consistency. In: Proc. of the 32nd USENIX Conf. on Security Symp. Anaheim: USENIX Association, 2023. 6929–6946.

[53] Hall DL, Llinas J. An introduction to multisensor data fusion. Proc. of the IEEE, 1997, 85(1): 6–23. [doi: 10.1109/5.554205]

[54] Chandrasekaran B, Gangadhar S, Conrad JM. A survey of multisensor fusion techniques, architectures and methodologies. In: Proc. of the 2017 Annual IEEE Region 3 Technical, Professional, and Student Conf. Concord: IEEE, 2017. 1–8. [doi: 10.1109/SECON.2017.7925311]

[55] Liu JS, Park J. "Seeing is not always believing": Detecting perception error attacks against autonomous vehicles. IEEE Trans. on Dependable and Secure Computing, 2021, 18(5): 2209–2223. [doi: 10.1109/TDSC.2021.3078111]

[56] Kai J, Schäfer M, Moser D, Lenders V, Pöpper C, Schmitt J. Crowd-GPS-Sec: Leveraging crowdsourcing to detect and localize GPS spoofing attacks. In: Proc. of the 2018 IEEE Symp. on Security and Privacy (SP). San Francisco: IEEE, 2018. 1018–1031. [doi: 10.1109/SP.2018.00012]

[57] Thilak KD, Amuthan A. DoS attack on VANET routing and possible defending solutions—A survey. In: Proc. of the 2016 Int'l Conf. on Information Communication and Embedded Systems (ICICES). Chennai: IEEE, 2016. 1–7. [doi: 10.1109/ICICES.2016.7518892]

[58] Petit J. Analysis of ECDSA authentication processing in VANETs. In: Proc. of the 3rd Int'l Conf. on New Technologies, Mobility and Security. Cairo: IEEE, 2009. 1–5. [doi: 10.1109/NTMS.2009.5384696]

[59] Kumar S, Mann KS. Prevention of DoS attacks by detection of multiple malicious nodes in VANETs. In: Proc. of the 2019 Int'l Conf. on Automation, Computational and Technology Management. London: IEEE, 2019. 89–94. [doi: 10.1109/ICACTM.2019.8776846]

[60] Patel KN, Jhaveri RH. Isolating packet dropping misbehavior in VANET using ant colony optimization. Int'l Journal of Computer Applications, 2015, 120(24): 5–9. [doi: 10.5120/21406-4161]

[61] Kamel J, Haidar F, Jemaa IB, Kaiser A, Lonc B, Urien P. A misbehavior authority system for Sybil attack detection in C-ITS. In: Proc. of the 10th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conf. (UEMCON). New York: IEEE, 2019. 1117–1123. [doi: 10.1109/UEMCON47517.2019.8993045]

[62] Xu YY, Lei M, Li M, Zhao MJ, Hu B. A new anti-jamming strategy based on deep reinforcement learning for MANET. In: Proc. of the 89th IEEE Vehicular Technology Conf. (VTC2019-Spring). Kuala Lumpur: IEEE, 2019. 1–5. [doi: 10.1109/VTCSpring.2019.8746494]

[63] Narayanadoss AR, Truong-Huu T, Mohan PM, Gurusamy M. Crossfire attack detection using deep learning in software defined ITS networks. In: Proc. of the 89th IEEE Vehicular Technology Conf. (VTC2019-Spring). Kuala Lumpur: IEEE, 2019. 1–6. [doi: 10.1109/VTCSpring.2019.8746594]

[64] Gruebler A, McDonald-Maier KD, Alheeti KMA. An intrusion detection system against black hole attacks on the communication network of self-driving cars. In: Proc. of the 6th Int'l Conf. on Emerging Security Technologies (EST). Braunschweig: IEEE, 2015. 86–91. [doi: 10.1109/EST.2015.10]

[65] Ali S, Nand P, Tiwari S. Detection of wormhole attack in vehicular ad-hoc network over real map using machine learning approach with preventive scheme. Journal of Information Technology Management, 2022, 14: 159–179. [doi: 10.22059/jitm.2022.86658]

[66] Khanapuri E, Chintalapati T, Sharma R, Gerdes R. Learning-based adversarial agent detection and identification in cyber physical systems applied to autonomous vehicular platoon. In: Proc. of the 5th IEEE/ACM Int'l Workshop on Software Engineering for Smart

Cyber-physical Systems (SEsCPS). Montreal: IEEE, 2019. 39–45. [doi: 10.1109/SEsCPS.2019.00014]

[67] Sargolzaei A, Crane CD, Abbaspour A, Noei S. A machine learning approach for fault detection in vehicular cyber-physical systems. In: Proc. of the 15th IEEE Int'l Conf. on Machine Learning and Applications (ICMLA). Anaheim: IEEE, 2016. 636–640. [doi: 10.1109/ICMLA.2016.0112]

[68] Othmane LB, Weffers H, Mohamad MM, Wolf M. A survey of security and privacy in connected vehicles. In: Benhaddou D, Al-Fuqaha A, eds. Wireless Sensor and Mobile Ad-hoc Networks: Vehicular and Space Applications. New York: Springer, 2015. 217–247. [doi: 10.1007/978-1-4939-2468-4_10]

[69] Ali I, Lawrence T, Li FG. An efficient identity-based signature scheme without bilinear pairing for vehicle-to-vehicle communication in VANETs. Journal of Systems Architecture, 2020, 103: 101692. [doi: 10.1016/j.sysarc.2019.101692]

[70] Ali Alheeti KM, Gruebler A, McDonald-Maier K. Intelligent intrusion detection of grey hole and rushing attacks in self-driving vehicular networks. Computers, 2016, 5(3): 16. [doi: 10.3390/computers5030016]

[71] Lu XZ, Xiao L, Xu TW, Zhao YF, Tang YL, Zhuang WH. Reinforcement learning based PHY authentication for VANETs. IEEE Trans. on Vehicular Technology, 2020, 69(3): 3068–3079. [doi: 10.1109/TVT.2020.2967026]

[72] Gomides TS, Kranakis E, Lambadaris I, Viniotis Y. Optimal control for platooning in vehicular networks. In: Proc. of the 2023 IEEE Int'l Conf. on Communications. Rome: IEEE, 2023. 6597–6602. [doi: 10.1109/ICC45041.2023.10279610]

[73] Xu H, Ji JQ, Zhu K, Wang R. Deep reinforcement learning for resource allocation in multi-platoon vehicular networks. In: Proc. of the 16th Int'l Conf. on Wireless Algorithms, Systems, and Applications. Nanjing: Springer, 2021. 402–416. [doi: 10.1007/978-3-030-86130-8_32]

[74] Chang S, Qi Y, Zhu HZ, Zhao JZ, Shen XM. Footprint: Detecting Sybil attacks in urban vehicular networks. IEEE Trans. on Parallel and Distributed Systems, 2012, 23(6): 1103–1114. [doi: 10.1109/TPDS.2011.263]

[75] Lu RX, Lin XD, Liang XH, Shen XM. A dynamic privacy-preserving key management scheme for location-based services in VANETs. IEEE Trans. on Intelligent Transportation Systems, 2012, 13(1): 127–139. [doi: 10.1109/TITS.2011.2164068]

[76] Junaidi DR, Ma MD, Su R. Secure vehicular platoon management against Sybil attacks. Sensors, 2022, 22(22): 9000. [doi: 10.3390/s22229000]

[77] Gu PWL, Khatoun R, Begriche Y, Serhrouchni A. Support vector machine (SVM) based Sybil attack detection in vehicular networks. In: Proc. of the 2017 IEEE Wireless Communications and Networking Conf. (WCNC). San Francisco: IEEE, 2017. 1–6. [doi: 10.1109/WCNC.2017.7925783]

[78] Gong J, Murguia C, Bayuwindra A, Cao JD. Resilient controller synthesis against DoS attacks for vehicular platooning in spatial domain. arXiv:2307.15874, 2023.

[79] Ravindran R, Santora MJ, Jamali MM. Multi-object detection and tracking, based on DNN, for autonomous vehicles: A review. IEEE Sensors Journal, 2021, 21(5): 5668–5677. [doi: 10.1109/JSEN.2020.3041615]

[80] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84–90. [doi: 10.1145/3065386]

[81] Bijjahalli S, Sabatini R, Gardi A. Advances in intelligent and autonomous navigation systems for small UAS. Progress in Aerospace Sciences, 2020, 115: 100617. [doi: 10.1016/j.paerosci.2020.100617]

[82] Waymo LLC. Waymo safety report: On the road to fully self-driving. 2017. https://storage.googleapis.com/sdc-prod/v1/safety-report/waymo-safety-report-2017-10.pdf

[83] Bojarski M, Del Testa D, Dworakowski D, Firner B, Flepp B, Goyal P, Jackel LD, Monfort M, Muller U, Zhang JK, Zhang X, Zhao J, Zieba K. End to end learning for self-driving cars. arXiv:1604.07316, 2016.

[84] Muñoz-González L, Biggio B, Demontis A, Paudice A, Wongrassamee V, Lupu EC, Roli F. Towards poisoning of deep learning algorithms with back-gradient optimization. In: Proc. of the 10th ACM Workshop on Artificial Intelligence and Security. Dallas: ACM, 2017. 27–38. [doi: 10.1145/3128572.3140451]

[85] Suciu O, Mărginean R, Kaya Y, Daumé H III, Dumitraş T. When does machine learning FAIL? Generalized transferability for evasion and poisoning attacks. In: Proc. of the 27th USENIX Conf. on Security Symp. Baltimore: USENIX Association, 2018. 1299–1316.

[86] Shafahi A, Huang WR, Najibi M, Suciu O, Studer C, Dumitras T, Goldstein T. Poison frogs! Targeted clean-label poisoning attacks on neural networks. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 6106–6116.

[87] Zhu C, Huang WR, Li HD, Taylor G, Studer C, Goldstein T. Transferable clean-label poisoning attacks on deep neural nets. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 7614–7623.

[88] Szegedy C, Zaremba W, Sutskever I, Bruna J, Erhan D, Goodfellow I, Fergus R. Intriguing properties of neural networks.

arXiv:1312.6199, 2014.

[89] Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 2018, 35(1): 53–65. [doi: 10.1109/MSP.2017.2765202]

[90] Dumford J, Scheirer W. Backdooring convolutional neural networks via targeted weight perturbations. In: Proc. of the 2020 IEEE Int'l Joint Conf. on Biometrics (IJCB). Houston: IEEE, 2020. 1–9. [doi: 10.1109/IJCB48548.2020.9304875]

[91] Rudd EM, Rozsa A, Günther M, Boult TE. A survey of stealth malware attacks, mitigation measures, and steps toward autonomous open world solutions. IEEE Communications Surveys & Tutorials, 2017, 19(2): 1145–1172. [doi: 10.1109/COMST.2016.2636078]

[92] Costales R, Mao CZ, Norwitz R, Kim B, Yang JF. Live Trojan attacks on deep neural networks. In: Proc. of the 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020. 3460–3469. [doi: 10.1109/CVPRW50498.2020.00406]

[93] Zhang QX, Ma WC, Wang YJ, Zhang YY, Shi ZW, Li YZ. Backdoor attacks on image classification models in deep neural networks. Chinese Journal of Electronics, 2022, 31(2): 199–212. [doi: 10.1049/cje.2021.00.126]

[94] Tang RX, Du MN, Liu NH, Yang F, Hu X. An embarrassingly simple approach for Trojan attack in deep neural networks. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. ACM, 2020. 218–228. [doi: 10.1145/3394486.3403064]

[95] Li YC, Hua JY, Wang HY, Chen CY, Liu YX. DeepPayload: Black-box backdoor attack on deep learning models through neural payload injection. In: Proc. of the 43rd IEEE/ACM Int'l Conf. on Software Engineering (ICSE). Madrid: IEEE, 2021. 263–274. [doi: 10.1109/ICSE43902.2021.00035]

[96] Tramèr F, Zhang F, Juels A, Reiter MK, Ristenpart T. Stealing machine learning models via prediction APIs. In: Proc. of the 25th USENIX Conf. on Security Symp. Austin: USENIX Association, 2016. 601–618.

[97] Truong JB, Maini P, Walls RJ, Papernot N. Data-free model extraction. In: Proc. of the 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 4769–4778. [doi: 10.1109/CVPR46437.2021.00474]

[98] Shokri R, Stronati M, Song CZ, Shmatikov V. Membership inference attacks against machine learning models. In: Proc. of the 2017 IEEE Symp. on Security and Privacy (SP). San Jose: IEEE, 2017. 3–18. [doi: 10.1109/SP.2017.41]

[99] Fredrikson M, Jha S, Ristenpart T. Model inversion attacks that exploit confidence information and basic countermeasures. In: Proc. of the 22nd ACM SIGSAC Conf. on Computer and Communications Security. Denver: ACM, 2015. 1322–1333. [doi: 10.1145/2810103.2813677]

[100] Zhu LG, Liu ZJ, Han S. Deep leakage from gradients. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 14774–14784.

[101] Lin SC, Zhang YQ, Hsu CH, Skach M, Haque ME, Tang LJ, Mars J. The architectural implications of autonomous driving: Constraints and acceleration. In: Proc. of the 23rd Int'l Conf. on Architectural Support for Programming Languages and Operating Systems. Williamsburg: ACM, 2018. 751–766. [doi: 10.1145/3173162.3173191]

[102] Cheng ZY, Wu BY, Zhang ZY, Zhao JJ. TAT: Targeted backdoor attacks against visual object tracking. Pattern Recognition, 2023, 142: 109629. [doi: 10.1016/j.patcog.2023.109629]

[103] Zhang KY, Song X, Zhang CH, Yu S. Challenges and future directions of secure federated learning: A survey. Frontiers of Computer Science, 2022, 16(5): 165817. [doi: 10.1007/s11704-021-0598-z]

[104] McMahan B, Moore E, Ramage D, Hampson S, Aguera y Arcas B. Communication-efficient learning of deep networks from decentralized data. In: Proc. of the 20th Int'l Conf. on Artificial Intelligence and Statistics. Fort Lauderdale: PMLR, 2017. 1273–1282.

[105] Peri N, Gupta N, Huang WR, Fowl L, Zhu C, Feizi S, Goldstein T, Dickerson JP. Deep K-NN defense against clean-label data poisoning attacks. In: Proc. of the 16th European Conf. on Computer Vision. Glasgow: Springer, 2020. 55–70. [doi: 10.1007/978-3-030-66415-2_4]

[106] Rosenfeld E, Winston E, Ravikumar P, Kolter JZ. Certified robustness to label-flipping attacks via randomized smoothing. In: Proc. of the 37th Int'l Conf. on Machine Learning. Virtual Event: JMLR.org, 2020. 8230–8241.

[107] Xiao P, Li YY, Li XH. Design and implementation of firewall based on MOST. Science and Technology & Innovation, 2009, 25(21): 57–58, 61 (in Chinese with English abstract). [doi: 10.3969/j.issn.1008-0570.2009.21.024]

[108] Wu YH. Research on vehicle CAN network intrusion detection system based on neural networks [MS. Thesis]. Chengdu: Chengdu University of Information Engineering, 2018 (in Chinese).

[109] Wei K, Li J, Ding M, Ma C, Yang HH, Farokhi F, Jin S, Quek TQS, Vincent Poor H. Federated learning with differential privacy: Algorithms and performance analysis. IEEE Trans. on Information Forensics and Security, 2020, 15: 3454–3469. [doi: 10.1109/TIFS.2020.2988575]

[110] Wang JX, Guo S, Xie X, Qi H. Protect privacy from gradient leakage attack in federated learning. In: Proc. of the 2022 IEEE Conf. on Computer Communications. London: IEEE, 2022. 580–589. [doi: 10.1109/INFOCOM48880.2022.9796841]

[111] Phong LT, Aono Y, Hayashi T, Wang LH, Moriai S. Privacy-preserving deep learning via additively homomorphic encryption. IEEE Trans. on Information Forensics and Security, 2018, 13(5): 1333–1345. [doi: 10.1109/TIFS.2017.2787987]

[112] Manzoor SI, Jain S, Singh Y, Singh H. Federated learning based privacy ensured sensor communication in IoT networks: A taxonomy, threats and attacks. IEEE Access, 2023, 11: 42248–42275. [doi: 10.1109/ACCESS.2023.3269880]

[113] Bolte JA, Bar A, Lipinski D, Fingscheidt T. Towards corner case detection for autonomous driving. In: Proc. of the 2019 IEEE Intelligent Vehicles Symp. (IV). Paris: IEEE, 2019. 438–445. [doi: 10.1109/IVS.2019.8813817]

[114] Klischat M, Liu EI, Holtke F, Althoff M. Scenario factory: Creating safety-critical traffic scenarios for automated vehicles. In: Proc. of the 23rd IEEE Int'l Conf. on Intelligent Transportation Systems (ITSC). Rhodes: IEEE, 2020. 1–7. [doi: 10.1109/ITSC45102.2020.9294629]

[115] Kim S, Liu M, Rhee JJ, Jeon Y, Kwon Y, Kim CH. DriveFuzz: Discovering autonomous driving bugs through driving quality-guided fuzzing. In: Proc. of the 2022 ACM SIGSAC Conf. on Computer and Communications Security. Los Angeles: ACM, 2022. 1753–1767. [doi: 10.1145/3548606.3560558]

[116] Feng S, Sun HW, Yan XT, Zhu HJ, Zou ZX, Shen SY, Liu HX. Dense reinforcement learning for safety validation of autonomous vehicles. Nature, 2023, 615(7953): 620–627. [doi: 10.1038/s41586-023-05732-2]

附中文参考文献:

[1] 张行, 孙航. GB/T 40429—2021《汽车驾驶自动化分级》分析. 中国汽车, 2022, (5): 3–5, 7.

[107] 肖鹏, 李媛媛, 李晓红. 车载 MOST 网络防火墙的设计与实现. 微计算机信息, 2009, 25(21): 57–58, 61. [doi: 10.3969/j.issn.1008-0570.2009.21.024]

[108] 吴贻淮. 基于神经网络的车载 CAN 网络入侵检测系统的研究 [硕士学位论文]. 成都: 成都信息工程大学, 2018.

郗来乐(1998－), 男, 硕士生, 主要研究领域为智能网联汽车安全, 物联网安全.

孙玉砚(1982－), 男, 博士, 高级工程师, 主要研究领域为工控安全, 物联网安全, 车联网.

林声浩(1999－), 男, 博士生, 主要研究领域为车联网安全, 自动驾驶安全.

朱红松(1973－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为物联网安全, 物联网安全, 网络空间安全测量, 智能安全.

王震(1999－), 男, 博士生, 主要研究领域为车联网, 深度强化学习.

孙利民(1966－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为物联网及其安全, 工业控制系统安全, 网络空间安全.

谢天鸽(1994－), 男, 博士生, 主要研究领域为车联网安全, 网络空间安全.