

面向物联网系统的 APT 攻击活动预测方法

成翔^{1,2}, 匡苗苗¹, 张佳乐¹, 陈玮彤¹, 李云¹, 杨宏宇^{3,4}

¹扬州大学信息工程学院(人工智能学院) 扬州 中国 225127

²中国民航大学民航飞联网重点实验室 天津 中国 300300

³中国民航大学安全科学与工程学院 天津 中国 300300

⁴中国民航大学计算机科学与技术学院 天津 中国 300300

摘要 随着信息技术及应用的快速发展, 各类物联网系统也面临着大量的信息安全风险和隐患, 网络信息安全事件频繁发生。其中, 以“震网病毒”攻击事件为代表的高级持续性威胁(Advanced Persistent Threat, APT), 凭借其特有的高级性、持续性和针对性, 成功窃取了政府、金融机构、大型企业等组织信息系统中的机密数据(或破坏了这些关键信息系统), 这不仅阻碍了各行业生产经营活动的正常运行, 还严重影响到了社会公共秩序安全乃至国家安全。鉴于 APT 攻击活动发生频率较低, 单个设备很难获取充足的 APT 攻击数据, 对攻击活动预测模型进行训练。针对上述 APT 攻击活动预测模型训练数据不足且分散的问题, 本文提出了一种面向物联网系统的 APT 攻击活动预测方法(APT activities prediction method for IoT systems, APTPIS)。该方法首次应用联邦学习机制实现了物联网系统中可疑活动特征的聚合, 并且无须预设关联规则。为了实现隐私保护特性, 攻击活动预测模型采用了差分隐私数据扰动机制, 在物联网终端设备数据特征中加入了 Laplacian 随机噪声, 从而最大限度地保护了终端设备的数据隐私。利用训练后的攻击预测模型, 设计了一种 APT 攻击活动预测流程, 通过将系统当前产生的日志数据输入预测模型, 可预测后续 APT 攻击活动出现的概率。仿真实验结果表明, APTPIS 方法能够在保障终端设备数据隐私的前提下准确、高效地预测出物联网系统中后续 APT 攻击活动出现的概率。

关键词 高级持续性威胁; 联邦学习; 攻击活动预测; 物联网; 差分隐私

中图分类号 TP391 DOI号 10.19363/J.cnki.cn10-1380/tn.2025.07.12

APT Attack Activity Prediction Method for Iot Systems

CHENG Xiang^{1,2}, KUANG Miaomiao¹, ZHANG Jiale¹, CHEN Weitong¹, LI Yun¹, YANG Hongyu^{3,4}

¹School of Information Engineering, Yangzhou University, Yangzhou 225127, China

²Key Laboratory of Flying Internet, Civil Aviation University of China, Tianjin 300300, China

³School of Safety Science and Engineering, Civil Aviation University of China, Tianjin 300300, China

⁴School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China

Abstract With the rapid development of information technology and application, all kinds of network information systems are also coping with a large number of information security risks and hidden danger, and all kinds of network information security incidents occur frequently. Represented by the “Stuxnet”, Advanced Persistent Threats (APT) with its characteristic of advanced, sustainability, and targeted, have theft the confidential data (destroyed the critical information system) of government, financial institutions, large enterprises and other organizations. It not only arises all kinds of significant security risks in the information systems, hinders the normal operation of industry and business, and also seriously affects public security and even national security. As the APT attacks usually will not launch frequently, it is difficult for a single device to capture sufficient APT attack data, the data scale is not enough for training the attack activity prediction mode. Aiming at the problem of insufficient and scattered APT attack prediction data, we proposed an APT activities prediction method for IoT systems, named APTPIS to predict the probability of subsequent APT attacks occurred in IoT systems. It is the first time to apply a federated learning mechanism to aggregate suspicious activities features in the IoT systems, where the APT prediction phase does not need any correlation rules. Moreover, to achieve privacy-preserving property, we further adopt a differentially private data perturbation mechanism to add the Laplacian random noise to the IoT device training data features, so as to achieve the maximum protection of private data. Using the trained attack prediction model, a kind of APT attack activity prediction process is designed. Inputting the log data generated in the system into the prediction model, the probability of subsequent APT attack can be predicted. The experimental results show that APTPIS can accurately and efficiently predict the probability of subsequent APT attack activities occurring in the IoT system.

通讯作者: 成翔, 博士, 实验师, Email: huozhai9527@126.com。

本课题得到江苏省基础研究计划青年基金项目(No. BK20230558)、中国民航大学民航飞联网重点实验室开放基金(No. MHFLW202304)、新疆维吾尔自治区自然科学基金项目(No. 2024D01A40)资助。

收稿日期: 2022-09-14; 修改日期: 2024-02-29; 定稿日期: 2025-06-11

Key words advanced persistent threat; federated learning; attack activity prediction; Internet of Things; differential privacy

1 引言

随着信息技术的不断发展, 各行各业的生产经营活动逐渐依赖于信息技术服务所提供的便捷性和高效性。物联网系统作为一种极具代表性的信息系统, 使得传统生产和经营活动在时间和空间上的局限性得以打破。然而, 在物联网技术及应用快速发展的同时, 各类物联网系统中仍然存在着大量的信息安全风险和隐患。在各种类型的攻击活动中, 高级持续性威胁(Advanced Persistent Threat, APT)作为一种高级的多步攻击行为具有渗透性、隐蔽性和针对性的特点, 给各类物联网系统造成了严峻的威胁^[1]。为了防范这种日益增长且复杂的安全风险, 研究人员和机构提出了各种安全检测技术, 如入侵检测技术、恶意代码检测技术、漏洞检测技术等。然而, APT 攻击通常采取分步渗透长期潜藏的策略逐步实现窃取系统中机密数据的目的。如果仅仅针对恶意 APT 攻击行为进行检测, 则很难满足对物联网系统有效防御的高水准安全需求, 还应该分析 APT 攻击活动的演化过程, 并预测后续攻击活动出现的可能性。

攻击活动预测主要用于在获取攻击意图的基础上进一步深入分析相邻步骤攻击活动之间的逻辑关系, 从而推断攻击活动后续的变化趋势。目前, 攻击预测的热点问题涉及以下四类问题: 攻击/入侵预测、攻击投射、攻击意图推演和网络安全态势预测^[2-3]。攻击/入侵预测作为一种普遍采用的方法, 主要致力于解决预测是否会有恶意活动发生。攻击投射和攻击意图推演对入侵检测技术的依赖性比较高, 主要目标是预测攻击者下一步可能采取的攻击活动以及攻击者最终的攻击目的。网络安全态势预测本质上是一个与网络态势感知相关的概念理论, 主要对网络中恶意活动和漏洞数量的波动情况进行预测。

具体预测方法方面, Luo 等人在文献[4]中提出了对垂直联邦学习中的模型预测进行特征推断攻击的问题, 根据单个预测输出, 提出了对逻辑回归和决策树模型的两种具体攻击。进一步设计了一种基于攻击者积累的多个预测输出的通用攻击方法, 以处理神经网络和随机森林等复杂模型。Rajagopal 等人在文献[5]提出方法的目标是通过使用一种称为 stacking 的集成技术来获得可靠的预测。实验表明随机森林、逻辑回归、K 近邻和支持向量机等算法的组合在实时数据集上的预测结果优于模拟数据集。

Deng 等人在文献[6]中基于拒绝服务攻击, 提出了一种新的事件触发预测控制方案, 该方案将基于模型的网络控制系统、预测控制方案和事件触发方案相结合。提出的周期性拒绝服务攻击, 以及时间约束拒绝服务攻击, 对于两种攻击, 所提方案具有不同的预测步骤。Amaizu 等人在文献[7]提出了一种复合的、高效的分布式拒绝服务检测框架。该框架由两个结构不同的深度神经网络模型连接得到的模型组成, 以产生一个新的分类器, 能够预测各种类型的分布式拒绝服务和良性流量。Latif 等人在文献[8]为运用机器学习技术, 对拒绝服务(DoS)、恶意操作、恶意控制、数据类型探测、间谍、扫描、错误设置等网络安全攻击进行了预测。提出了一种基于轻量级随机神经网络的工业物联网攻击预测方案。所提方案利用最新的安全相关数据集, 以较高的准确率检测 IoT 攻击, 并减少了预测时间。Wallace 等人在文献[9]定义了通用的对抗性触发器: 与输入无关的标记序列, 当连接到数据集中的任何输入时, 触发模型产生特定的预测, 提出一种梯度引导的 token 搜索, 找到成功触发目标预测的短触发序列。Zhang 等人在文献[10]提出一种新的攻击方法, 称为生成模型-反演攻击, 可以高成功率地反演深度神经网络。利用部分公共信息, 这些信息可以非常通用, 通过生成对抗网络(GANs)学习分布先验, 并使用它来指导反演过程。该文还提供了理论分析, 表明了模型的预测能力与其对反转攻击的脆弱性之间的基本联系。Zhang 等人在文献[11]研究轨迹预测模型的对抗鲁棒性, 提出了一种新的对抗攻击, 通过扰动正常的车辆轨迹来最大化预测误差。实验表明, 对抗性预测将预测误差提高了 150%以上。Du 等人在文献[12]使用差分隐私实现智能边缘的机器学习策略。首先通过引入 Laplacian 机制, 设计了两种满足差分隐私保护的算法 OPP 和 OJP; 此外, 考虑了关联数据集的隐私保护问题, 并通过理论分析证明了关联数据集的差分隐私保护。

由于本文提出的 APT 攻击活动预测方法使用到了差分隐私联邦学习机制, 本部分还将对具有隐私保护能力的深度学习方法进行归纳总结。为了优化深度神经网络的效率, 模型划分技术应运而生^[13], 该技术可以将松散耦合的隐含层^[14]分配给第三方进行模型训练。同时, 为了防止恶意服务器窃取用户的隐私数据, 研究者从不同角度对差分机器学习的隐私保护需求进行了深入研究^[15]。Zhu 等人在文献[16]

采用特定机制分析的方法研究了隐私会计问题。提出了支配对分布的概念,证明了每种机制的隐私配置都具有紧支配对的特征,并推导了一些有用的支配对代数,包括自适应合成和抽样放大。在算法上,提出了一种分析型傅里叶会计,象征性地表示了支配对的特征函数,具有类似 rdp 的自然组成,并允许利用现成的数字工具。Kandpal 等人在文献[17]通过强调序列级复制是最近提出的针对语言模型的隐私攻击成功的一个重要因素,向这一目标迈出了一步。专注于基于生成的记忆,其中生成模型通过从其训练数据中生成精确样本来泄漏信息。在学习语言模型中基于生成的记忆时,将模型的生成行为与通过记忆完美拟合训练数据的模型的预期行为进行了比较。这种机制对于保护用户数据隐私具有非常明显的作用。为了防止每个用户的更新通过学习模型共享泄露给不可信的第三方, Qi 等人在文献[18]提出一种基于联盟链的联邦学习框架,在没有中心化模型协调器的情况下实现去中心化、可靠和安全的联邦学习。为了相同的研究目的, Feldman 等人在文献[19]通过洗牌实现隐私放大的简单且近乎最优的分析。然而,上述所有这些隐私解决方案都必须依赖于可靠的聚合服务器,这些服务器的主要任务为在全局模型参数中增添干扰噪声,并安全地为每个用户分配包含噪声的模型参数。这意味着聚合服务器可以读取任意一个用户的私有模型参数,如果聚合服务器为不可信设备,则用户的数据隐私将无法受到保护。因此,在基于联邦学习的 APT 攻击活动预测中,必须采用一种切实可行的机制来保护物联网设备的数据隐私不会受到任何第三方的影响。

纵观目前已有的研究成果,预测物联系统中的 APT 攻击活动将主要面临以下挑战: (1) 攻击模式复杂多变。传统的攻击活动预测方法一般首先会设置一定的攻击规则和模式,然后对已产生的告警信息(攻击活动)进行分析和推理,得到攻击活动出现的概率。然而, APT 攻击活动具有攻击模式变化多样的特性,并且经常利用零日漏洞发起攻击,无法确定完备的攻击模式。因此,基于规则和模式的攻击活动预测方法无法满足 APT 攻击活动准确预测的需求。(2) 预测数据不足且分布不均衡。作为一种会利用各类技术手段分步进行实施的攻击行为, APT 攻击涉及的攻击模式是非常多的。对于单个安全管理组织(或物联网系统中单个攻击预测主机)很难获取包含 APT 所有攻击阶段的攻击数据及攻击模式。此外,物联网系统中的不同区域可能会面对不同类

型的 APT 攻击活动,这些攻击活动所涉及的数据量也是很小的。因此,单个主机设备无法获得充足的预测数据对预测模型进行有效的训练。(3) 数据孤岛问题。传统的云计算架构可实现将不同安全管理组织收集的预测数据整合起来进行预测模型训练。然而,这种集中式的训练方式会存在数据隐私泄露的问题。如何实现在不泄露数据隐私的前提下实现预测数据共享是构建 APT 攻击活动预测模型过程中面临的巨大挑战。(4) 物联网终端设备资源受限。物联网终端设备一般只具有特定的功能,因此其存储容量和计算能力都非常有限。将大规模的数据分析和处理服务直接分配给物联网终端设备是很难实现的。为了应对上述问题与挑战,本文提出了一种面向物联网系统的 APT 攻击活动预测方法(APT activities prediction method for IoT systems, APTPIS),用于预测物联网系统中后序 APT 攻击活动出现的概率。本文的主要贡献和创新点如下:

(1) 本文提出的 APTPIS 方法将联邦学习机制首次应用于物联网系统中可疑攻击活动特征的聚合过程。利用基于 Louvain 方法的日志群落挖掘,在初始阶段,每一个实例代表一个独立群落,根据量化参数的定义,将日志实例移入相关联的日志群落,区分恶意与良性的日志群落,从而确定 APT 可疑活动特征。物联网终端设备联合边缘服务器,借助已检测到的恶意日志数据与告警信息对 APT 攻击活动预测模型进行训练,仅将预测模型的参数更新上传至安全服务云。安全服务云对收到的参数更新进行加权平均,得到预测模型的全局参数并分发给各边缘服务器,实现了一种不需要预先设定攻击模式的 APT 攻击活动预测方法,可对 APT 攻击的演化过程进行准确预测。

(2) 为了保护物联网终端设备的隐私数据不被非可信边缘服务器窃取,本文采用了一种差分隐私数据扰动机制,将 Laplacian 随机噪声增加至物联网终端设备训练数据特征之中,实现了对物联网终端设备数据隐私最大程度的保护。

(3) 本文提出了一种基于边缘计算框架进行 APT 攻击预测模型的训练。边缘服务器不仅可以分担物联网终端设备的计算开销,还可以减少物联网终端设备与安全服务云之间的通信开销。

下文主要包括:第 2 节给出了物联网系统中 APT 攻击活动预测过程的整体概览,对 APTPIS 方法进行了整体描述并提供了一种将此方法应用于物联网环境的边缘计算架构;第 3 节对 APTPIS 方法的技术细节进行了详细描述,其中包括联邦学习方

法和 APT 攻击活动预测流程。第 4 节对本文节所提出方法的实验与评估分析过程进行了描述。第 5 节对本文节的内容进行了总结分析并得出了一些结论。

2 方法模型与架构

2.1 整体架构

由于使用云计算架构进行物联网环境中联邦学习过程会面临以下两个主要问题: 首先, 资源受限的物联网终端设备无法胜任频繁的模式训练过程; 其次, 大规模分散的物联网终端设备在联邦学习过程中会与安全服务云之间产生大量的通信开销。为此, 本文还给出了一种边缘计算系统架构(在大多场景适用), 可将 APTIS 方法中所涉及的模型拆分并分别部署在边缘计算架构中的安全服务云、边缘服务器以及物联网终端设备中(如图 1 所示)。物联网系统中所有的终端设备都参与到了安全服务云提供的联邦学习协议当中, 从而获得的 APT 攻击活动预测服务。由于恶意边缘服务器或非可信的安全服务云有可能参与到联邦学习过程当中, 物联网终端设备中的隐私数据信息有可能被这些恶意实体所窃取。为此, 本文将本地预测模型拆分为终端设备部分和边缘部分, 两个部分的预测模型共同构成了本文提出的 ICP-GRU 模型。其中, ICP 模型(Google 提出的 Inception 模型)是一种卷积神经网络, 这一模型可被部署在物联网终端设备中, 用于从大规模日志数据中提取系统日志中的重要特征。以 ICP 模型的输出数据作为训练数据, 部署在边缘服务器中的 GRU 模型(Gated Recurrent Units 模型)可实现对 APT 攻击活动演化过程的学习。整合以上两种模型而构成的 APTIS 方法的主要学习目标并不是对 APT 攻击活动进行检测, 而是预测当前出现的 APT 攻击活动发展到下一攻击阶段的概率, 有助于实现对整个物联网系统安全态势的预测。这种 APT 攻击活动预测过程中涉及的实体如下文所述:

1) 物联网终端设备

物联网终端设备包含多种类型物联网系统中的终端处理设施设备, 例如工控机、智能监测设备等。各类物联网终端设备都具有一定的通信和计算能力, 可利用部署在其中的 ICP 模型完成对本地日志数据特征的提取。提取到的数据特征可通过基站传输至边缘服务器进行后续的训练过程。

2) 边缘服务器

边缘服务器所具备的计算资源和存储资源要远强于物联网终端设备, 他们通常会被部署在云数据中心与物联网终端设备之间, 作为一种计算节点为

终端设备提供计算外包服务。边缘服务器在基于联邦学习的 APT 攻击活动预测过程中主要负责对 GRU 模型进行训练。联邦学习作为一种高效通信且具有隐私保护能力的分布式机器学习方法, 可指派边缘服务器负责将物联网终端设备提供的日志数据特征用于对本地 GRU 模型进行训练。边缘服务器会将训练得到的参数更新上传至安全服务云, 而无需向云端提供完整的数据信息。以这种方式向云数据中心提供 APT 攻击活动预测所需的训练数据可以很好的保证不同区域内的日志数据相互隔离, 使得物联网终端设备的数据隐私可以得到有效保护。

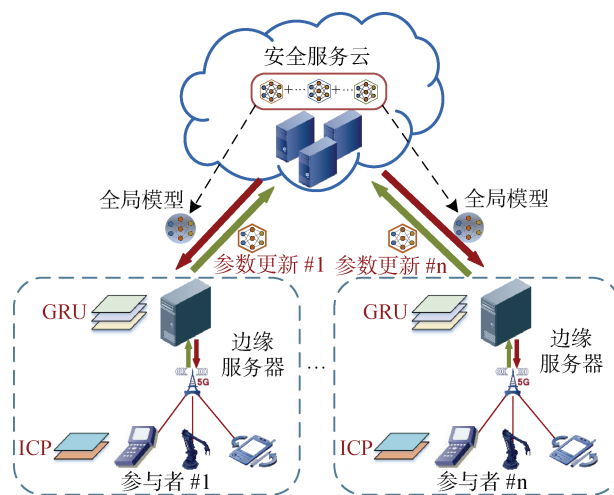


图 1 APTIS 方法系统架构

Figure 1 APTIS system architecture

3) 安全服务云

安全服务云收到各边缘服务器上传的本地参数更新后会进行聚合操作, 并将聚合得到的全局模型参数分发回各边缘服务器, 用于 APT 攻击活动预测。以这种机制进行的模型训练过程中, 各边缘服务器协同其管辖的物联网终端设备作为联邦学习参与者只需与安全服务云进行模型参数更新的交互, 无须涉及完整的日志数据信息。各边缘服务器可以属于不同的安全管理组织, 在不担心泄露其数据隐私的情形下向云数据中心提供出现在其系统中的 APT 攻击活动预测模型参数。云数据中心负责储存和管理这些预测模型参数, 并为各联邦学习参与者提供集聚全体贡献的全局模型。利用这种分布式的机器学习机制, 很好的解决了 APT 攻击数据分布不均衡的问题, 打破了 APT 攻击活动预测过程中的“数据孤岛”。

2.2 威胁模型及假设

APT 攻击是一类具有相同特性的攻击行为的总称, 在不同的 APT 攻击情境中攻击者会采用不同的攻击手段。但由于这些 APT 攻击情境一般具有相同

的最终目的, 其攻击过程可总结概括为以下步骤。首先, APT 攻击者会通过各种手段完成对目标系统防御体系的首次突破, 试图潜入目标系统, 获取系统权限。然后, 攻击者会通过 C&C 服务器建立连接来向目标系统中的已入侵实体发送攻击指令。接着, 攻击者会开始搜集系统内的各类信息资产用于提升权限和在系统中横向移动。最终, 攻击者会逐渐接近关键核心信息资产, 达到窃取机密隐私数据或者破坏系统核心基础设施的目的。在这种攻击过程中, 攻击者会在目标系统中驻留相当长的一段时间, APT 攻击活动也就具有了步步推进逐渐演化的特性。本文提出的 APTIS 方法就是通过对 APT 攻击活动演化过程所产生的各种日志数据进行学习, 获取 APT 攻击活动的演化特征, 得到可用于 APT 攻击活动预测的 ICP-GRU 模型。当系统中出现新的 APT 攻击活动时便可将相关日志实例输入到 ICP-GRU 模型中, 预测后序 APT 攻击活动出现的概率。

在 ICP-GRU 模型涉及的实体中, 物联网终端设备可被视为可信的实体, 它们通过与其他物联网终端设备协作执行模型训练过程而受益于最终形成的安全服务。然而, 作为第三方边缘服务器和安全服务云的可信性则有待商榷^[20]。具体而言, 边缘服务器和安全服务云会遵从联邦学习协议准确计算各类学习结果, 但他们同时也有可能试图获取物联网终端设备日志中包含的隐私数据。除此之外, 本文不考虑在预测模型训练过程中会有实体发起投毒攻击^[21]。最后, 本文还假设 ICP-GRU 模型所获取的日志数据可

涵盖系统中所有与 APT 攻击相关的活动, 同时这些数据没有被攻击者篡改。

3 APT 攻击活动预测方法设计

3.1 联邦学习模型

本文提出一种高效且具有差分隐私特性的联邦学习模型—ICP-GRU 模型(模型训练过程如图 2 所示), 实现对物联网系统中 APT 攻击活动的预测。ICP-GRU 模型依靠基于边缘计算的物联网系统架构, 可将边缘计算协议分发给各参与者。为了在提高模型预测效果的同时保证物联网终端设备日志数据的隐私, 可将 ICP-GRU 模型拆分为两部分。其中 ICP 模型被部署在物联网终端设备中, GRU 模型被部署在边缘服务器中。这种部署方式还能有效减小物联网终端设备的计算开销^[22]。物联网终端设备只需要通过 ICP 模型提取日志数据中的轻量级数据特征, 然后在提取的特征中添加特定的 Laplace 噪声并将干扰后的数据特征传输至部署于边缘服务器上的 GRU 模型。这种具有差分隐私功能的模型训练机制可以有效保护物联网终端设备的数据隐私。部署在边缘服务器上的 GRU 模型主要负责提取 APT 攻击活动的演化特征, 最终将训练得到的参数更新传输至云数据中心。云数据中心在收到各边缘服务器上传的参数更新后进行聚合平均操作, 将最终得到的全局模型参数, 并逐级分发给各边缘服务器和物联网终端设备, 物联网终端设备和边缘服务器便可在本地进行后续的 APT 攻击活动预测。

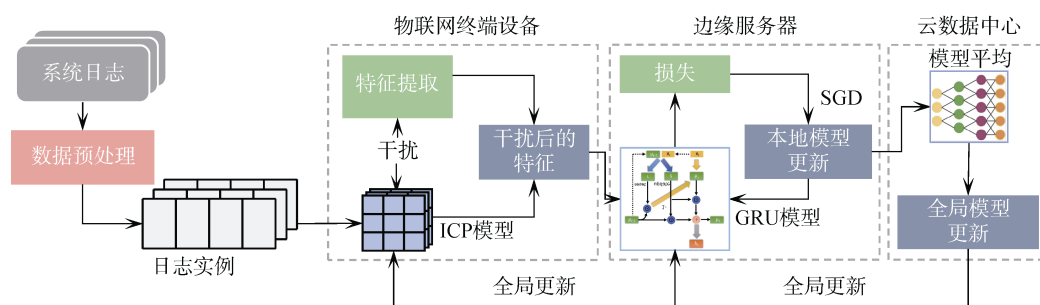


图 2 ICP-GRU 模型
Figure 2 ICP-GRU model

为了向基于联邦学习的 APT 攻击活动预测过程提供合适的训练数据, 本文还对物联网终端设备中日志数据进行了预处理操作。由于日志数据中的某些特征是字符型数据(例如, pname、q_domain 等), 在将数据集输入神经网络之前需要将所有字符型特征转换为数值型特征。同时, 不同特征的维度也是大不相同的, 且数值区间跨度也非常大。一些高维度特征数据可能权重较大, 从而忽略了低维度数据中隐藏

的信息。因此, 本方法在正式开始模型训练之前对原始数据进行了预处理操作。

3.1.1 ICP 模型与特征干扰

随着卷积神经网络(Convolutional Neural Networks, CNN)在图像处理领域的快速发展和出色应用, 这一技术在其他技术领域也逐渐得到了很好的应用。然而, 传统的 CNN 仅致力于提取单一的局部特征而忽略了对多种局部特征的聚合。为了解决这一问题, 谷

表 1 ICP 参数
Table 1 ICP parameter

Type	Patch size/stride	Depth	#1×1	#3×3 reduce	#3×3	#5×5 reduce	#5×5	Pool proj
Convolution-1	7×7/2	1						
Max pool-1	3×3/2	0						
Convolution-2	3×3/1	2		64	192			
Max pool-2	3×3/2	0						
Inception-(3a)		2	64	96	128	16	32	32
Inception-(3b)		2	128	128	192	32	96	64
Maxpool-3	3×3/2	0						
Inception-(4a)		2	192	96	208	16	48	64
Inception-(4b)		2	160	112	224	24	64	64
Inception-(4c)		2	128	128	256	24	64	64
Inception-(4d)		2	112	144	288	32	64	64
Inception-(4e)		2	256	160	320	32	128	128
Maxpool-4	3×3/2	0						
Inception-(5a)		2	256	160	320	32	128	128
Inception-(5b)		2	384	192	384	48	128	128
Average pool-5	7×7/1	0						
Dropout(40%)		0						
Linear		1						
Softmax		0						

歌公司在 GoogLeNet 中提出了一种叫做 Inception 的卷积神经网络。Inception 对 1×1 、 3×3 和 5×5 卷积核进行聚合, 并进行池化操作。多重卷积核从不同维度对数据特征进行提取并进行融合, 可以更好的提取日志实例中的关键特征。本文提出的 ICP 模型采用了 Inception 模型的基本结构对不同规模的日志实例进行卷积操作提取关键特征可以更加智能的捕获日志实例中隐藏的特征。

完成数据预处理后, 日志实例根据时间顺序以数据流的形式输入到 ICP 卷积模型中进行模型训练。根据数据之间的上下文相关性, 将数据流划分为固定大小的向量, 每个向量包含 n 个数据块。对每个向量进行处理后, 可以形成一个 $n \times m$ 维特征矩阵, 其中 m 表示每个数据的特征个数。ICP 通过 1×1 、 2×2 、 3×3 的卷积核和 2×2 的 max-pooling 来提取日志实例中的特征。在特征提取过程中, 为了匹配输出矩阵的宽度和高度, 需要使用相同的卷积。由于 ICP 在执行卷积操作时会产生大量的计算开销。因此, 可将 1×1 卷积设置在 max-pooling 操作之后, 并在 1×1 卷积之后设置 2×2 卷积和 3×3 卷积, 以降低特征维数, 加快计算速度。

联邦学习方法通过在本地训练敏感数据, 可以为物联网终端设备提供基本的数据隐私保护。然而, 这些敏感数据(例如数据特征和梯度)仍有可能被非

可信安全服务云或边缘服务器窃取。因此, 针对非可信安全服务云和边缘服务器, 迫切需要提出一种有效的保护机制来保护物联网终端设备的数据隐私。为了应对上述挑战实现隐私保护, 本文将差分隐私机制应用于 ICP 模型中, 提出了一种差分隐私机制, 在 ICP 模型提取出的数据特征中增加了 Laplacian 随机噪声, 并在基于边缘计算的物联网系统中构建了私有联邦学习协议。这样非可信的边缘服务器就无法得到物联网终端设备日志实例中的实际特征, 有效保护了物联网终端设备的敏感数据隐私。ICP 模型可以视为一种确定性函数: $x_l = F(x_r)$, 其中 x_r 代表包含物联网终端设备隐私信息的日志实例, x_l 代表 ICP 模型第 l 层的输出数据。根据差分隐私的基本定义, 全局敏感度查询 $f: D \rightarrow R$ 如公式(1)所示。

$$\Delta f = \max_{d \in D, d' \in D} \|f(d) - f(d')\| \quad (1)$$

其中 D 表示输入域, d 为其中的具体数据点; 而 D' 表示另一个输入域, d' 为其中的具体数据点。 F 表示的是输入 x_r 到输出 x_l 的转换函数, 它是一个确定性的映射。而 f 是一个查询函数, 将输入域 D 中的数据映射到输出域 R 中。

3.1.2 GRU 模型

门控循环单元(Gated Recurrent Unit, GRU)是一

种类类似于长短期记忆人工神经网络(Long Short Term Memory networks, LSTM)的循环神经网络, 可用于解决反向传播中的长期记忆和梯度问题。本文在 ICP-GRU 模型中采用了门控循环单元一方面是为了获得与 LSTM 相同的训练效果, 另一方面也是为了希望通过节省计算资源达到提高预测准确性的目的。输入日志实例表示为 19 维向量, 特征输入向量少在表现方面可能略劣于传统的机器学习方法, 但由于通过时间窗口划分的日志数据, 输入的向量具有时序关系, 且 GRU 模型能捕捉序列中的长期依赖关系的优势对比传统机器学习明显。因此, 即使在输入向量较少的情况下, 也能够达到较好的训练效果。Transformer 等较新的模型在处理特征复杂且长期依赖关系, 训练结果更优, 但同时产生的费用更高, 文中的特征序列参数较少, 相对简单, GRU 相较新模型具有更高的效率以及更低的费用。

与传统的循环神经网络相似, 由前序节点传递得到的隐藏状态 h^{t-1} 和当前输入 x^t 共同组成了 GRU 的输入项。通过将 h^{t-1} 和 x^t 相结合, 可以得到当前节点的输出项 y^t 和 h^t , 并将二者输入到后序节点。在初始阶段, 复位门(reset gate)和更新门(update gate)可由 h^{t-1} 和 x^t 计算得到。公式(2)和公式(3)分别给出了复位门(其数值由 r 表示)和更新门(其数值由 z 表示)的计算方法, 其中 δ 代表 Sigmoid 函数。

$$r = \delta(W^r x^t + U^r h^{t-1}) \quad (2)$$

$$z = \delta(W^z x^t + U^z h^{t-1}) \quad (3)$$

获得门信号后, 复位门可用于计算复位数据 h^{t-1} (计算方法如公式 4 所示)。将 h^{t-1} 和 x^t 输入到激活函数 $\tanh(\cdot)$ 中, 可保证输出数据的范围在 -1 至 1 之间。从而, h^t 可通过当前输入的数据 x^t 和前序节点传递得到的隐藏状态 h^{t-1} 计算得到(如公式 5 所示)。将 h^t 添加到当前隐藏状态, 便可实现对当前状态的记忆。

$$h^{t-1} = h^{t-1} \odot r \quad (4)$$

$$h^t = \tanh(W^h x^t + U^h (h^{t-1} \odot r)) \quad (5)$$

最后, GRU 同时进行遗忘和记忆两种操作。通过当前隐藏状态和位于 0 至 1 之间的更新门数值可以通过 $(1-z) \odot h^{t-1}$ 计算之前有多少记忆被遗忘, 并通过 $z \odot h^t$ 计算 h^t 保留了多少当前节点的信息。更新门的值越接近于 1, 越多的记忆会得到保留。反之, 更新门的值越接近于 0, 越多的记忆会被遗忘。通过多轮迭代, 最终状态可由公式(6)表示。

$$h^t = (1-z) \odot h^{t-1} + z \odot h^t \quad (6)$$

经过 ICP 模型从大量日志实例中提取的多维度数据特征以序列化的形式输入得到 GRU 模型之中, 从而可通过选择记忆或遗忘来学习其中的时序特征。通过梯度反向传播方法不断更新模型参数, 经过多轮迭代最终便可得到一个有效的 APT 攻击活动预测模型。

表 2 GRU 参数
Table 2 GRU Parameter

Symbol	Parameter	Values
N	# of Neurons	['64,128,256,512,1024']
η	Learning Rate	[0.1,0.5,1,2,5,10]
L	Sliding Window Steps	[3,6,12,18,24]
	Optimizer	['adam']
	Loss	['MSE']
BS	Batch Size(b)	[16,32,64,128,256]
E	Epochs	['500,1000']
	Activation Function	['Sigmoid']
	Dropouts	N/A
	Weight Regularizer	N/A
	Kernel_INITIALIZER	['Normal']
	Last Activation	['RELU']

3.1.3 联邦学习流程

本文提出的 APTIS 方法可应用于基于边缘计算的物联网系统, 具体的 APT 攻击活动预测模型的联邦学习流程主要包括以下步骤(如图 3 所示):

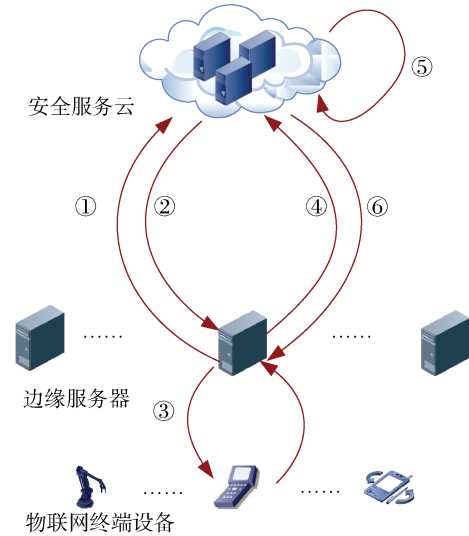


图 3 联邦学习流程

Figure 3 Federated process

1) 边缘服务器作为联邦学习参与者首先向云数据中心请求获得 ICP-GRU 初始模型。

2) 当安全服务云收到来自边缘服务器的请求后将初始的 ICP-GRU 分发至边缘服务器。边缘服务器收到模型后将其中的 ICP 模型分发至所管辖的物联网终端设备。

3) 当物联网终端设备中出现告警后, 将告警发生前一段时间内的收集到的日志实例输入入到 ICP 模型中进行训练, 在提取到的数据特征中添加 Laplace 噪声, 并将干扰后的数据特征上传至边缘服务器。

4) 边缘服务器将收到的物联网终端设备日志实例特征输入 GRU 模型进行训练, 然后将训练完成后得到的本地参数更新上传至安全服务云。

5) 安全服务云对所有边缘服务器上传的本地参数更新进行聚合操作。

6) 安全服务云再次将聚合得到的全局参数更新分发至各边缘服务器。经过多轮迭代后安全服务云可聚合得到最终收敛的全局模型。

最终, 安全服务云将最终收敛的 ICP-GRU 全局模型分发至加入到联邦学习过程的各边缘服务器, 可用于后续 APT 攻击活动预测。APTPIS 方法中联邦学习过程的伪代码如表 3 所示, 伪代码中使用到的符号在表 4 进行了解释说明。

3.2 APT 攻击活动预测流程

APT 攻击者对目标系统的攻击时间跨度一般会持续较长一段时间且每步攻击之间会间隔较长一段时间, 使得防御者必须持续地监控系统中出现的恶意行为。然而, 这种攻击特性对于防御者也是一把双刃剑, 不同攻击步骤之间的间隔为防御者预留了足够的时间预测攻击者的下一步攻击活动。当检测到 APT 攻击情境的某步攻击活动并生成相应的告警时, 可以对相关的日志数据进行分析, 预测攻击者发起下一步攻击活动的概率, 从而在 APT 攻击者达到最终攻击目标之前实施适当的防御措施。本小节将首先进行日志实例构建过程的研究, 然后给出 APT 攻击活动预测的具体过程。

3.2.1 日志实例构建

被 APT 攻击入侵后的物联网终端设备中生成的日志数据可用于 APT 攻击活动预测模型的训练。当在物联网系统中检测到某步 APT 攻击活动时, 安全监测设备将产生相应的告警信息。虽然识别出的 APT 攻击活动已经触发了告警信息, 但仍有大量相关的未被检测到的非攻击性恶意行为, 这些行为可能会成为攻击者发起下一步攻击活动的跳板。这些非攻击性恶意活动通常会以系统日志的形式留存在目标系统中。因此, 需要对系统日志信息中的特征进行提取和学习, 挖掘这些特征与后续攻击活动之间的关系。从而, 在出现类似的系统日志时可对后续攻

击活动进行有效预测和预警。为了对不同类型的日志数据中的参数属性进行规范化处理, 首先需要将日志数据构建为日志实例。日志的类型主要取决于物联网终端设备中装载了何种操作系统, 不同的系统会产生不同类型的日志。本文假设所有物联网终端设备中都装载了 Windows Embedded Compact (Windows CE) 操作系统, 那么终端设备中可采集的日志类型如表 5 所示。

为了有效地处理数据, 需要将日志数据转换为统一的格式。为此, 本文选取了 19 个参数用于构建日志实例(如表 6 所示), 每个日志实例可表示为一个 19 维矢量: $A(I(\log)_m) = (a_1, a_2, \dots, a_{18}, a_{19})$ 。不同类型的日志实例包含的有效参数值可能存在差异性, 若某个日志实例仅包含 a_1 、 a_3 和 a_5 三个有效参数, 那么矢量中其他位置将设为“0”。

表 3 APTPIS 方法中联邦学习过程的伪代码

Table 3 The APTPIS pseudo-code

APTPIS 联邦学习算法

```

1: procedure 物联网终端设备
2:   矩阵变换:  $x_r' = x_r \odot I_n$ ;
3:   特征提取:  $x_l = F(x_r')$ ;
4:   限定边界:  $x_l' = x_l / \max\left(1, \frac{\|x_l\|_\infty}{B}\right)$ ;
5:   添加噪声:  $\bar{x}_l = x_l' + \text{Lap}(B/\delta I)$ ;
6:   上传  $\bar{x}_l \leftarrow \bar{x}_l$  至边缘服务器;
7: end procedure
8: procedure 边缘服务器
9:   EVENT: 收到添加噪声的特征  $\bar{x}_l$ 
10:   初始化操作:  $w_t^{(i)} \leftarrow w_t^{(global)}$ ;
11:   for 每个  $E_k \in (1, \dots, E)$  do
12:     for 批处理  $b \in B$  do
13:        $g_t^{(i)} = \nabla_{w_t} L(w_t, \bar{x}_l)$ ;
14:        $w_{t+1}^{(i)} = w_t^{(i)} - \sigma \cdot g_t^{(i)}$ ;
15:        $\Delta w_{t+1}^{(i)} = w_{t+1}^{(i)} - w_t^{(i)}$ ;
16:     end for
17:     上传  $\Delta w_{t+1}^{(i)}$  至安全服务云;
18:   end procedure
19: procedure 安全服务云
20:   EVENT: 收到本地模型更新  $\Delta w_{t+1}^{(i)}$ 
21:   初始化全局参数  $w_t^{(global)}$ ;
22:   for 所有  $i \in [1, n]$  do
23:      $w_{t+1}^{(global)} = w_t^{(global)} + \frac{1}{n} \sum_{i=1}^n \Delta w_{t+1}^{(i)}$ ;
24:   end for
25:   发送  $w_{t+1}^{(global)}$  至各边缘服务器;
26: end procedure

```

表 4 参数说明

Table 4 The notations used in the pseudo-code

符号	说明
x_r	物联网终端设备原始数据
x_r'	矩阵变换后的 x_r
$\bar{x}_i$	干扰后的特征
I_n	零位矩阵
x_l	神经网络第 1 层输出
x_l'	经过边界限定后的 x_l
l	神经网络的层级
$F(x_r')$	深度神经网络
$L(w_l, x_l')$	目标损失函数
B	限定边界
E	本地迭代
δ	本地学习效率
$w_t^{(i)}$	物联网终端设备 i 在 t 时刻的参数值
$g_t^{(i)}$	物联网终端设备 i 在 t 时刻的梯度
$\Delta w_{t+1}^{(i)}$	t 时刻本地模型更新
$w_t^{(global)}$	t 时刻全局模型参数
σ	噪声规模

表 5 APT 攻击活动预测模型及到的日志类型

Table 5 Log types from different recorders

编号	日志类型	提供日志数据的应用程序
1	HTTP	Internet explorer
2	Object access	Audit
3	DNS	Tshark
4	Authentication	Syslogd
5	Process create	Audit
6	WFP connect	Audit

3.2.2 APT 攻击活动预测过程

当 ICP-GRU 模型完成对目标物联网系统中 APT 攻击行为演化过程的学习后, 可用于预测后序 APT 攻击活动出现的概率。通过我们之前的研究工作^[24]可知, APT 攻击情境中可被检测到的攻击活动主要属于以下 4 个攻击步骤: 首次突破防御、C&C 通信、资产/数据挖掘和数据外泄。在上述 4 个攻击步骤间隔中产生的日志数据首先可通过我们之前的研究工作^[23]提出的 LICM 方法进行日志群落挖掘。挖掘得到的带有标签的日志实例可用于 ICP-GRU 模型的训练。完成训练后的模型可根据以下流程进行 APT 活动预测(如图 4 所示)。

假设在时间点 t_1 和 t_2 时刻分别在物联网系统检测到某个 APT 攻击情境中第 i 步和第 $i+1$ 步 APT 攻击行为, 那么这两步攻击涉及的物联网终端设备需

使用时间窗口 $[t_1 - \tau, t_2]$ 内的所有标记过的日志实例进行 ICP-GRU 模型训练。选择这一时间的数据作为预测模型训练数据是由于这段时间内出现的网络活动对于促成后序 APT 攻击活动起到了潜在的推动作用。经过一定时间的积累, 便可完成 ICP-GRU 模型的训练。当有新的 APT 告警出现时, 可将这一告警所涉及的物联网终端设备中后续出现的日志数据输入到已完成联邦学习的 ICP-GRU 模型中, 便可计算得到后序 APT 攻击活动出现的概率。当概率值超过警报阈值 λ 时, 该物联网系统内将很有可能出现后序 APT 攻击活动, 安全管理人员应产生警觉。

4 实验仿真与评估分析

为了对本章提出的 APTIS 方法进行详细、客观的评价, 实现了一个联邦学习原型系统, 运用先前已完成的工作^[24]中重建的 7 个典型 APT 攻击情境 (Op-Clandestine Fox、Hacking Team、APT on Taiwan、Tibetan and HK、Op-Tropic Trooper、Russian Campaign 和 Attack on Aerospace) 相关日志数据对 ICP-GRU 模型进行训练并评估其预测效果。Op-Clandestine Fox 指的是 IE 浏览器中的一个漏洞, 如果访问者使用 ie6 及以上版本, 恶意网站的所有者就可以完全访问访问者的电脑。进入电脑后, 黑客可以进行一些恶意活动, 比如安装应用程序, 甚至将

表 6 日志类涉及的参数

Table 6 Log instance features extracted from log records

编号	日志类型	参数名	描述
1	日志 1-6	timestamp	日志生成的时间
2	日志 3	q_domain	DNS 请求域名
3	日志 3	r_ip	DNS 解析 IP 地址
4	日志 2、5、6	pid	16 位进程 ID
5	日志 5	ppid	16 位父进程 ID
6	日志 2、4、5、6	pname	进程名
7	日志 6	h_ip	主机 IP 地址
8	日志 6	h_port	主机端口号
9	日志 6	d_port	目标端口号
10	日志 6	d_ip	目标 IP 地址
11	日志 6	type	请求/响应
12	日志 1	get_q	GET 路径
13	日志 1	post_q	POST 路径
14	日志 1	res_code	响应代码
15	日志 1	h_domain	主机域名
16	日志 1	referer	请求 URI 的引用
17	日志 1	res_loc	重定向位置
18	日志 2	acct	访问策略
19	日志 2	objname	实体名

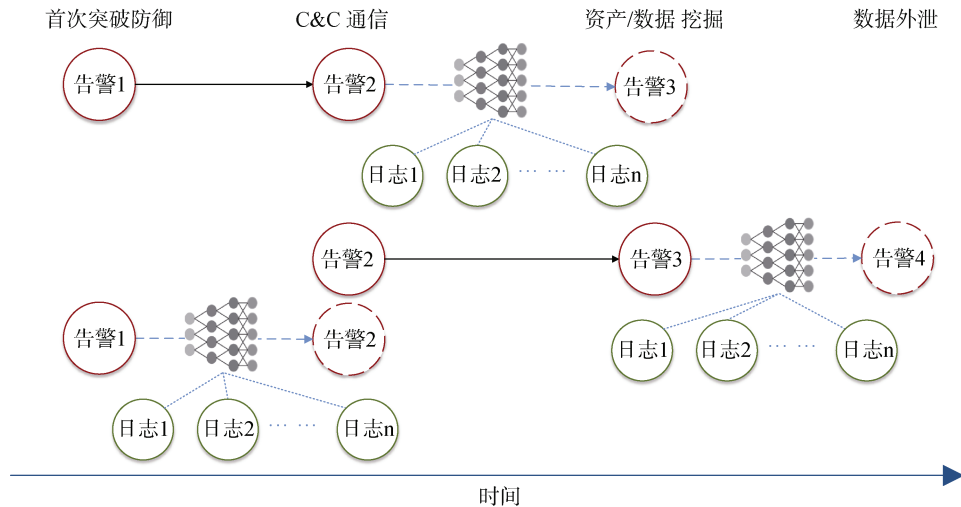


图 4 ICP-GRU 模型训练流程
Figure 4 ICP-GRU model training process

受感染的电脑当作自己的电脑使用。Hacking Team 是一个意大利的网络间谍公司, 曾经提供专门的间谍软件和服务给政府和执法机构。该公司的数据在 2015 年遭受了一次大规模的泄露, 导致其内部文档和源代码被公开。APT on Taiwan, Tibetan and HK 指的是针对台湾、西藏和香港地区的高级持久性威胁活动。Op-Tropic Trooper 是一个网络攻击活动的名称, 针对东南亚地区的政府和军事目标。该活动涉及使用定制的恶意软件和网络间谍工具进行监控和间谍活动。Russian Campaign 是指来自俄罗斯的网络攻击活动。俄罗斯被认为拥有一些高度先进和有组织的网络攻击团体, 参与了各种类型的网络攻击活动, 包括政治、经济和军事领域。Attack on Aerospace 指的是针对航空航天行业的网络攻击活动。航空航天领域是一个重要的目标, 因为它涉及敏感的技术、知识产权和国家安全。

4.1 实验数据获取与实验环境构建

4.1.1 数据集获取

目前, 对于典型的 APT 攻击情境尚无标准的日志数据集和告警数据集, 本实验室选择使用在之前工作中重建的 APT 攻击情境和带标记的日志实例作为实验数据, 对 APTIS 方法的预测效果进行评估分析。表 7 给出了实验数据中包含的 APT 攻击情境数据集及各个数据集的细节内容。

通过重建每个典型的 APT 攻击情境, 收集其攻击过程中产生的日志数据和触发的告警信息便可得到本实验需要的 7 个 APT 攻击活动数据集。这些 APT 场景是由不同的 APT 攻击团队设计和发起的, 不同的攻击团队利用到了不同的漏洞, 采取了不同的攻

击策略。本文实验选择使用属于不同 APT 攻击情境的多个数据集, 将其中一部分数据样本作为 ICP-GRU 模型的训练数据, 另一部分数据样本作为预测检验样本, 可充分验证本文提出的 APT 攻击活动预测方法对不同 APT 攻击情境的预测效果。

表 7 实验中生成的数据集

Table 7 Datasets used in our experiments			
数据集名称	数据格式	训练样本数量	检验样本数量
Op-Clandestine Fox	19 × 1	5600	1200
Hacking Team	19 × 1	5800	1240
APT on Taiwan	19 × 1	5750	1232
Tibetan and HK	19 × 1	6300	1350
Op-Tropic Trooper	19 × 1	6100	1307
Russian Campaign	19 × 1	7800	1671
Attack on Aerospace	19 × 1	4900	1050

4.1.2 实验环境的设置

为了使仿真实验环境尽可能的贴合真实物联网系统的特性, 本实验利用 4 台相同的主机(Intel Core i7-8550u 2.53GHz CPU, 16GB RAM)部署了一个基于边缘计算的联邦学习系统。具体的实验环境如图 5 所示。第一台主机可作为安全服务云, 其他 3 台主机作为边缘服务器。在 3 台主机上部署了相同的计算资源的原因是由于还在其中上分别设置了 8 台虚拟机作为虚拟物联网终端设备(充当虚拟物联网设备的虚拟机中均部署了 Windows Embedded Compact 操作系统), 共 24 台物联网终端设备。每台边缘服务器主机中一半的计算资源被虚拟物联网终端设备所占用, 每台边缘服务器拥有 8GB 内存资源而每台虚拟物联

网络端设备仅拥有 1GB 内存资源。这样的计算资源分配方式和真实的物联网系统环境是非常相似的。其中硬件具备足够的计算能力以及存储能力, 支持所选择的边缘计算框架的要求。日志利用基于 Louvain 方法的日志群落挖掘, 区分恶意与良性日志群落。数据传输使用 HTTP/HTTPS, 提供加密和安全性。本文提出的 APTPIS 方法就是为了解决计算资源受限的物联网终端设备对于预测模型训练的困难。虽然本实验没有实现真实的物联网系统, 但是每个虚拟物联网终端设备仅具有少量的计算资源并没有对其进行完备的安全加固, 可以充分地模拟出物联网系统在应用 APTPIS 时的预测效果。

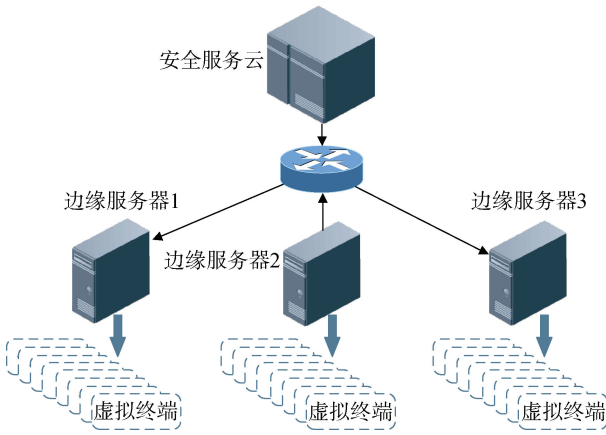


图 5 仿真实验环境

Figure 5 Simulation experiment environment

为了实现 APTPIS 方法中的联邦学习训练过程, 可将带有标签的日志实例根据攻击情境分配给各虚拟物联网终端设备并将 APT 攻击情境分发至各边缘服务器。随后, 虚拟物联网终端设备可将日志实例中的训练样本输入本地的 ICP 模型, 完成本地训练后再将提取出的干扰后的数据特征传送至边缘服务器中的 GRU 模型, 逐级完成联邦学习训练过程。在联邦学习过程中本实验采用了以下参数配置: 在本地训练中设定了 $E=20$ 的训练轮数; 将初始的训练速率设置为 $\delta=0.1$ 。经过多轮迭代训练后的 ICP-GRU 模型可用于预测虚拟物联网设备中的检验样本。

4.2 APT 攻击活动预测效果评估

4.2.1 预测准确性评估

为了评估 APTPIS 方法的预测准确性, 实验中设定了两个评估指标, 分别为 F_1 分值和误报率(FPR, False Positive Rate), 这两个指标的形式化描述如公式 7 和公式 8 所示。本文没有采用比较常用的 $Recall = TP/(TP + FN)$ 和 $Precision = TP/(TP + FP)$ 作为评估指标是因为这两个指标在某些情况下是存

在互斥性的, 需要一个调和平均的方法平衡各自的缺陷^[24]。本部分主要通过变化警报阈值 λ 的情况下测量 F_1 分值和 FPR 的数值来分析此预测模型对于不同 APT 攻击情境预测的准确性。其中, λ 值越低则后序 APT 攻击活动被预测并向安全管理人员发出警告的可能性也就越大, 但同时出现误报率的可能性也会相应增加。因此, 需要通过实验结果分析得到合适的警报阈值 λ 。

$$F_1 = \frac{2Recall \cdot Precision}{Recall + Precision} \quad (7)$$

$$FPR = \frac{FP}{FP + TN} \quad (8)$$

APTPIS 方法对于 7 种典型的 APT 攻击情境的预测准确性结果如图 6 和图 7 所示。根据实验结果可知, 随着警报阈值 λ 的增加 F_1 和 FPR 都呈逐渐减小的趋势。这是由于 λ 值越低则日志实例被视为 APT 攻击活动前序活动所需的概率值也就越低, 随之更多的良性日志实例会被视为 APT 攻击活动前序活动。通过观察实验结果可知, 当警报阈值 λ 设为 0.75 时, 各种 APT 攻击情境的 F_1 值可大约保持在 0.8 以上且 FPR 不会超过 5%。

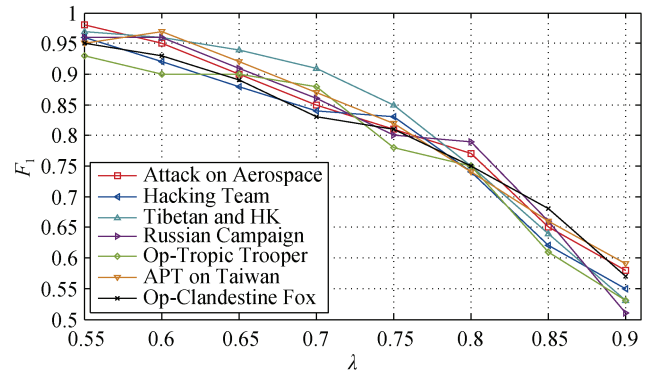


图 6 F_1 分值随警报阈值变化的结果

Figure 6 The F_1 score changing with alarm threshold

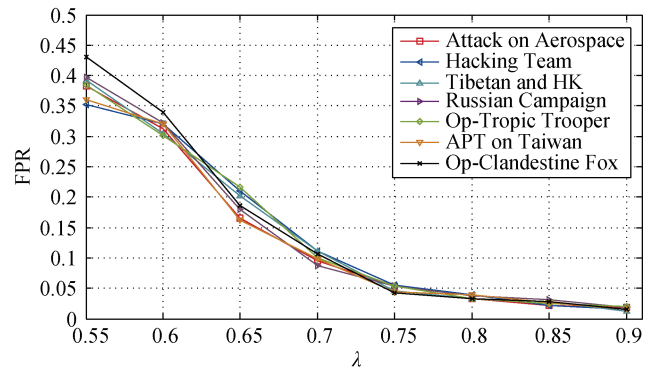


图 7 FPR 随警报阈值变化的结果

Figure 7 The FPR changing with alarm threshold

4.2.2 联邦学习效果评估

为了验证基于联邦学习的 ICP-GRU 模型随物联网终端设备数量变化时的预测效果, 本部分将虚拟物联网终端设备的数量在 3 至 24 之间进行变化, 观测不同 APT 攻击情境的预测效果。本实验将物联网终端设备与边缘服务器间的本地训练轮次设置为 20 次, 并将边缘服务器与云数据中心间的训练轮次设置为 5 次。这样, 每个物联网终端设备中的预测模型将经历 100 轮的迭代训练。设定这样的训练次数是比较充足的, 这是由于在集中式学习架构中可通过

98 轮的迭代完成对 ICP-GRU 模型的训练。本实验将数据集集中的数据随机分配到每个虚拟物联网设备中, 并且根据参与联邦学习模型的设备数量的变化观察 ICP-GRU 模型的预测效果。根据 APT 攻击情境的不同, 重复了 7 次上述实验过程。当警报阈值 λ 设为 0.75 时, 7 种 APT 攻击情境的预测效果如图 8 所示。通过分析实验结果可以发现, 随着参与到联邦学习的物联网终端设备数量的增加 FPR 值会明显降低, 但同时不会对 F_1 分值造成巨大的影响。从而可以证明联邦学习的方式可以有效降低预测模型的误报率。

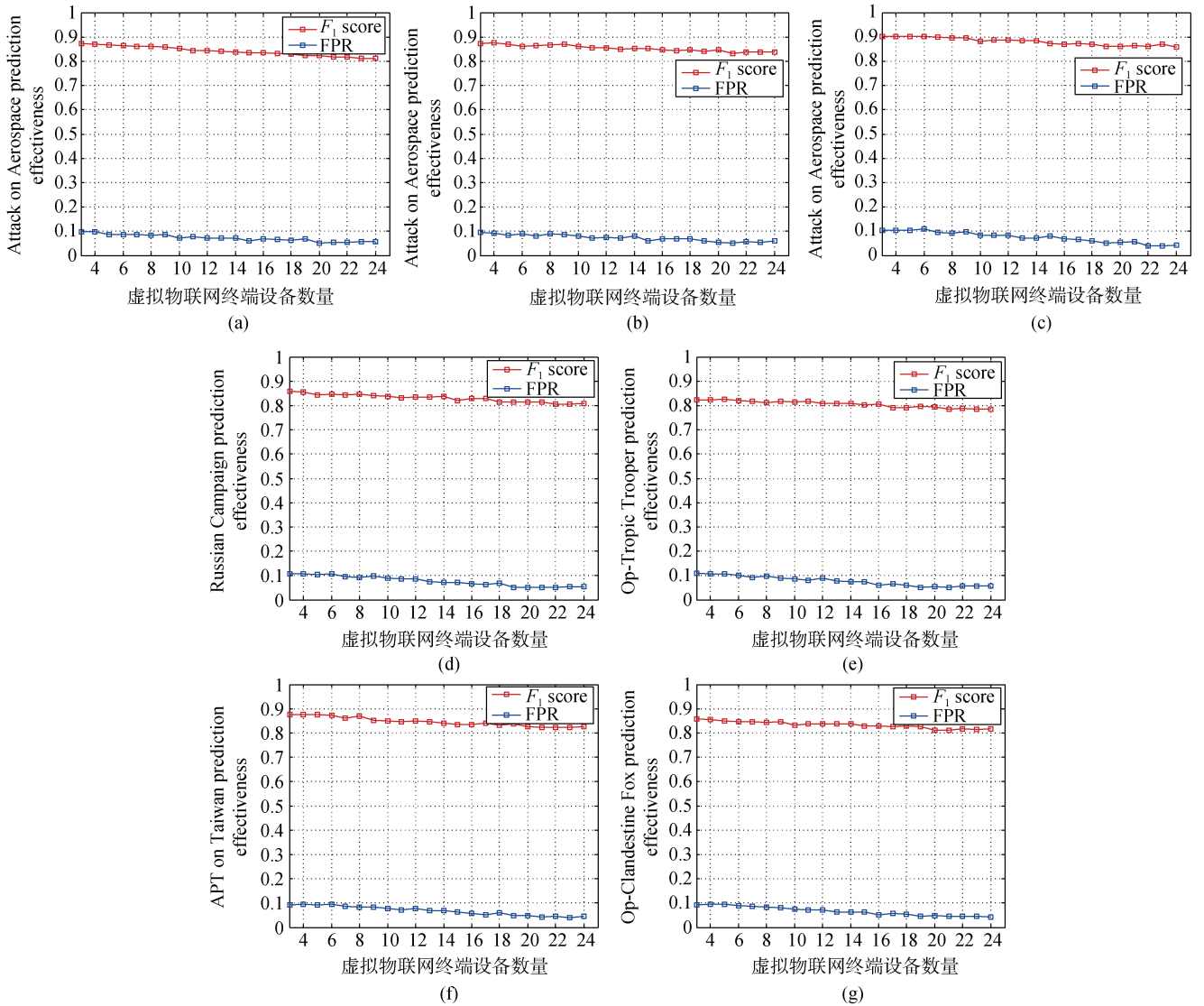


图 8 F_1 分值和 FPR 随虚拟物联网终端设备数量的变化

Figure 8 The F_1 score and FPR changing with IoT devices quantity

(注: (a)Attack on Aerospace (b)Hacking Team (c)Tibetan and HK (d)Russian Campaign (e)Op-Tropic Trooper (f)APT on Taiwan (g)Op-Clandestine Fox)

本文提出的 APTPIS 方法由于为在模型过程中没有向边缘服务器和云数据中心发送本地训练数据, 从而更好的保护了物联网终端设备中的数据隐私。然而, 联邦学习过程与集中式学习相比仍然具有一

定的缺陷和局限性, 例如其预测准确性不可避免的会稍劣于集中式学习方法。为了准确度量出这种差距, 可将完整的训练数据分配至 6、12、18、或 24 个虚拟物联网设备中对 ICP-GRU 模型重新进行训

练。以其中一台边缘服务器作为示例, 在实验过程中, 各设备被攻击的情况如表 8 所示。

表 8 设备被攻击的情况

设备	1	2	3	4	5	6	7	8
Op-Clandestine	●	●	○	●	○	○	○	○
Fox	○	●	○	○	○	○	●	○
Hacking Team	○	●	●	○	●	○	○	○
APT on Taiwan	●	○	●	●	○	○	●	○
Tibetan and HK	●	●	●	○	●	○	●	○
Op-Tropic	○	○	○	●	○	○	●	○
Trooper	○	○	○	●	○	○	●	○
Russian Cam-	○	○	○	●	○	○	●	○
paign	○	○	○	●	○	○	○	○
Attack on	○	○	○	○	○	○	○	○
Aerospace	○	○	○	○	○	○	○	○

表 9 给出了不同数量物联网终端设备的联邦学习模型与集中式学习模型的预测准确性对比。

表 9 联邦学习与集中式学习模型预测效果的对比

评估指标	集中学习	联邦学习			
		6 个终端	12 个终端	18 个终端	24 个终端
F_1	89.8%	88.7%	86.5%	84.5%	81.4%
FPR	4.85%	4.87%	4.89%	4.94%	4.96%

测量值 F_1 值和 FPR 为当报警阈值 λ 设为 0.75 时, 不同学习模型对于 7 种 APT 攻击情境的平均准确性。通过对比实验结果可知, 与集中式学习模型相比, 随着参与到联邦学习过程中物联网终端设备数量的增加 F_1 值呈小幅减小趋势, FPR 值未出现明显变化。 F_1 值在联邦学习过程中呈现出的损失主要是由于 ICP-GRU 模型训练数据中增加了干扰噪声, 使得模型预测的准确性有所下降, 可以通过调整报警阈值 λ 进行补偿。这种预测准确度损失, 作为保证物联网终端设备数据隐私的代价也是可以接受的。

GRU 模型与其他的模型进行了对比, 对比实验结果如下图 9 所示。MLP 没有显示记忆性能, 适合处理独立和静态的数据。RF 模型适用于结构化数据, 不适合处理序列数据。GCN 模型与 GRU 模型都可进行预测工作, 但泛化能力相对较弱。

5 结论

本文提出了一种基于联邦学习的 APT 攻击活动预测方法(APTPIS), 并给出了一种可将该方法应用

于物联网系统边缘计算架构。该方法通过分布式学习的方式实现了对于系统中出现的 APT 攻击活动演化过程的学习, 可用于预测系统中出现后序 APT 攻击活动的概率。运用该方法可以在保证终端设备数据隐私的前提下, 助力于整个系统的安全态势预测。评估分析实验结果表明, APTPIS 有效预测出了后序 APT 活动出现的概率, 且具有可接受的预测准确性和较低的误报率。

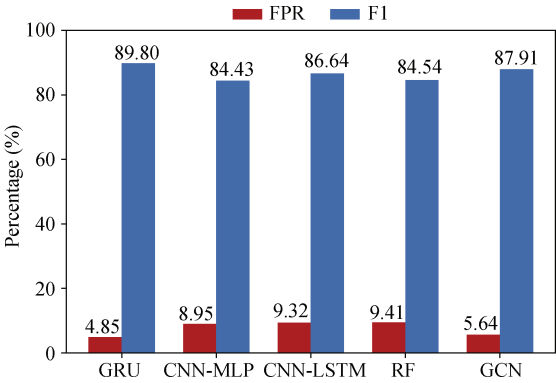


图 9 GRU 与其他模型对比

Figure 9 Comparison between GRU and other models

参考文献

[1] Han W J, Xue J F, Wang Y, et al. APTMalInsight: Identify and Cognize APT Malware Based on System Call Information and Ontology Knowledge Framework[J]. *Information Sciences*, 2021, 546: 633-664.

[2] Xuan C D. Detecting APT Attacks Based on Network Traffic Using Machine Learning[J]. *Journal of Web Engineering*, 2021, 20(1): 171-190.

[3] Wilkens F, Ortmann F, Haas S, et al. Multi-Stage Attack Detection via Kill Chain State Machines[EB/OL]. 2021: 2103.14628. <https://arxiv.org/abs/2103.14628v1>.

[4] Luo X J, Wu Y C, Xiao X K, et al. Feature Inference Attack on Model Predictions in Vertical Federated Learning[C]. *2021 IEEE 37th International Conference on Data Engineering*, 2021: 181-192.

[5] Rajagopal S, Kundapur P P, Hareesha K S. A Stacking Ensemble for Network Intrusion Detection Using Heterogeneous Datasets[J]. *Security and Communication Networks*, 2020, 2020(1): 4586875.

[6] Deng Y R, Yin X X, Hu S L. Event-Triggered Predictive Control for Networked Control Systems with DoS Attacks[J]. *Information Sciences*, 2021, 542: 71-91.

[7] Amaizu G C, Nwakanma C I, Bhardwaj S, et al. Composite and Efficient DDoS Attack Detection Framework for B5G Networks[J]. *Computer Networks*, 2021, 188: 107871.

[8] Latif S, Zou Z, Idrees Z, et al. A Novel Attack Detection Scheme for the Industrial Internet of Things Using a Lightweight Random Neural Network[J]. *IEEE Access*, 2020, 8: 89337-89350.

[9] Wallace E, Feng S, Kandpal N, et al. Universal Adversarial Trig-

- gers for Attacking and Analyzing NLP[EB/OL]. 2019: 1908.07125. <https://arxiv.org/abs/1908.07125v3>.
- [10] Zhang Y H, Jia R X, Pei H Z, et al. The Secret Revealer: Generative Model-Inversion Attacks Against Deep Neural Networks[C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 250-258.
- [11] Zhang Q Z, Hu S T, Sun J C, et al. On Adversarial Robustness of Trajectory Prediction for Autonomous Vehicles[C]. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 15138-15147.
- [12] Du M, Wang K, Xia Z Q, et al. Differential Privacy Preserving of Training Model in Wireless Big Data with Edge Computing[J]. *IEEE Transactions on Big Data*, 2020, 6(2): 283-295.
- [13] Liao Z F, Hu W B, Huang J W, et al. Joint Multi-User DNN Partitioning and Task Offloading in Mobile Edge Computing[J]. *Ad Hoc Networks*, 2023, 144: 103156.
- [14] Chen S T, Gollakota A, Klivans A R, et al. Hardness of Noise-Free Learning for Two-Hidden-Layer Neural Networks[EB/OL]. 2022: 2202.05258. <https://arxiv.org/abs/2202.05258v3>.
- [15] Baksi A. Machine Learning-Assisted Differential Distinguishers for Lightweight Ciphers[M]. *Classical and Physical Security of Symmetric Key Cryptographic Algorithms*. Singapore: Springer Singapore, 2022: 141-162.
- [16] Zhu Y Q, Dong J S, Wang Y X. Optimal Accounting of Differential Privacy via Characteristic Function[EB/OL]. 2021: 2106.08567. <https://arxiv.org/abs/2106.08567v3>.
- [17] Kandpal N, Wallace E, Raffel C. Deduplicating Training Data Mitigates Privacy Risks in Language Models[EB/OL]. 2022: 2202.06539. <https://arxiv.org/abs/2202.06539v3>.
- [18] Qi Y H, Hossain M S, Nie J T, et al. Privacy-Preserving Blockchain-Based Federated Learning for Traffic Flow Prediction[J]. *Future Generation Computer Systems*, 2021, 117: 328-337.
- [19] Feldman V, McMillan A, Talwar K. Hiding among the Clones: A Simple and nearly Optimal Analysis of Privacy Amplification by Shuffling[C]. *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science*, 2022: 954-964.
- [20] Poongodi M, Bourouis S, Ahmed A N, et al. A Novel Secured Multi-Access Edge Computing Based VANET with Neuro Fuzzy Systems Based Blockchain Framework[J]. *Computer Communications*, 2022, 192: 48-56.
- [21] Fang M H, Cao X Y, Jia J Y, et al. Local Model Poisoning Attacks to Byzantine-Robust Federated Learning[EB/OL]. 2019: 1911.11815. <https://arxiv.org/abs/1911.11815v4>.
- [22] Zhang J L, Zhao Y C, Wang J Y, et al. FedMEC: Improving Efficiency of Differentially Private Federated Learning via Mobile Edge Computing[J]. *Mobile Networks and Applications*, 2020, 25(6): 2421-2433.
- [23] Cheng X, Zhang J L, Chen B. Cyber Situation Comprehension for IoT Systems Based on APT Alerts and Logs Correlation[J]. *Sensors*, 2019, 19(18): 4045.
- [24] Zimba A, Chen H S, Wang Z S, et al. Modeling and Detection of the Multi-Stages of Advanced Persistent Threats Attacks Based on Semi-Supervised Learning and Complex Networks Characteristics[J]. *Future Generation Computer Systems*, 2020, 106: 501-517.



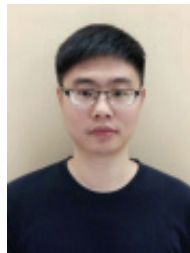
成翔 于 2021 年在南京航空航天大学计算机科学与技术专业获得博士学位。现任扬州大学信息工程学院实验师, CCF 会员。研究领域为网络与系统安全、网络安全态势感知、APT 攻击检测。Email: huozhai9527@126.com



匡苗苗 于 2023 年起在扬州大学软件工程专业攻读硕士学位。研究领域为网络与系统安全、网络安全态势感知、APT 攻击检测。Email: 515087254@qq.com



张佳乐 于 2021 年在南京航空航天大学计算机科学与技术专业获得博士学位。现任扬州大学信息工程学院讲师, CCF 会员。研究领域为人工智能安全、联邦学习、边缘计算。Email: jjialezhang@yzu.edu.cn



陈玮彤 于 2021 年在南京师范大学地图学与地理信息系统专业获得博士学位。现任扬州大学信息工程学院讲师, CCF 会员。研究领域为数据安全、地理信息安全。Email: wtchen@yzu.edu.cn



李云 现任扬州大学信息工程学院教授, CCF 会员、江苏省计算机学会高级会员, 江苏省计算机学会计算机应用专委会副主任委员。研究领域为数据挖掘、机器学习、并行与分布式处理。Email: liyun@yzu.edu.cn



杨宏宇 于 2003 年在天津大学计算机应用技术专业获得博士学位。现任中国民航大学安全科学与技术学院、计算机科学与技术学院教授, CCF 会员。研究领域为网络与系统安全。Email: hyyang@cauc.edu.cn