深圳国际金融科技大赛
技术文档【人工智能赛道】

［项目名称］

基于联邦学习的车联网攻击类型检测

［2023 年 12 月 17 日］

［团队名称］

卷不动队

目录

1 项目背景

1.1 需求和现状

1.1.1 联邦学习

随着互联网、物联网和边缘计算的蓬勃发展，数据越来越分散地存储在不同的设备和地点。在这个背景下，个人数据透明化引发了人们对隐私保护的日益增长的关切。不少行业和应用对如何处理敏感数据提出了更高的标准。传统的集中式学习方法在这样的环境下面临着数据传输和隐私保护等挑战。其次，随着各行各业数据规模的逐渐增大和维度的提高，有些场景下将所有数据传输到中央服务器进行模型训练会导致巨大的通信开销。此外，一些应用对模型的实时性要求日益提高，例如自动驾驶、智能驾驶辅助等。传统的集中式学习方法已经不能满足当前需求，于是联邦学习作为一种分布式学习算法框架应运而生，受到了学术界的广泛关注。

联邦学习本质上是一种分布式学习的算法框架，它允许分布在不同计算节点的设备之间在不共享原始数据的前提下协同训练模型。在联邦学习中，模型的训练仅在本地设备上进行，只有模型的更新参数被传输至主服务器，从而大大降低了隐私泄漏的风险，同时，也减少了对中央服务器的通信需求，有效降低了企业的成本。联邦学习的另一优势是允许模型在本地实时更新，以适应动态环境，为对实时性要求较高的技术

提供了可靠的支持，如自动驾驶、无人机等，这使得联邦学习成为当前需求下的一种先进而可行的解决方案。

1.1.2 车联网

车联网是一种基于移动车辆之间通信的网络，由车内网络、车间网络和车载移动互联网三者融合构建[1]，被定义为“人-车-路-云”之间通信连接的大规模分布式系统，车就是车辆，路就是路测基站、云就是云服务端。车联网最开始于 20 世纪 90 年代初被提出，最初关注车辆之间的通信，然而由于当时通信技术和计算能力的限制，车联网的发展受到了一定的制约。

近年来，随着移动互联网的普及，尤其是汽车行业智能化趋势的深入，搭载网联终端的汽车数量逐渐增多，网联汽车规模预计将会突破 9000 亿元。车联网因此进入了新的发展阶段，汽车行业在不断发展的同时逐步朝着数字化和信息化方向迈进。此外，随着智能交通系统的兴起，自动驾驶逐渐普及，无人驾驶车辆通过车联网进行数据交互，实现了实时获取和分享车辆状态、周围环境、交通状况、紧急信息等数据的功能，进一步提升了行车的安全性和效率。

车联网的未来发展将受益于新兴技术的涌现。如边缘计算、人工智能等。这些技术将进一步推动车联网的智能化水平，促使交通系统向更为高效、安全、环保的方向迈进。这标志着车联网在未来将成为连接车辆、路况、云端的关键驱动力，推动整个交通生态系统向更智能的方向演进。

1.1.3 国内外研究现状

为了充分发挥车联网数据的潜在价值，研究者们致力于探索安全可行的数据共享方法。在这一背景下，联邦学习作为一种前沿的机器学习方法，为车联网安全性和隐私性提供了新的思路。研究者 Pokhrel[2]等人提出的基于联邦学习的车联网 TCP 性能改进方案，通过将车辆视为移动数据中心，局部交换输入、输出以及学习参数，实现了对数据的隐私保护，在提高车联网系统性能的同时，强调了对个人敏感信息的敏感性。这一方法的成功应用为车联网系统的整体性能提升树立了榜样。

车联网的进一步应用体现在自动驾驶方面，在该领域，联邦学习也得到了广泛尝试。研究者们不仅致力于通过本地传感器数据训练车辆的本地模型，更通过联邦学习的方式仅共享模型参数，从而在确保数据隐私的前提下提高整个自动驾驶系统的性能[3]。这一方法为车辆之间的知识共享提供了一种高效且安全的途径。

全球范围内，研究者们正在积极研究如何通过联邦学习来解决车联网数据共享的安全性和隐私性问题。涉及的研究方向包括网络通信、自动驾驶、车载学习模型等多个领域。这些研究成果为构建更加安全可靠的车联网系统提供了全球性的创新方案。通过扩展对国内外研究现状的阐述，我们更全面地了解到联邦学习在车联网安全领域的广泛应用，并为本项目提供了更多的启示和借鉴价值。

1.1.4 对抗性系统

对抗性系统能够在面对恶意行为、攻击或不利环境时做出适应性的反应，旨在识别、防御或应对各种对抗性威胁，保持其功能、安全性和性能。

表 1 对抗性系统

<table>
<tr><th rowspan="2">等级</th><th rowspan="2">名称</th><th rowspan="2">定义</th><th colspan="4">主体</th></tr>
<tr><th>对抗操作</th><th>感知判断</th><th>任务支援</th><th>作用域</th></tr>
<tr><td>L0</td><td>人工对抗</td><td>由防御者完全人工对抗攻击者</td><td rowspan="2">人类</td><td rowspan="3">人类</td><td rowspan="3">人类</td><td>无</td></tr>
<tr><td>L1</td><td>辅助对抗</td><td>由机器完成已知攻击的攻击检测和攻击防御，其余由人类操作</td><td rowspan="5">部分</td></tr>
<tr><td>L2</td><td>低度自主对抗</td><td>由机器完成已知攻击的攻击检测和攻击防御，并具备能感知未知威胁、感知误报、感知漏报中的一种，其余由人类操作</td><td rowspan="4">机器</td></tr>
<tr><td>L3</td><td>中度自主对抗</td><td>由机器完成所有的对抗操作（攻击检测、攻击防御、自主感知未知威胁、误报漏报自主感知、对抗）</td><td rowspan="3">机器</td></tr>
<tr><td>L4</td><td>高度自主对抗</td><td>由机器完成所有的对抗操作，根据系统要求，人类不一定提供所有的应答（中间过程非必须人类参与），只能作用于限定的特定的场景（如网络域、主机域等）</td><td>机器</td></tr>
<tr><td>L5</td><td>完全自主对抗</td><td>由机器完成所有的对抗操作，根据系统要求，人类不一定提供所有的应答，不限定特定的场景</td><td>全域</td></tr>
</table>

1.1.5 需求来源

车联网发展迅速，它的隐私保护特别是针对潜在的网络攻击，受到人们密切关注。为了应对这一挑战，车联网需要进行联邦学习，允许在多个本地数据源上进行模型训练，而无需将数据传输到中心服务器。这一特性

使得联邦学习成为处理隐私敏感数据的理想选择，同时也有助于提高模型的泛化性能。在车联网攻击检测中，这意味着车辆之间可以共享攻击信息，而无需泄露具体的车辆数据，从而构建更加健壮的安全模型。

首先，联邦学习通过车辆共同参与模型训练，扩大了训练数据规模。车辆在不同地点、时间和场景下运行，捕捉各种攻击模式，由此整合的来自多个车辆的数据，能够更全面地了解潜在威胁，生成更鲁棒的攻击检测模型。其次，联邦学习解决了车辆面临的数据不平衡问题。不同车辆可能面临不同类型的攻击，联邦学习使车辆共同学习各种攻击类型的信息，使模型更通用，适应不同攻击场景。另外，联邦学习还有助于提高模型的隐私性。在传统的集中式学习中，为了在中心服务器上进行模型训练，车辆需要将其数据传输到服务器，这可能涉及到敏感的位置信息、驾驶行为等隐私数据，而联邦学习将模型的训练推送到本地，只共享模型参数的更新，大大减少了隐私泄露的风险。

然而，尽管联邦学习在车联网攻击检测中有诸多优势，但也面临一些挑战。首先是通信开销的增加，由于需要在车辆之间传递模型参数更新，因此可能导致较大的通信负担。其次是模型聚合的问题，即如何合并来自不同车辆的模型参数更新，以保证整体模型的性能。这需要设计有效的聚合算法，以确保在模型更新过程中不丢失重要信息。

总的来说，联邦学习在车联网攻击检测中的应用为解决车辆安全性和隐私性之间的平衡提供了一种新的思路。通过充分利用分布式数据、降低

通信开销、提高模型泛化性能，联邦学习为车联网系统提供了更加可靠和健壮的安全保障。然而，随着技术的不断发展，对于联邦学习在车联网安全中的应用还需要进一步的研究和改进，以应对不断演变的网络威胁。

1.1.6 项目创意

我们的项目创意源于对车联网安全性和隐私性的深刻认识，旨在通过创新性的联邦学习方法，构建一套安全可靠、隐私友好的入侵检测系统。该系统将有效应对车联网系统面临的日益复杂的网络攻击和威胁，同时注重保护车主和车辆的隐私。

（1）创意亮点

①联邦学习实时更新：通过引入联邦学习，使车辆在本地进行模型训练，通过共享模型更新，实现入侵检测的实时改进。

②差异隐私技术保护隐私：结合差异隐私技术，确保在信息共享的过程中，每个车辆的个体隐私得到充分的保护。

③用户友好的界面：设计直观、用户友好的界面，让车联网系统管理员能够轻松监控系统状态，设置安全策略，并及时响应安全事件。

（2）创新性贡献

①实时入侵检测：以实时性为核心，突破传统入侵检测的时效性限制，为车联网系统提供更灵活、更及时的安全防护。

②隐私保护：将差异隐私技术融入联邦学习框架，为车主提供极高水

平的隐私保护，树立隐私友好型车联网系统的标杆。

③分层联邦学习：在车联网的实际应用中，我们将路测基站定位为边缘聚合节点，构建了一个巧妙的分层局部聚合中心节点体系。每辆车在这个系统中充当着独立的客户端，负责将生成的数据首先发送给就近的路测基站进行局部聚合。这种分层的局部聚合中心节点的设计充分借鉴了 FL[4] 和 VEC[5]的优化思想，使得整个系统在维持高效性能的同时更具灵活性。通过局部聚合，每个车辆在通信开销上得到了显著的减轻，尤其是避免了在整个系统范围内频繁进行云端模型聚合所带来的性能下降。随后，这些局部聚合的结果通过多轮迭代逐步传输到中心云端服务器，实现了任务处理的整体优化。这个设计方案不仅考虑了车辆任务的多样性，根据数据属性进行了智能化的局部处理，还充分考虑了车辆的分类，确保每辆车在任务处理上都能以最适合的方式参与系统运作。

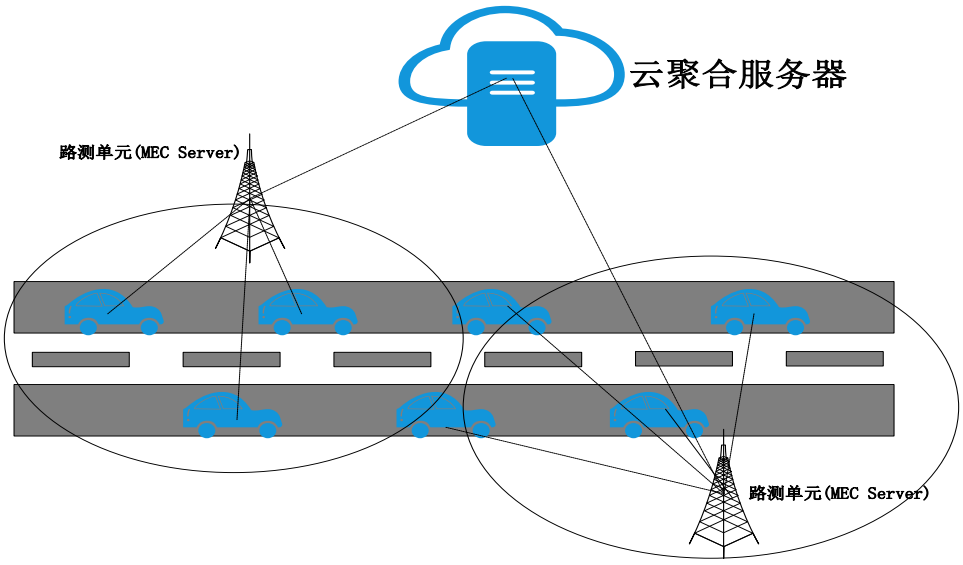

图 1 分层联邦学习

（3）创意背后的动机：

①社会责任：在数字化时代，车联网系统的安全性和隐私性已成为社会的重要问题。我们的项目致力于为车主和车辆提供更安全、更隐私保护的数字交通环境。

②推动技术发展：通过融合联邦学习、差异隐私等前沿技术，推动车联网安全领域的技术发展，为未来数字化交通社会打下坚实基础。

1.2 总体目标

1.2.1 用户界面

设计和实现一个直观、用户友好的界面，以便用户能够轻松地监控车联网系统的安全状态。用户界面应该提供以下功能：

（1）实时监控：显示车联网系统当前的安全状态，包括正常行为和检测到的潜在入侵类型。

（2）可视化展示：使用图表、图形或其他可视化元素清晰地呈现安全事件和模型的性能指标，使用户能够迅速理解系统的整体状况。

（3）用户交互：提供用户与系统交互的功能，例如设置安全策略、查看历史安全事件等。确保用户能够方便地与系统进行沟通和反馈。

1.2.2 初步验证

在车联网行业中进行初步验证，以确保开发的系统能够有效地检测和防范各种入侵类型，并满足实际应用的需求。

（1）性能评估：对系统进行全面的性能评估，包括准确性、实时性和鲁棒性。使用真实世界的数据集进行测试，模拟各种场景。

（2）实际应用测试：在仿真车联网环境中进行测试，观察系统在实际使用情况下的表现。可以包括在车辆之间进行通信时的网络入侵、恶意代码注入等。

（3）用户反馈：收集系统使用者的反馈，了解他们对系统的满意度以及可能存在的改进点。

（4）与行业标准对比：将系统的性能与车联网安全领域的行业标准进行比较，以确保系统在安全性和性能方面达到或超过行业要求。

1.3 所需软硬件条件

表 2 软硬件条件表

名称	配置
操作系统	Ubuntu22.04
CPU	InterCorei9-7920XCPU@2.90GHZ
显卡	NVIDIAGeForceRTX3090Ti
内存大小	64G
硬盘大小	512G
仿真开发工具	Pycharm、OMNeT++、SUMO、Veins
开发平台	FATE

2 系统分析

2.1 需求分析

2.1.1 适用场景

（1）车辆通信网络：实时监控车辆之间的通信网络的状态，包括网络拥塞、信道质量等，防范网络入侵和数据篡改，通过本地学习和模型共享，车辆可以迅速适应网络动态变化，提高网络的实时性和性能。这种策略使得车辆能够在运行中即时调整，有效降低网络拥塞风险，提高通信信道质量，为车辆通信安全性提供了可靠保障。

（2）车载系统：联邦学习可以检测和缓解车辆通信网络中的对抗攻击，以确保车辆通信网络的安全性。通过这一技术，能够有效地防范恶意代码注入和远程攻击，从而保护车载系统免受各类威胁。这项创新性的安全措施旨在保障车辆内部系统的稳健性和安全性，确保其正常运行。通过联邦学习的实施，我们能够在车联网中建立一个更为安全和可信赖的环境，阻止潜在的威胁，提高整个车载系统的抗攻击能力。

2.1.2 社会价值

（1）提升车联网安全：通过实时入侵检测和联邦学习的方法，提高车联网系统的整体安全性，降低恶意攻击和入侵的风险，从而有效地降低网络攻击成功率。通过这些措施，我们能够更好地防范交通事故、车辆劫持以及其他可能的恶意行为。实时入侵检测允许我们迅速发现并应

对潜在的威胁，确保系统在面对恶意攻击时能够迅速做出反应。这种综合的安全策略有助于维护车联网生态系统的稳定性，为用户提供更为安全可靠的使用环境。

（2）保护用户隐私：联联邦学习通过在本地设备上训练模型，仅分享模型更新，避免原始数据传输，利用差异隐私技术，确保在入侵检测过程中保护车主和车辆的隐私数据，增强用户信任感。避免敏感数据的不必要传播，为用户提供了更为可靠的隐私保护机制。

（3）降低维护成本：及时发现入侵事件，有助于提前采取措施，降低系统被攻击的风险，减少因安全问题而产生的维护和修复成本。通过迅速响应入侵威胁，能够最小化潜在的损害，避免问题扩大化导致更严重的安全漏洞。这种主动性的安全措施不仅有助于降低维护成本，同时也为系统的健康运行提供了更有力的保障。

（4）改善交通流和道路安全：通过实时监测车辆通信网络的状态，优化交通流和减少交通拥堵，从而提高道路安全性。这种实时的监控机制使得系统能够更灵敏地感知交通状况的变化，为交通流的优化提供了有力支持。联邦学习的协同特性使得车辆之间能够更加紧密地协作，尤其在紧急情况下，能够更及时地做出反应。这种协作机制的实施有助于减少事故发生的可能性，提高整体交通的安全性。通过车辆之间的即时信息共享，创造更为智能、更为安全的道路环境。

2.1.3 主要功能

（1）联邦学习入侵检测：通过多设备的协同建模，使车辆能够在本地进行模型训练。这种方式仅分享模型更新，实现了入侵检测的不断改进，同时有效保护了隐私。这一方法不仅能够提升入侵检测的效能，还确保了系统的隐私安全性。

（2）差异隐私保护：差异隐私技术通过在本地设备上引入噪声来模糊个体贡献。这一方法确保了在模型更新的过程中，个体车辆的隐私数据得到了有效的保护，避免了模型更新泄露敏感信息的风险。通过这种隐私安全的措施，联邦学习在保障个体隐私的同时，有效提升了整体的隐私安全水平。

（3）实时监控和警报：联邦学习通过在本地设备实时更新模型，为车联网提供了即时而有效的安全状态监控。在这个监控系统中，根据预设规则，联邦学习能够迅速发出及时的警报，以应对潜在的安全威胁。并使得车联网的实时安全性得以显著提高，系统能够更为灵敏地应对各种安全挑战。这种实时监控和预警机制为车联网的整体安全提供了一层强有力的保障，能够使安全问题能够得到及时发现和处理。

2.1.4 主要性能

（1）准确性：系统应具备高准确性，能够有效地检测各类入侵，并减少误报率，以防止误报给用户带来不必要的困扰。以防止误报给用户带来不必要的困扰。

（2）实时性：实时更新模型和实时监控，确保系统能够及时响应新型入侵和威胁，立即采取必要的措施。通过保持实时性，系统能够更加灵活地适应动态的安全环境，为用户提供更为可靠的安全保护。

（3）可扩展性：能够轻松扩展到大规模车联网系统，保持性能稳定。通过具备良好的可扩展性，系统能够灵活地应对未来可能的需求增长，为车联网系统的规模扩大提供了可靠的支持。这种性能的稳定性对于确保系统在面对不断增长的数据和用户负载时仍然能够保持高效运行至关重要。

2.2 可行性分析

2.2.1 经济可行性

（1）成本评估：本项目初期由学生组成的团队负责，我们充分认识到项目的成功实施不仅仅依赖于创新性的技术方案，还需要对项目成本进行详细的估算和合理的控制。在项目启动阶段，我们进行了全面的成本估算，覆盖硬件、软件开发、人力资源等各个方面的费用，以确保项目在预算内高效运作。成本评估过程如下：

表 3 成本预算

项目成本	预算与介绍
人力资源费用	前期由学生团队成员开发、测试和管理，无直接薪资支出。 中后期招聘开发团队，预计支出：100000。

软件和工具成本	开发工具和许可：选择免费或学生许可的开发工具，如FATE。
项目测试	系统集成测试：学生团队内部进行应用开发和测试，无额外费用。 性能测试优化：利用开源工具，最小化费用。
模型训练和优化	云端计算资源：申请学校提供的云计算资源或争取学术资源资助，最大支出：1000。
项目管理和咨询	前期项目经理和团队由学生团队担任，无额外费用。 中后期聘请外部咨询，预计支出 100000。

初步成本分析表明，项目的投资是可控的，且在财务承受范围内。经过精心的估算，我们项目初期总成本无直接支出，中后期项目总成本支出预算为 200000-250000。

（2）市场调研：通过市场调研，我们了解了车联网安全市场的潜在规模、增长趋势和竞争格局等。市场研究结果表明，项目在当前市场中有望占据一席之地。

①市场规模：全球车联网行业经过近二十年的发展，已经逐渐成熟并形成了相对稳定的产业链合作和商业模式运营。根据《2024 年中国车联网行业深度研究报告》，2021 年全球车联网市场规模达 1430 亿美元，预计 2023 年将达 1865 亿美元。2022 年中国车联网市场规模达 3878 亿元，

近五年年均复合增长率为 33.67%。随着车联网的普及，对安全性的需求不断增加，为联邦学习在该领域的广阔市场提供了良好的机遇。

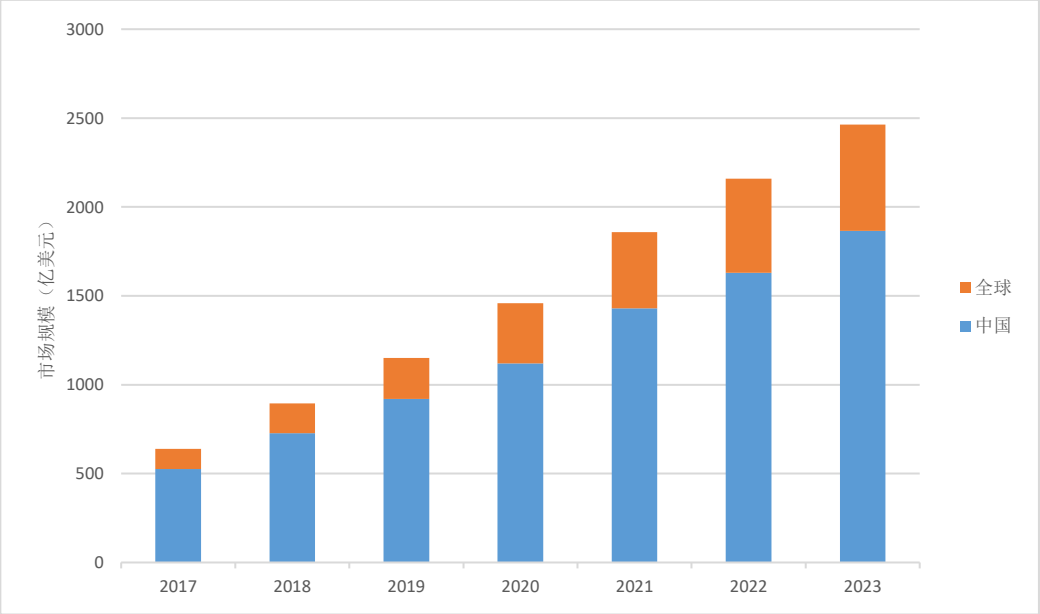

图 22017-2023 车联网市场规模图

②需求趋势：随着车联网的迅速普及，车辆通信网络的安全需求愈发显著。消费者对于数据隐私的关切不断提升，对网络攻击和信息泄露的担忧也逐渐加深。这一趋势在对实时监控、威胁检测以及隐私保护等方面的需求上表现得尤为迫切。在中国，车联网用户规模呈现出显著增长，从 2012 年的 400 万辆攀升至预计的 2023 年将达到 9057 万辆。这个增长趋势引起了各方广泛的关注，不仅是由于其便利性和创新性，更因为用户对安全性和隐私保护有强烈的需求。

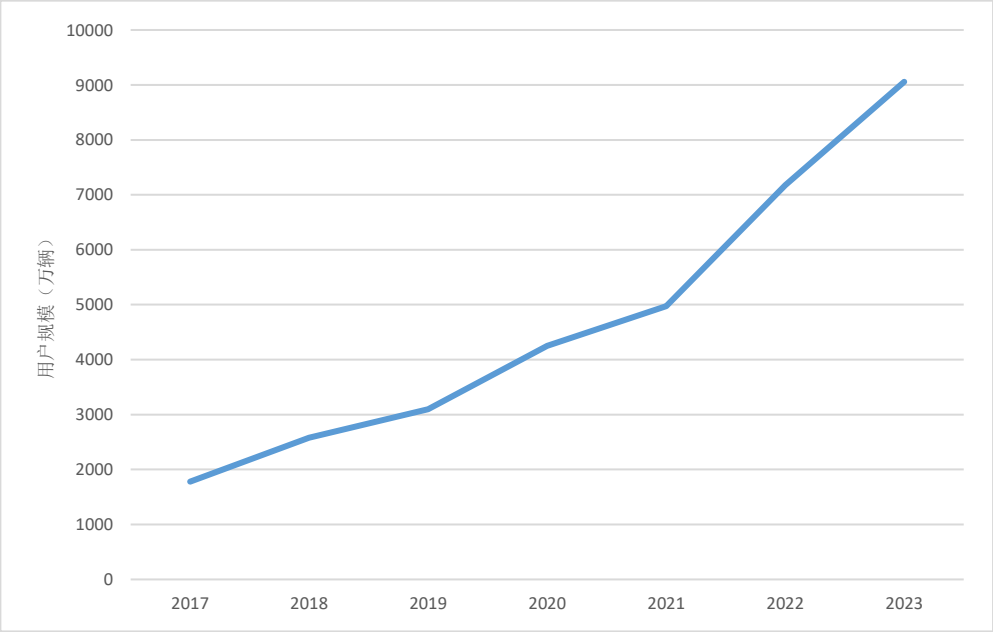

图 32016-2023 中国车联网用户规模

③竞争格局：目前，车联网安全领域的竞争格局呈现出两个明显的趋势，即传统安全解决方案提供商与新兴联邦学习技术公司的激烈竞争。传统公司专注于成熟的安全技术，而新兴公司则通过引入联邦学习技术，强调数据隐私、实时监控和攻击检测等创新功能。尽管领先公司在安全性能和技术创新方面占据优势，但新兴公司以其灵活的解决方案和先进技术成功吸引了广泛的关注。

随着市场的不断演变，竞争焦点逐渐集中在服务质量、隐私保护和技术创新上。这种竞争格局有望推动整个行业的不断进步，为用户提供更高水平的安全保障。传统公司在积极应对市场变化的同时，不可避免地要考虑融入更创新的元素，以满足用户对于更安全、更隐私保护的需求。相对地，新兴公司则需继续改进其解决方案的成熟度，以确保在竞争中保持竞

争力。整体而言，这种双趋势的竞争格局为车联网安全领域带来了更多的选择和机遇。

④潜在客户群：潜在客户群主要包括各类主体，涵盖车辆制造商、车队管理公司、政府机构，以及那些对车联网安全格外关注的企业。车辆制造商着眼于保障其生产车辆通信网络的安全性，以提升整体车辆品牌的安全形象。对于车队管理公司而言，他们渴望实现车队运营的高效与安全，因此对实时监控和攻击检测的需求显得尤为迫切。在政府机构这一层面，整体道路安全和信息基础设施的保障成为焦点，因此对车联网安全解决方案的需求也逐渐增加。那些特别关注车联网安全的企业，尤其是那些依赖车辆通信网络进行业务的公司，同样是潜在的合作伙伴。通过满足这一广泛的客户需求，项目有望建立起广泛而深刻的合作伙伴关系，从而促进本产品在市场上获得广泛的应用。本项目的目标是通过为各行业提供全面而创新的车联网安全解决方案，为客户提供极具价值的保障，进而推动整个车联网安全领域的不断发展。

⑤风险与挑战：首先，竞争激烈是主要要面对的挑战，因为传统的安全解决方案提供商已经在市场上占据了相当大的份额，这意味着需要通过创新和差异化来突破市场局面。法规合规性是第二个风险，尤其是在车联网领域，法规变化较为迅速，因此，需要时刻关注法规的更新，确保解决方案符合最新的法规要求。技术风险也是一个重要的考量因素，特别是在联邦学习模型的开发和实施阶段，需要应对可能出现的性能和安全性等方

面的问题，确保技术能够稳健地应对各种挑战。

在克服这些挑战的过程中，团队将注重多方面，包括但不限于市场营销战略的制定、技术创新的推动、法规合规性的确保。这将有助于提高本产品在市场中的竞争力，并助力更好地应对激烈的竞争环境。

⑥市场机会：在当前车联网安全领域，联邦学习技术的崭新优势为项目开创了独特的市场机遇。首先，联邦学习的核心优势之一是其在隐私保护方面。相较于传统的集中式学习模型，联邦学习在本地设备上进行模型训练，避免了敏感数据的集中存储，从而最大限度地减少了数据隐私泄露的风险。这不仅能够满足用户对于数据隐私的高度关切，还为解决方案奠定了可持续的市场地位。

其次，联邦学习在分布式环境下展现出的协同能力提供了独特的市场差异化关键。通过实现多方协作，可以提供个性化的安全解决方案，满足不同客户群体的特定需求。这种灵活性和个性化将成为本项目在市场中脱颖而出的关键因素，引领着创新与竞争力的新标准。

此外，随着车联网行业的迅速发展，对于实时监控和攻击检测等创新功能的需求也在不断升级。在这一趋势下，本项目有望通过整合联邦学习技术，为市场提供更为先进、更为智能的解决方案，满足用户对于车联网安全的不断提升的需求。

综上所述，联邦学习技术不仅为本项目开创了市场机会，还提供了构建创新、个性化、高效安全的解决方案。通过充分发挥这一技术的优势，

将在激烈的市场竞争中保持领先地位，并为客户提供卓越的价值。

⑦政策支持：近年来，我国出台了一系列与车联网相关的政策，鼓励发展智慧交通、智能网联汽车、自动驾驶等领域。在《车联网网络安全和数据安全标准体系建设指南》等政策的推动下，我国不断建设车联网协同服务综合监测平台，推动城市交通基础设施、交通载运工具、环境网联化和协同化发展。

2.2.2 技术可行性

技术验证：进行了详尽的技术验证，包括联邦学习和差异隐私技术的实际应用。通过小规模原型测试，验证了这些技术在车联网入侵检测中的可行性，并发现了可能的技术挑战。

3 系统设计与实现

3.1 概要设计

3.1.1Car-Hacking 数据集介绍

由于现代车辆已经能作为边缘计算设备参与联网，因此保护车载网络免受网络攻击成为一个重要问题。CAN（ControllerAreaNetwork）总线是目前车辆中最常见的通信总线之一。CAN 数据集包含 CAN 总线发送和接收的消息，以及时间戳等信息。但是由于 CAN 协议缺乏安全功能，车辆容易受到攻击。消息注入攻击是一种具有代表性的攻击类型，它注入捏造的消息来欺骗原始 ECU 或导致故障。因此 Car-Hacking 数据集可用来训练攻击检测模型，提高车联网信息安全。

Car-Hacking 数据集包括了一系列与汽车系统安全相关的数据。该数据集由多个攻击类型数据集构成：

表 4 数据集

攻击名称	攻击描述
Dos 攻击	每 0.3 毫秒注入“0000” CAN ID 消息
Fuzzy 攻击	每 0.5 毫秒注入完全随机的 CAN ID 和 DATA 值消息
Gear 攻击	每 0.1 毫秒注入 Gear 相关的 CAN ID 消息
RPM 攻击	每 0.1 毫秒注入 RPM 相关的 CAN ID 消息

数据集包含每 300 次消息注入入侵。每次入侵执行 3 到 5 秒，每个子数据集总共有 30 到 40 分钟的 CAN 流量。CAN 数据属性如下：

①TimeStamp

时间戳数据，表示每条 CANID 消息注入系统的时间；

②CAN ID

CANID 是车辆发送的十六进制 CAN 报文的标识符（例如 043f），每辆车都通过 CAN 总线控制每一个车辆元器件，CANID 代表了对应的发送方和接受方，每一个车辆厂商有自己的 CAN 解析协议；

③DLC

DATA 数据的字节数，大小为从 0 到 8，为 16 进制数据；

④DATA[0]-DATA[7]

CAN 消息的特征值，为 16 进制数据，单位为字节数；

⑤Flag

标志数据，值为 T 或 R，T 代表攻击消息，R 代表正常消息。

图为 Car-Hacking 数据集中进入控制器局域网(CAN)的消息数量统计，可以看出 Normal 消息仍是 CAN 中的主要数据，在大量的 Normal 消息中准确识别各类攻击消息则是我们的目标。

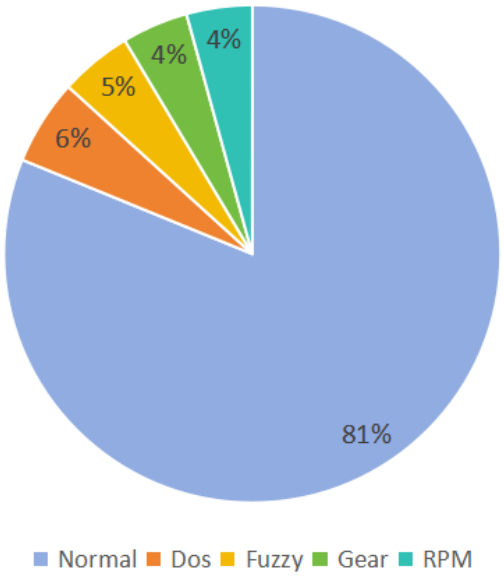

图 4 消息数量统计

本项目选取 TimeStamp 数据与 DATA[0]～DATA[7]等 9 类数据作为攻击消息检测模型的训练指标开展模型训练。

3.1.2 系统流程设计

本项目利用 FATE 平台与 OMNeT++、Veins 和 SUMO 等计算机网络仿真平台开发了一个基于分层卷积神经网络模型的车联网攻击检测系统，系统中内置了开展车联网攻击检测的分层卷积神经网络预训练模型。本系统利用 OMNeT++、Veins 和 SUMO 模拟汽车行驶过程中的车辆、路测基站等边缘计算平台间的通信过程，采用分层联邦的思想，将路测基站作为边缘聚合器聚合车辆等边缘计算平台的数据和参数，再将路测基站的数据聚合到云端，提高通信效率和检测精度。

系统的主要运行流程为：对于各类平台间的无标签通信数据，系统首

先会检验该条消息的数据特征(TimeStamp 和 DATA[0]～DATA[7]，不接受缺失值)是否完整，若消息中存在缺失值，会进行数据补全处理；通过数据处理通道后，系统加载预训练后的分层卷积神经网络分类器模型对消息进行检测分类：最后将预测结果呈现在仿真界面上，一次检测任务完成。图 5 为车联网攻击检测系统的简易流程图：

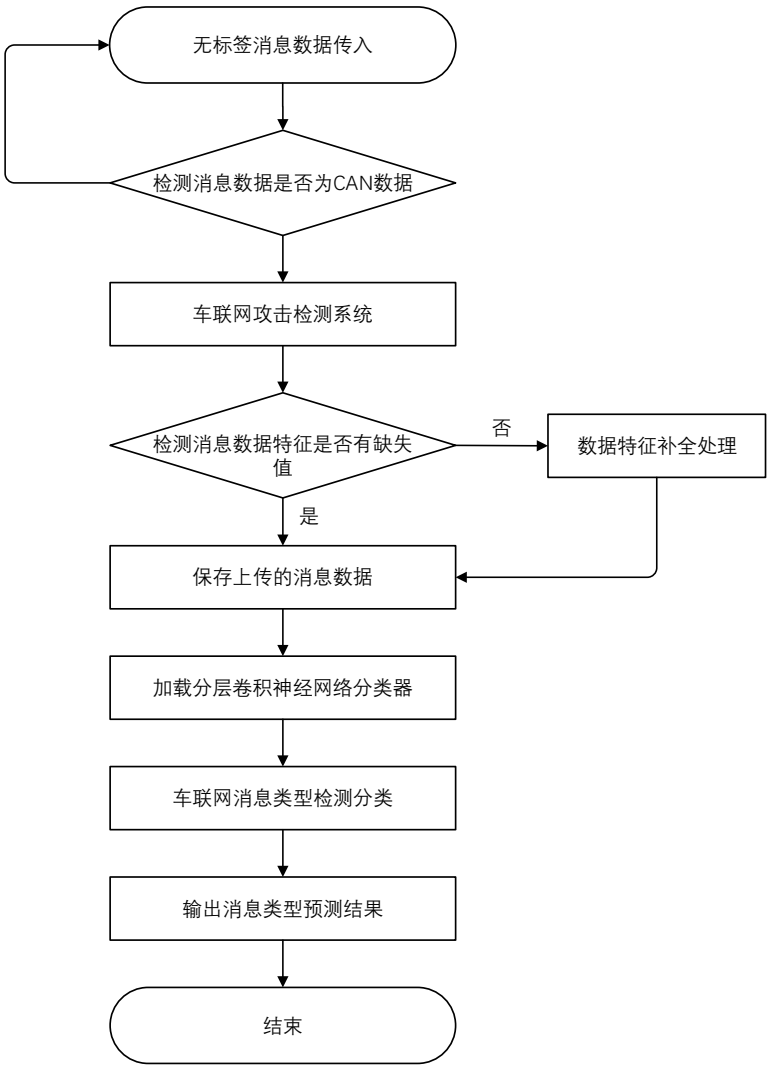

图 5 检测系统流程图

3.1.2 系统模块设计

本系统的核心为分层卷积神经网络分类器，它是基于 FATE 平台进行开发的。FATE 平台中内置了丰富的机器学习算法，通过例如名为 xx_xx 的 json 文件进行各类数据与命令的提交，降低了代码开发的难度，用户可通过各类模块和接口的调用进行个性化的机器学习算法设计，因此该分类器和系统都实现了高度的模块化。接下来介绍系统的模块设计：

data_deal：消息数据进入系统前的数据处理，包括去标签、数据补全，进制转换等；

upload_x_conf：将消息数据上传至 FATE 平台，指定文件地址，表名等；

homo_dsl：用于检测任务的配置文件。使用特定的语法和关键字，描述检测任务的各个方面，包括数据预处理、模型选择、算法参数等；

homo_mutli_label_predict：conf 文件设计分类器检测模块，用于指定 FATE 平台的行为和功能，使用同态加密技术，在不暴露原始数据的情况下，对新传入的消息数据进行多标签预测。最后利用 flowsubmit 命令执行预测分类任务，具体的模块间调用关系如图 6：

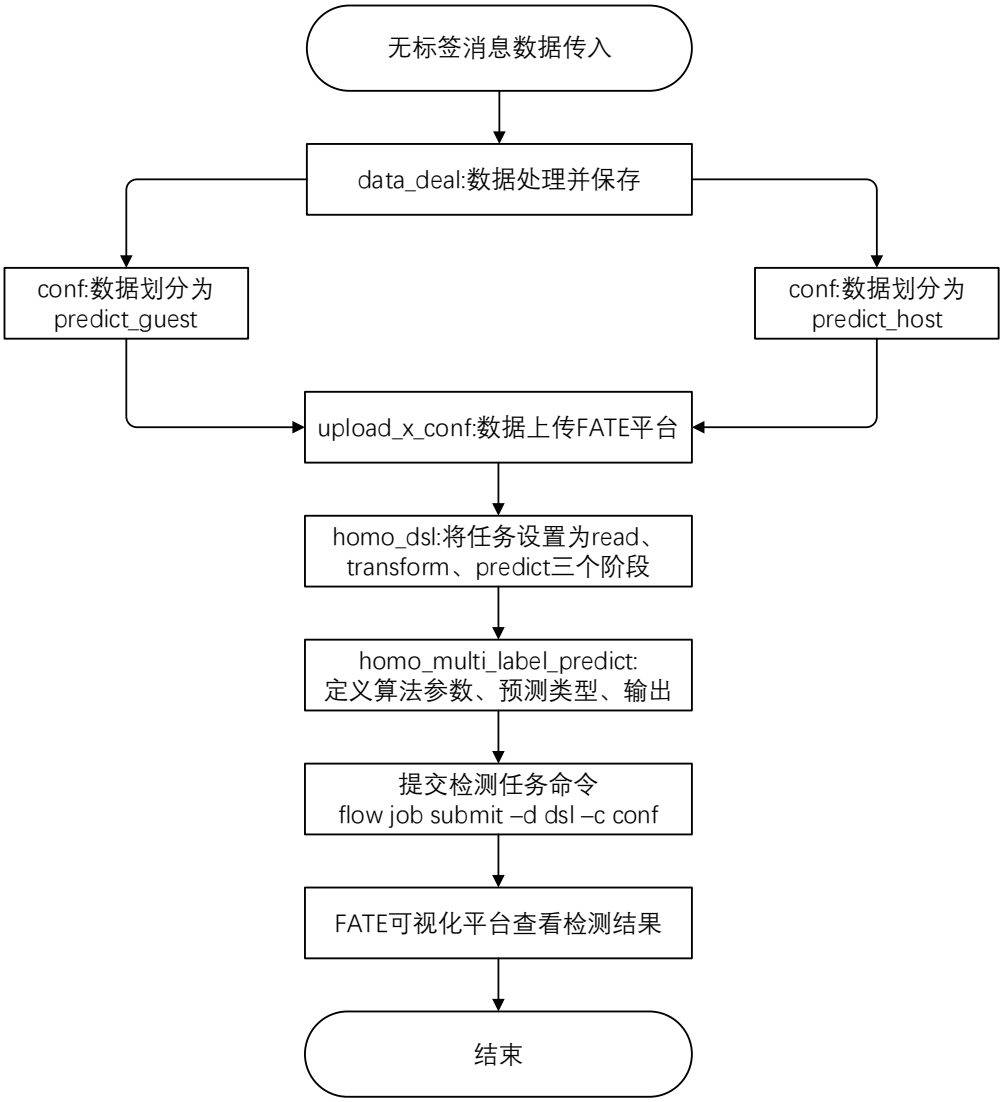

图 6 模块间调用关系

3.1.3 系统交互界面设计

我们充分考虑现实应用中车联网设备的异构网络环境。分别在不同的主机上模拟十台车联网设备作为联邦学习的客户端参与训练，为每一台模拟设备分配一定的计算资源。十台客户端设备都已经配置好了训练数据集和测试数据集。配置具体的测试服务端界面如图 7 所示：

图 7 服务端界面

当客户端要参与联邦学习时，点开运行界面首先在设备编号中输入自身的设备编号，编号范围 1-10 分别代表着 10 台模拟车联网设备。之后在服务端口中需要输入联邦学习服务端的 IP 地址以及开放的 gRPC 通信端口，载入在模拟车联网设备中预存的通信帧训练数据集之后点击参与训练。此时，下方窗口将会返回与服务端的通信结果，和本地训练进度条。模拟车联网设备就成功的参与了联邦学习。在数据集预处理时，我们将原始数据集分为了训练集和测试集两部分，其中测试集占总数据集的 20%。

模拟车联网设备在运行界面中选择入侵检测按钮的时候，通信帧数据集将会载入总数据量 20%的测试数据集，通过训练模型对于测试数据集

的识别准确率展示基于鲁棒性联邦学习车联网入侵检测系统的性能。

3.1.4 系统仿真

为了验证基于联邦学习的车联网攻击类型检测方案的有效性与可行性，我们使用 OMNeT++、Veins 和 SUMO 协同仿真平台生成了车联网流量场景，具体的设计参数如表 5

表 5 设计参数

参数	数值
仿真区域	3km*3km
路测基站通信范围	100m
车辆通信范围	100m
仿真时间	200s
发送数据大小	80bits
发送速率	6Mbs

为了验证联邦学习对于车联网交通通信的影响，我们使用仿真平台对于方案进行了评估。仿真使用德国 Erlangen 市的真实地图模拟道路交通，实时模拟车辆、路测基站、云服务器等计算平台间的两两通信。我们分别模拟了 20 辆、40 辆、60 辆汽车在不同速度下通信测试了平均丢包率(APLR)和平均通信延迟(ACD)的变化，仿真如图 8 所示：

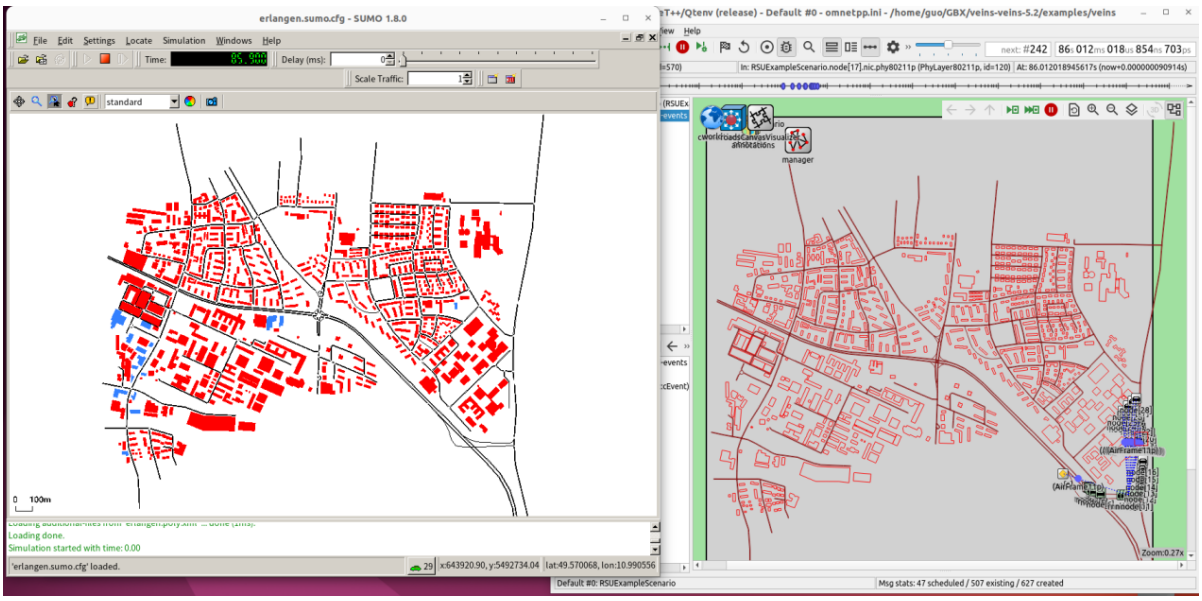

图 8 仿真测试

在仿真测试之中 APLR 是总发包数与车辆在模拟时间内接受的总包数量之比的平均值，图 9 展示了车速和车辆数量对于 APLR 的影响。当车辆数量一定时，APLR 随着车速的减少而升高。车速较低时，由于路测基站和车辆的通信范围内持续时间变长，数据包传输量较高，易使报文排队时间超过阈值导致丢包所以丢包率升高。在车速较高时，数据包传输量较低，通信双方在通信范围内停留时间也较短，所以丢包率也较低。在车速一定的时候，通常 APLR 与车辆数量同步。

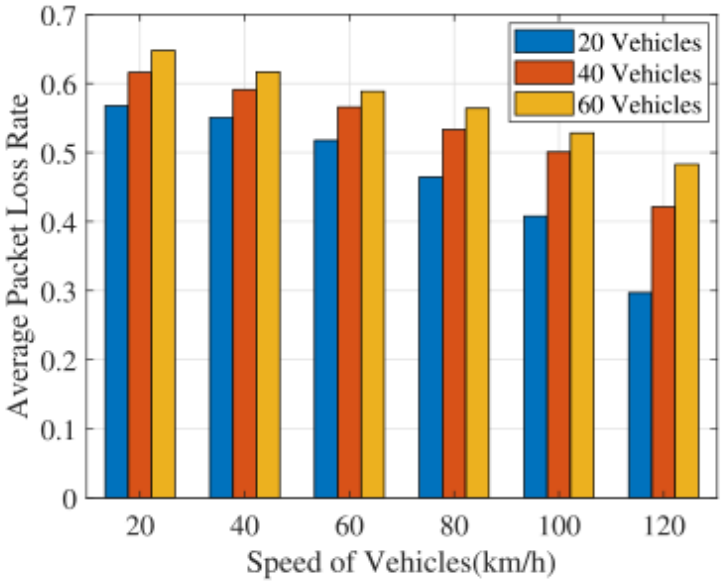

图 9 车速和车辆数量对于 APLR 的影响

在仿真测试中 ACD 被定义为边缘计算节点路测基站在进行局部聚合将全局模型广播到通信范围内参与学习的客户端汽车时间之间的平均值。车速和车辆数量对于 ACD 的影响如图 10 所示。

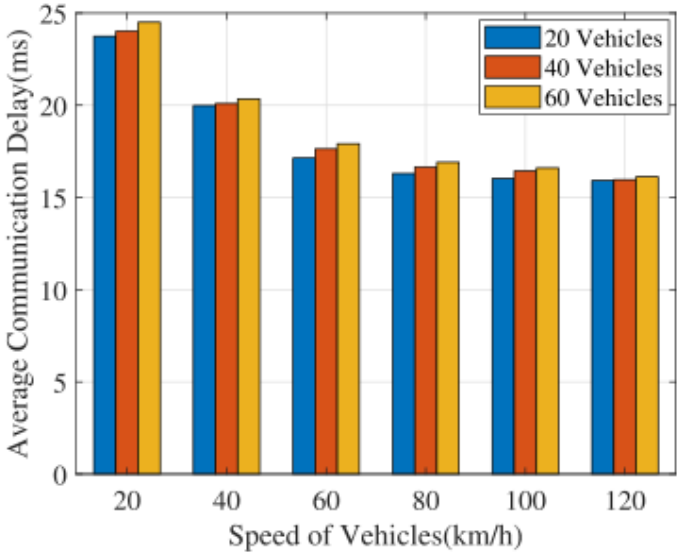

图 10 车速和车辆数量对于 ACD 的影响

3.2 详细设计

主要介绍系统内置的分层卷积神经网络分类器和横向联邦任务的训练策略设计、训练数据处理、算法设计、和算法模块内部的调用关系。

3.2.1 训练策略设计

在一般的神经网络多分类任务中，模型往往只能学习到训练集内的已知数据特征。例如 A 车辆在本地数据中只学习到了 Dos、Fuzzy 和 Gear 攻击，那么在遭受 RPM 类型的攻击消息就会被误判，对于车联网安全造成严重威胁。结合联邦学习可以在不泄露原始数据的情况下进行模型信息聚合优化的特点，我们希望系统能从聚合后的模型中学习到新的信息，从而能够增强抵御未知类型的攻击的能力。

因此，利用 FATE 平台进行训练时，guest 与 host 方的数据特征空间不必保持完全一致，在分配数据集时，不必要求每个子集都有 5 种消息数据类型。各个参与方在本地完成训练后，本地模型并不能检测未知攻击类型，但是在经过 arbiter 聚合后会学习到其他 client 模型的参数，返回本地后，模型得到了抵御本地未知攻击类型的能力。

3.2.2 数据预处理

本系统所用训练数据均来自于 Car-Hacking，均为 CAN 消息数据，该类型消息数据包括 Timestamp、CANID、DATA[0]~DATA[7]及 Flag 共计 12 列数据。为了更好地利用 FATE 平台开展联邦学习任务，系统需要

适配 FATE 平台数据处理格式，对 Car-Hacking 数据集进行数据预处理：

（1）选取训练指标特征：在 FATE 平台开展 CSV 数据分类，需要每条数据均包含至少 3 类特征：Sid（数据 ID）、Label（数据标签）、Features（数据特征）。对 Car-Hacking 数据集研究后，我们发现车联网消息的各类攻击不仅与数据特征 DATA[0]~DATA[7]相关，也与时间戳（TimeStamp）间隙有关，因此我们选定了 TimeSatmp、DATA[0~7]9 个特征作为训练数据特征 Features。原有的 CANID 并不能直接作为数据 ID 使用，因此对每条消息使用自然顺序进行编号得到新的 Sid。

（2）数据特征处理：Car-Hacking 数据集的数据特征 DATA[0 ~7]为 16 进制数据，而 TimeStamp 时间戳数据则是 10 进制数据，因此为了数据统一，我们将所有数据转化为 10 进制数据；同时由于 DLC（字节数）的大小范围为 0~8，在 Dos 和 Fuzzy 攻击中有 2 字节大小和 5 字节大小的数据，因此我们把所有数据特征补全为 8 字节，把缺失的 DATA[]用 0 填充；据观察，CAN 数据流的时间基本集中在大致相近的 30-40 分钟，因此对时间戳数据进行预处理，统一减去所有攻击的最早开始时间，再作为特征参与训练。

（3）数据标签 Label 处理：在每类消息的数据集中，注入消息（攻击）的 Flag 为 T，正常消息的 Flag 为 R，为了利用 FATE 平台开展横向联邦学习任务，使得系统学习到各种攻击的特征，我们需要一个统一的

Label 化处理，采用常见的阿拉伯数字进行 Label 的整理。令正常运行的 CAN 消息的 label 为 0，Dos 攻击为 1，Fuzzy 攻击为 2，gear 攻击为 3，RPM 攻击则为 4，在 Sid 列后加一列，作为每条消息数据的标签 label。

（4）数据集整合及分配：由于我们需要系统能具有抵抗未知类型攻击的能力，因此针对 3.2.1 中我们提出的训练策略，结合 FATE 平台的训练数据格式，我们对数据集进行随机划分，取 80%的数据为训练集，剩余 20%数据为测试集，设置了 4 个 client 参与分层卷积神经网络的模型训练。对于训练集，将 5 类消息数据随机分配到 4 份训练集 train_dataset_a，train_dataset_b，train_dataset_c，train_dataset_d 中。剩余的 20%数据无需分为两份，所有参与方共用一个测试集 test_dataset_a 进行模型测试。

3.2.3 分层卷积神经网络分类器算法关键设计

卷积神经网络（CNN）在分类任务中的优势主要体现在对数据特征的学习能力、参数共享和稀疏交互、适应大规模数据以及层级特征表示等方面。CNN 能够自动学习数据中的特征，无需手动设计特征提取器。通过参数共享和稀疏交互，降低了模型复杂度并提高了泛化能力。CNN 适应大规模数据，通过多层次的抽象特征表示，CNN 能够提高分类准确性。因此我们利用 CNN 的优势设计分层卷积神经网络分类器，接下来分层卷积神经网络分类器的具体设计：

（1）数据处理并输入：由于 Car-Hacking 数据集的数值特征均为数值型数据，可以直接将其转换为向量表示。将数值特征按照样本顺序排

列，每个消息数据为一个向量。将向量序列作为输入传入分类器。

（2）卷积、池化和全连接操作：检测数据先进行一次卷积操作，随后利用分层卷积的方法，在第二部分，也就是从二次卷积运算开始，分层分别进行两次卷积和两次池化操作并将结果传至下一层。总共进行五次卷积操作和四次池化操作，然后进行全连接操作。具体的全连接层设计如下：

① 利用输入层完成数据输入工作：具有 512 个神经元的全连接层（Denselayer），激活函数为 ReLU（RectifiedLinearUnit），输入形状为(9,)，表示输入特征的维度为 9。图 11 为输入层的定义：

```
model_nn = keras.Sequential()
model_nn.add(layers.Dense(512, activation='relu', kernel_regularizer=regularizers.l2(0.01), input_shape=(9,)))
model_nn.add(layers.BatchNormalization())
model_nn.add(layers.Dropout(0.5))
```

图 11 输入层的定义

②利用隐藏层完成特征提取、数据压缩、特征处理任务：

第一层隐藏层：具有 256 个神经元的全连接层，激活函数为 ReLU。

第二层隐藏层：具有 128 个神经元的全连接层，激活函数为 ReLU。

第三层隐藏层：具有 64 个神经元的全连接层，激活函数为 ReLU。

第四层隐藏层：具有 32 个神经元的全连接层，第五层隐藏层：具有 16 个神经元的全连接层，激活函数为 ReLU。激活函数为 ReLU。每个隐藏层都使用 ReLU 作为激活函数，有助于引入非线性，以更好地捕获数据中的复杂关系和模式，图为 12 为全连接层的定义：

```
model_nn.add(layers.Dense(256, activation='relu', kernel_regularizer=regularizers.l2(0.01)))
model_nn.add(layers.BatchNormalization())
model_nn.add(layers.Dropout(0.5))

model_nn.add(layers.Dense(128, activation='relu', kernel_regularizer=regularizers.l2(0.01)))
model_nn.add(layers.BatchNormalization())
model_nn.add(layers.Dropout(0.5))

model_nn.add(layers.Dense(64, activation='relu', kernel_regularizer=regularizers.l2(0.01)))
model_nn.add(layers.BatchNormalization())
model_nn.add(layers.Dropout(0.5))

model_nn.add(layers.Dense(32, activation='relu', kernel_regularizer=regularizers.l2(0.01)))
model_nn.add(layers.BatchNormalization())
model_nn.add(layers.Dropout(0.5))
```

图 12 全连接层定义

③利用输出层完成类别输出：具有 5 个神经元的全连接层，激活函数为 Softmax。这是一个多分类问题的输出层，其中 Softmax 函数用于将输出转换为表示概率分布的形式，输出的每个元素表示对应类别的概率。输入层有 9 个特征，输出层有 5 个类别，分别为四种车联网攻击类型和一种正常数据类型。

④L2 正则化、批归一化层和 Dropout 丢弃层：在分层神经网络分类器模型中，我们加了一些防止过拟合和加速训练的操作。在除输出层外的每一层都加入了 L2 正则化，其作用分别是是防止过拟合；批归一化层，起作用是加速训练过程，增强模型的稳定性；Dropout 层，它使得该层以 50%的概率丢弃神经元，有助于防止过拟合。全连接层设计如图 13 所示：

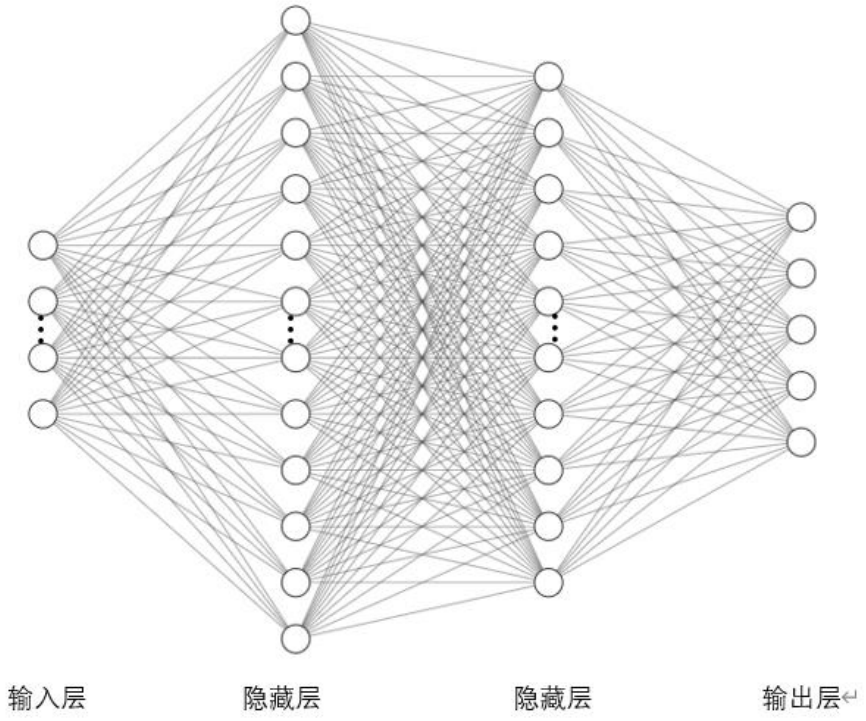

图 13 全连接层设计

（3）合并计算：第三部分工作是对第二部分分层卷积操作得到的输出数据使用 Tensorflow 中的 Concat 函数进行合并计算。

（4）反向传播：通过反向传播的方法优化网络参数，从而不断迭代训练优化直到使模型达到良好的收敛效果。

本系统采用的分层卷积神经网络分类器有三个部分组成：第一部分为数据输入模块。作为数据输入由一个数据输入层、一个卷积层组成；第二部分为分层聚合模块。作为数据的处理由 4 个卷积层、4 个池化层、2 个全连接层组成；第三部分为数据输出模块。作为数据的整合分类输出，由一个 Concat 函数用于合并卷积结果、一个 Softmax 层和一个输出层组成。其中 Conv1、Conv2、Conv43 个卷积层使用 PRelu 激活函数，

PRelu 激活函数是针对 Relu 函数的改进型，该方法有效解决了 Relu 函数在输入为负数的时候，不被激活的情况。另外由于 CNN 在训练过程中会出现过度拟合的现象，为了提高 CNN 的泛化能力，所以在全连接层 FC1_layer 和 FC2_layer 使用 L2 正则化方法和 Dropout 层。随后使用 Concat 函数对分层计算的结果进行合并，最后经由 Softmax 层对入侵检测模型方法的输出结果进行分类操作。

将预处理后的测试数据集输入基于 Normal、Dos、Fuzyy、Gear、RPM5 类消息数据训练的车联网攻击检测分类器模型。分类器对检测到的新样本进行分类检测，并输出作为检测结果的混淆矩阵。模型结构图如图 14 所示。

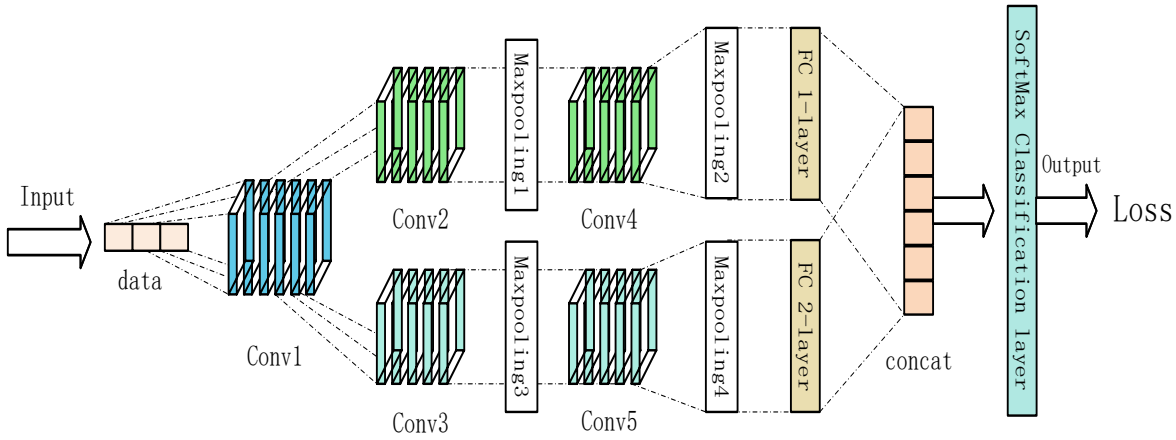

图 14 模型结构图

模型训练过程中的参数如表 6 所示：

表 6 模型训练参数

参数	数值
训练迭代次数	50
批处理大小	128
Early_Stopeps	0.0001
学习率	0.05
优化器	Adam
损失函数	CrossEntropy

3.2.4 训练过程中模块调用关系

分层卷积神经网络分类器是在 FATE 平台上进行开发训练的，通过 FATE 平台命令简介和模块化的特点，我们在 FATE 的联邦机器学习库中内置的 conf 文件 homo_nn_multi_predict.json 中设计了我们的算法，接下来介绍训练过程中的模块调用关系：

data_deal：数据训练前的数据处理，包括去标签、数据补全等；

upload_x_conf：消息数据上传至 FATE，指定文件地址，表名等；

homo_dsl：用于检测任务的配置文件。使用特定的语法和关键字，描述检测任务的各个方面，包括数据预处理、模型选择、算法参数等；

homo_mutli_label_train：conf 文件设计分类器检测模块，用于指定 FATE 平台的行为和功能，在该文件中我们设计了 3.2.3 中的分层神经网络分类器模型，最后利用 flow submit -c 命令执行预测分类任务，具体的

模块间调用关系如图 15 所示：

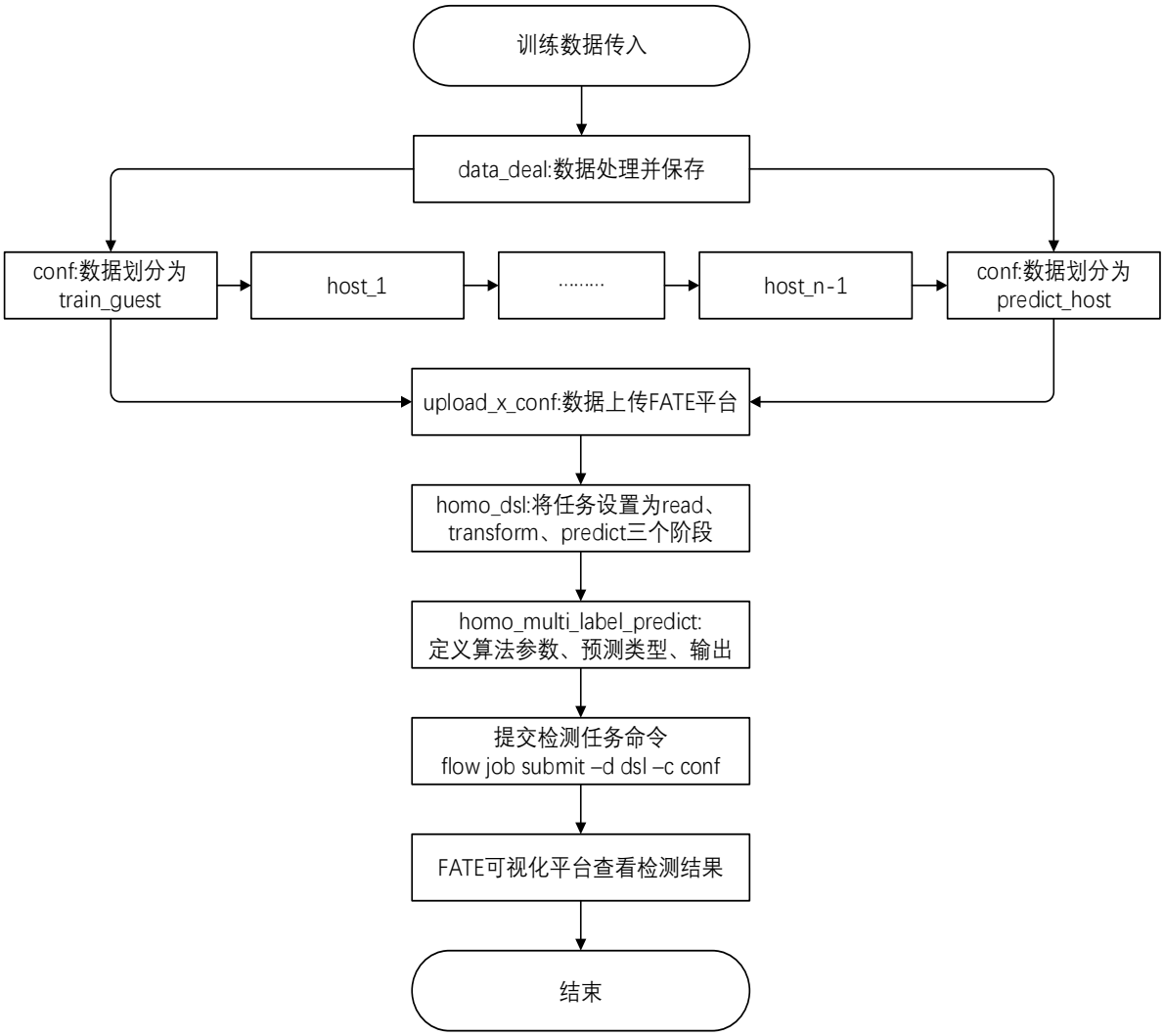

图 15 模块间调用关系图

3.3 系统测试

3.3.1 测试方案

在搭建完车联网攻击检测系统后，我们设计了验证车联网攻击检测系统精度的系统检测方案。

在验证车联网攻击检测系统精度的过程中，我们将每台客户端模拟为

有一定计算能力的 IoT 设备。每台 IoT 设备之中会预先存有 CAN 数据集，其中包括 Normal 数据，和经过 Dos、Fuzzy、RPM 等四种不同恶意攻击的异常 CAN 数据。每一个 IoT 设备可以使用 CAN 数据集参与车联网攻击检测系统的训练过程，在规定好的训练轮次之后对于预先分配好的测试集进行测试。

3.3.2 测试结果

利用 FATE 平台进行车联网攻击检测系统的测试。利用 flow model deploy --model-id --model-version 命令部署预训练模型，利用 flow submit -c 提交测试任务，图 16 为 FATE 平台对 test_dataset_a 测试集进行测试的后台测试图，图 17 为部分测试结果展示图。本系统在面对各类攻击时在 FATE 上的平均测试准确度为 0.9362。

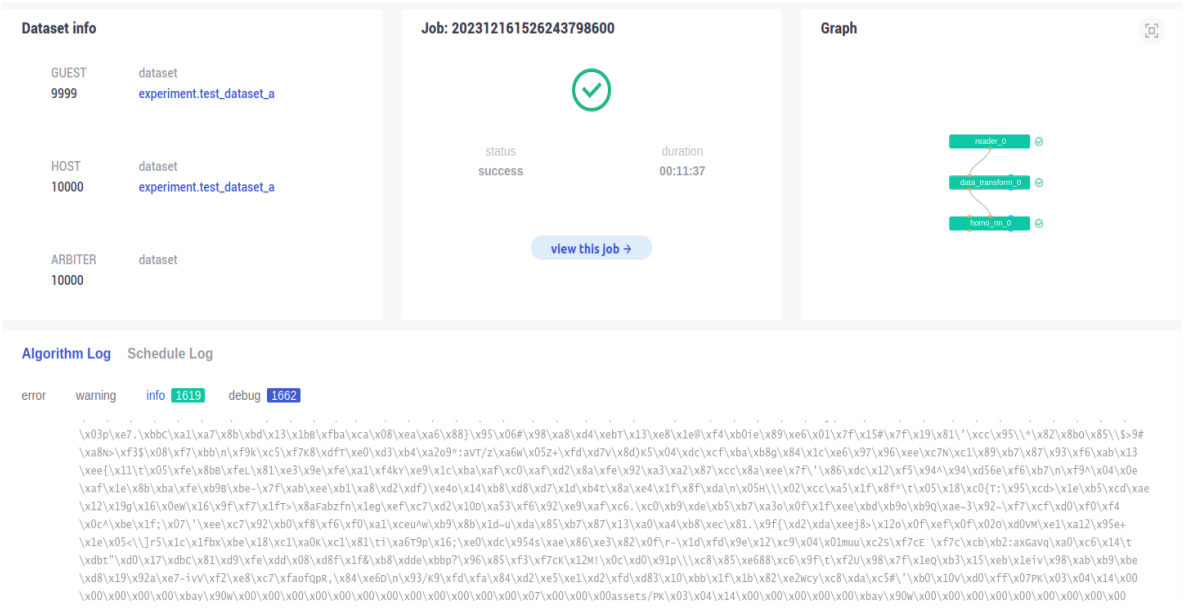

图 16 FATE 平台后台测试图

model output | data output | log

downLoad: Model Data

Outputting 3313894 instances (only 100 instances are shown in the table)

index	id	label	predict_result	predict_score	predict_detail	type
1	7772373	0	0	1	{"0":1,"1":3.2445920949975005…	predict
2	7772374	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
3	7772375	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
4	7772376	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
5	7772377	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
6	7772378	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
7	7772379	1	1	0.573541	{"0":0.39384984970092773,"1":…	predict
8	7772380	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
9	7772381	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
10	7772382	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
11	7772383	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
12	7772384	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
13	7772385	0	1	0.573541	{"0":0.39384984970092773,"1":…	predict
14	7772386	1	1	0.573541	{"0":0.39384984970092773,"1":…	predict
15	7772387	0	0	0.999979	{"0":0.9999797344207764,"1":0…	predict
16	7772388	0	0	1	{"0":1,"1":3.2445920949975005…	predict
17	7772389	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
18	7772390	0	0	1	{"0":1,"1":0,"2":0,"3":0,"4":0}	predict
19	7772391	1	1	0.573541	{"0":0.39384984970092773,"1":…	predict
20	7772392	0	0	1	{"0":1,"1":5.697796332921584e…	predict
21	7772393	1	1	0.573541	{"0":0.39384984970092773,"1":…	predict

图 17　部分检测结果展示图

由测试结果得出，各本地模型在聚合后，一定程度上提高了抵御未知风险的能力，面对各类攻击都保持着较高的准确度。证明我们的车联网攻击检测系统，可以有效的在各种情况下实现识别入侵攻击的功能。

4 总结

（1）联邦学习应用于车联网安全：

成功将联邦学习引入车联网入侵检测系统，实现了分布式学习，车辆之间通过本地训练模型并仅共享参数，从而在保护隐私的同时提高了检测性能。

（2）综合性入侵检测系统：

基于联邦学习的模型聚合，构建了综合性的入侵检测系统，具备实时监测、异常行为识别和迅速响应的功能，为车联网系统提供了全面的安全保障。

（3）将分层卷积神经网络模型引入 FATE 平台：

本项目在 FATE 平台中引入分层卷积神经网络模型，将其应用于数值数据处理中。传统上，分层卷积神经网络模型在中心化环境中的应用已取得显著成果，而本项目将其成功迁移到联邦学习框架，为数值数据处理任务的联邦学习提供了新的可能性。

（4）展望下一步工作：

① 生成对抗网络：

在后面的工作中考虑使用生成对抗网络去优化改进模型，将文本数据转化为图像数据进行模型学习，利用生成对抗网络优秀的图像生成能力，可以人为制造一些其他攻击供模型训练，使模型除了能够辨别已知的攻击

为还有一定的能力能够辨别一些未知的攻击，增强模型的实用性，再配合联邦学习解决数据孤岛的优异性，能够使用多模态的数据供模型学习使用，增强模型的学习能力。

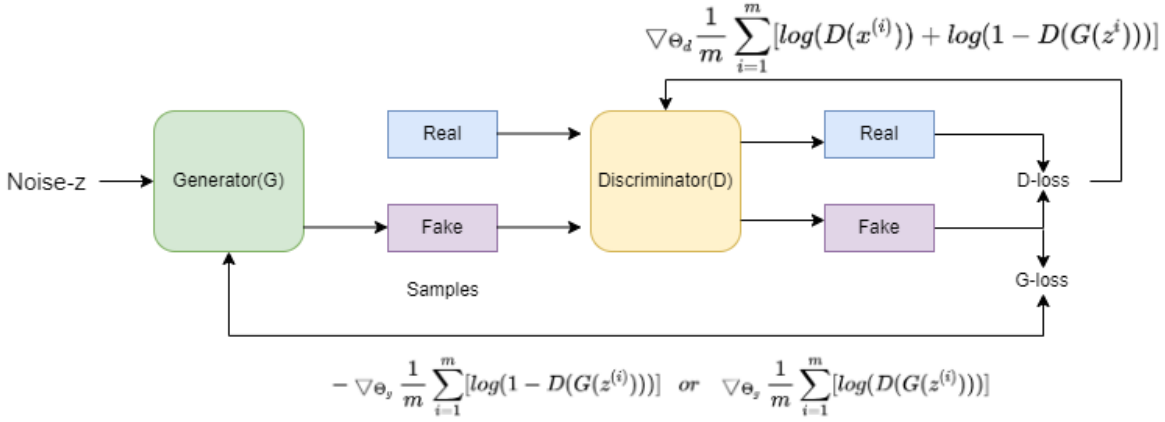

图 18 生成对抗网络结构

②拓展应用场景：

将联邦学习的理念扩展到其他车联网安全领域，例如智能交通管理和车辆身份认证。这样的拓展能够实现更广泛的应用，提升整个车联网生态系统的安全性。

③用户友好性：

在系统设计方面，致力于创建更加用户友好的界面和交互方式。通过提高系统的易用性，促进实际应用中的广泛采用。这样的努力将有助于用户更轻松地理解和使用系统，从而推动技术在实践中的更广泛应用。

希望在不断迭代和完善中，我们能够将这一创新推向更广泛的应用领域，为智慧出行的安全奠定坚实基础。

参考资料

[1] XiePWuBSunG.BAYHENN:CombiningBayesiandeeplearningandhomomorphicencryptionforsecureDNNinference[J].arXivpreprintarXiv:1906.00639,2019.[2]PokhrelSR,ChoiJ.ImprovingTCPPerformanceOverWiFiforInternetofVehicles:AFederatedLearningApproach[J].IEEETransactionsonVehicularTechnology,2020,PP(99):1-1.

[2] 武文涛,张志才,付芳.基于联邦学习的智能网联车驾驶策略优化研究[J].测试技术学报,2023.

[3] NguyenA,DoT,TranM,etal.DeepfederatedlearningforautonomousdrivingIEEEIntelligentVehiclesSymposium(IV).IEEE,2022:1824-1830.

[4] DengY,LyuF,RenJ,etal.SHARE:Shapingdatadistributionatedgeforcommunication-efficienthierarchicalfederatedlearning[C]//2021IEEE41stInternationalConferenceonDistributedComputingSystems(ICDCS).IEEE,2021:24-34.

[5] FanW,SuY,LiuJ,etal.Jointtaskoffloadingandresourceallocationforvehicularedgecomputingbasedonv2iandv2vmodes[J].IEEETransactionsonIntelligentTransportationSystems,2023,24(4):4277-4292.

[6] 俞建业.车联网分布式智能入侵检测系统研究与实现[D].南京理工大学,2021.DOI:10.27241/d.cnki.gnjgu.2021.001173