

隐私保护的加密流量检测研究

张心语, 张秉晟, 孟泉润, 任奎

(浙江大学网络空间安全学院, 浙江 杭州 310000)

摘要: 现有的加密流量检测技术缺少对数据和模型的隐私性保护, 不仅违反了隐私保护法律法规, 而且会导致严重的敏感信息泄露。主要研究了基于梯度提升决策树 (GBDT) 算法的加密流量检测模型, 结合差分隐私技术, 设计并实现了一个隐私保护的加密流量检测系统。在 CICIDS2017 数据集下检测了 DDoS 攻击和端口扫描的恶意流量, 并对系统性能进行测试。实验结果表明, 当隐私预算 ϵ 取值为 1 时, 两个数据集下流量识别准确率分别为 91.7% 和 92.4%, 并且模型的训练效率、预测效率较高, 训练时间为 5.16 s 和 5.59 s, 仅是 GBDT 算法的 2~3 倍, 预测时间与 GBDT 算法的预测时间相近, 达到了系统安全性和可用性的平衡。

关键词: 隐私保护; 加密流量检测; 梯度提升决策树; 差分隐私

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2021057

Study on privacy preserving encrypted traffic detection

ZHANG Xinyu, ZHANG Bingsheng, MENG Quanrun, REN Kui

School of Cyber Science and Technology, Zhejiang University, Hangzhou 310000, China

Abstract: Existing encrypted traffic detection technologies lack privacy protection for data and models, which will violate the privacy preserving regulations and increase the security risk of privacy leakage. A privacy-preserving encrypted traffic detection system was proposed. It promoted the privacy of the encrypted traffic detection model by combining the gradient boosting decision tree (GBDT) algorithm with differential privacy. The privacy-protected encrypted traffic detection system was designed and implemented. The performance and the efficiency of proposed system using the CICIDS2017 dataset were evaluated, which contained the malicious traffic of the DDoS attack and the port scan. The results show that when the privacy budget value is set to 1, the system accuracy rates are 91.7% and

收稿日期: 2020-07-01; 修回日期: 2020-09-22

通信作者: 张秉晟, bingsheng@zju.edu.cn

基金项目: 国家自然科学基金 (62032021, 61772236); 浙江省重点研发计划 (2019C03133); 阿里巴巴-浙江大学前沿技术联合研究所, 浙江大学网络空间治理研究所, 创新创业团队浙江省引进计划 (2018R01005); 移动互联网系统与应用安全国家工程实验室 2020 开放课题

Foundation Items: The National Natural Science Foundation of China (62032021, 61772236), Zhejiang Key R&D Plan (2019C03133), Alibaba-Zhejiang University Joint Institute of Frontier Technologies, Research Institute of Cyberspace Governance in Zhejiang University, Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2018R01005), 2020 Open Project of the National Engineering Laboratory of Mobile Internet System and Application Security

论文引用格式: 张心语, 张秉晟, 孟泉润, 等. 隐私保护的加密流量检测研究[J]. 网络与信息安全学报, 2021, 7(4): 101-113.
ZHANG X Y, ZHANG B S, MENG Q R, et al. Study on privacy preserving encrypted traffic detection[J]. Chinese Journal of Network and Information Security, 2021, 7(4): 101-113.

92.4% respectively. The training and the prediction of our model is efficient. The training time of proposed model is 5.16 s and 5.59 s, that is only 2-3 times of GBDT algorithm. The prediction time is close to the GBDT algorithm.

Keywords: privacy-preserving, encrypted traffic detection, gradient boosting decision tree, differential privacy

1 引言

随着互联网应用、物联网设备的日益发展和普及,大数据驱动的智能技术,如图像分类、推荐算法、语音识别、无人驾驶等推动社会信息化时代的变革。而大数据平台收集的数据通常包含用户的隐私和敏感信息,深入分析这些数据可能会导致用户隐私信息被挖掘和滥用,在大数据被合理利用的同时,如何保证数据安全和用户隐私,是政府、企业和个人面临的新的挑战。隐私保护的加密流量检测是一个重要研究方向。

“棱镜”计划和各类网络监控事件在全球范围内曝光,这使用户的隐私保护意识逐渐增强。为了保证通信安全和隐私,越来越多的网络流量采用数据加密技术,如安全套接字协议等,以此确保数据在流量包传输过程中不会被随意窃听。但同时,越来越多的攻击者使用加密流量的方式隐藏自己的行踪,防止被检测和识别。针对恶意加密流量的检测主要有两种方式:先解密后检测和不解密直接检测。

网管设备通常采用先解密流量的方式进行恶意流量检测,但这种检测方式需要先解密流量、检测流量,再重新加密,其中,解密操作的实现难度较高,而且会消耗大量计算资源,导致网络吞吐量和时延增大,降低设备的可用性。此外,解密加密流量的操作违背了对网络流量进行加密的初衷,解密后的流量可能以日志或临时存储文件的形式被泄露,大大增加了设备受攻击的可能性。解密加密流量还可能面临违反隐私保护法律法规的风险。

学术界提出了在不解密的情况下,对恶意加密流量进行检测的方案,即基于传统的统计分析和机器学习算法。由于加密流量技术只对载荷信息加密,而流特征仍然以明文形式传输,因此可以从网络流量包中提取数据特征,使用机器学习来训练并区分恶意流量和良性流量的行为特征,从而进行准确识别。这类方法具有较高的识别性

能和速度,可以较好地适用于各种网络环境,因此被大量应用于企业内部进行恶意流量检测,如思科公司的 Joy 系统^[1]。

为了应对大数据时代下的隐私泄露风险与挑战,国内外标准化组织在信息技术的不同领域给出了“隐私”的定义,并针对网络环境中的隐私保护制定了相关标准。欧盟于 2016 年通过的《通用数据保护规范》^[2],从个人信息的采集到信息的传输和使用,直到销毁,对信息的全周期有很明确的行为规范要求。个人信息采集时,应实行“最少采集”原则,不能非法采集数据;采集信息的目的达到后,需在一定期限之内予以销毁。对于违反规范的机构,政府将处以法律责任追究、罚款,甚至处以刑事责任。美国 2020 年实施的《加利福尼亚州消费者隐私法案》^[3]为消费者创建了访问权、删除权、知情权等一系列消费者隐私权利,并要求企业必须遵循相关义务。对违反隐私保护要求的企业,政府有权征收处罚。

国内也针对个人信息保护出台了相关法律法规。于 2017 正式施行的《网络安全法》^[4]系统性提出网络空间治理的法律法规,特别明确和强调了个人信息保护方面的要求,明确并强化了对公民个人信息安全的保护。新颁布的 2020 版《个人信息安全规范》^[5]对个人信息的采集、存储、使用都做了明确规范,并规定了个人信息主体具有查询、更正、删除、撤回授权同意、注销账户、获取个人信息副本权利等,对后续个人信息保护工作的开展将产生深远影响。全世界各组织对隐私保护标准的严格规范和定义,更突显了数据隐私保护的重要性。隐私保护不仅关系到个人隐私安全问题,而且会对机构的数据安全问题乃至国家的网络安全产生重大影响。

在大规模网络加密流量检测系统中,互联网服务供应商和企业面临着重大挑战。一方面,由于在复杂的网络环境中,加密流量分析变得越来越困难,这些组织往往需要将此类任务及必要的网络数据外包给第三方分析机构。另一方面,这

些组织通常不愿意与第三方共享网络流量数据。这是由于此类数据中包含用户的敏感信息,虽然检测过程中不需要收集用户发送的明文信息,但大量流量包的源 IP 地址结合目的 IP 地址,可以定位到某个具体用户;从网络流量数据中也可以推断出用户的网络配置信息等;攻击者利用这些信息进行拒绝服务攻击等。对于这类恶意攻击,检测系统难以控制,当前方案下,企业难以对用户个人信息的采集、传输和使用全过程进行保障,违反了《通用数据保护规范》和《个人信息安全规范》中对个人隐私保护的规定,一旦流量包被恶意利用,企业将面临高额罚款,甚至法律追究。

在上述情况下,流量特征匿名化的加密流量检测技术引起了极大关注,传统流量数据匿名化技术通过对原始数据进行变形处理,攻击者无法直接进行属性泄露攻击,常见的数据变形处理方法包括哈希、截断、替换、置换、IP 地址混淆、隐藏等。此类方法的优势在于处理速度快、计算成本低。但变形后的数据由于丢失了部分性质,无法进行复杂的网络流量分析。因此,如何权衡网络流量数据匿名化后的安全性和实用性问题,成为亟待解决的问题。设计一套隐私保护的加密流量检测算法,有助于企业在当前错综复杂的网络环境下,实现对用户个人信息的全过程保护,更好地保护企业流量数据资产,并且严格遵守隐私保护法律法规。

本文提出了一种基于差分隐私的加密流量检测技术,该技术既能保证加密流量数据包的隐私性,又能以较高的准确率实现恶意流量检测,主要贡献如下。

(1) 设计了一种隐私保护的加密流量检测系统,采用基于差分隐私的梯度提升决策树算法,规定了严格的灵敏度范围并有效地分配噪声,引入了基于梯度的数据过滤算法和几何叶剪枝算法。

(2) 实现了隐私保护的加密流量检测系统并进行实验测试,实验采用 CICIDS2017 数据集进行测试,在两种攻击的实验数据集上都取得了较好的效果,当 $\epsilon=1$ 时,恶意流量识别准确率分别可以达到 91.7% 和 92.4%。

2 方法介绍和安全模型

本文采用的加密流量检测是梯度提升的决策树算法,该算法在恶意流量识别领域能达到较高的准确率。隐私保护的安全模型方面,本文采用了差分隐私的扰动机制。本文系统分为流量数据采集、数据预处理、模型训练和评估验证这 4 个模块。本文搭建了一个完整的基于差分隐私的流量检测模型。

2.1 流量异常检测的概念

流量检测技术被广泛应用于完善入侵检测系统、提升网络通信服务质量等。传统的网络流量异常检测,是通过分析流量的地址、端口、字节等数据来监控网络流量,及时发现网络攻击行为,如通过僵尸网络实现的分布式拒绝服务 (DDoS, distributed denial of service) 攻击。

僵尸网络^[6]是从网络蠕虫、木马病毒、后门工具等传统恶意代码形式发展而来的攻击方法,黑客可以入侵网络空间中大量僵尸主机,通过远程命令和控制,实现大规模网络攻击,如分布式拒绝服务攻击和大量垃圾邮件发送、信息盗取等。其中,DDoS 攻击的目的是通过干扰正常的网络环境造成计算机服务不可用,典型形式是流量溢出,攻击者利用大量连接请求来消耗被攻击网络的服务资源,从而使被攻击者无法处理其他合法用户的请求。

端口扫描 (port scan) 是一种用于探测本地或远程端口是否开放的技术,黑客将其恶意利用为一种攻击探测手段,通过向目标主机发送数据包,并记录目标主机的回应情况来攻击目标主机。由于一个开放的端口即为通信通道,黑客可以了解目标主机存在哪些弱点,如开放的端口、运行的服务等,并将其作为入侵检测的突破点。隐蔽扫描技术能成功绕过防火墙,入侵检测系统,获取目标主机端口信息。

随着用户隐私保护意识的增强与加密协议 [如传输层安全 (TLS, transport layer security) 协议] 的广泛应用,加密流量急剧增加。恶意程序 (如僵尸网络、木马病毒),越来越多地通过加密技术来绕过防火墙和入侵检测系统,进行 DDoS 攻击、隐蔽扫描、窃取机密信息。传

统的流量异常检测技术在识别 DDoS 攻击、port scan 等异常流量时,准确性有所下降。因此,研究者^[7]采用机器学习技术,实现一种加密流量的异常流量识别方法。

2.2 加密流量检测技术

本节主要介绍基于机器学习的加密流量检测技术,该技术^[1]需要从实时流量和数据包中提取数据特征。恶意加密流量和正常的加密流量在一些特征上存在差异,主要表现在四类特征:数据包的时间序列特征、数据包的字节分布情况、数据包的报头特征、TLS 相关特征,并且这些特征的差异性不受流量加密的影响。在流量采集过程中,可以先分别提取这些流量特征^[1]。

(1) 数据包长度和时间序列:流量包的长度(以字节为单位)、流量包到达的时间间隔序列(以毫秒为单位)、流量包的发送方向。

(2) 数据包的字节分布情况:流量包中字节分布的熵、字节分布的均值和标准差。

(3) 数据包报头特征:流量包发送和到达的端口号、协议类型、流量包的源 IP 地址、流量包的目的 IP 地址。

(4) TLS 相关特征:TLS 扩展类型、客户端或服务端使用的密码套件列表、TLS 版本号、客户端的公钥长度。

统计每个加密流量包的上述特征后,常用的机器学习检测算法有逻辑回归算法^[8]、决策树算法[如随机森林算法^[9]、梯度提升决策树(GBDT, gradient boosting decision tree)算法^[10]]等。

本文主要使用梯度提升决策树算法对加密流量的检测进行训练,Boosting 是一种可以将弱学习器提升为强学习器的算法,其工作机制是先从初始数据集训练出一个基学习器(典型的如决策树),再根据基学习器预测的结果对训练样本进行调整,使基学习器分类错误的样本在后续训练过程中获得更多关注,再基于调整后的样本分布来训练下一个分类器。重复这一过程,直到基学习器的数量达到预先设定的值,再将所有基学习器进行加权相加。

GBDT 算法是基于 Boosting 的一种集成学习算法,其本质是下一轮的基学习器是建立在上一轮基学习器的损失函数的梯度下降方向,将多棵

决策树利用加法模型,组合成一棵梯度提升的决策树。令训练数据集 D 表示为 $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, $(\mathbf{x}_i, y_i)$ 为第 i 个输入的流量包样本, N 为训练数据集中实例的总数,其中 $\mathbf{x}_i$ 为输入的特征向量, y_i 为流量包标签, $A = \{a_1, a_2, \dots, a_m\}$ 为样本属性集合, m 表示每个样本的特征维度数)。在第 k 轮迭代中,假设给定了损失函数 $L(\theta)$,目的是训练出决策树 f_k ,使损失函数的近似值最小^[11]。

$$\sum_{i=1}^n L(y_i, f_k(\mathbf{x}_i)) = \sum_{i=1}^n \left[g_i f_k(\mathbf{x}_i) + \frac{1}{2} f_k^2(\mathbf{x}_i) \right] \quad (1)$$

g_i 是损失函数的一阶梯度值,决策树从根节点开始构建,直到达到最大深度。假设分裂后左右节点的实例集为 D_L 和 D_R 。令 $D = D_L \cup D_R$,则分裂产生的信息增益 $G(D_L, D_R)$ 为

$$G(D_L, D_R) = \frac{(\sum_{i \in D_L} g_i)^2}{|D_L| + \lambda} + \frac{(\sum_{i \in D_R} g_i)^2}{|D_R| + \lambda} \quad (2)$$

此时,GBDT 遍历所有特征,选择信息增益最大化的划分点。若当前节点不满足划分要求,如已经到达最大深度或增益小于零,则当前节点成为叶节点,并且最佳叶节点的值由式(3)得出。

$$V(D) = - \frac{\sum_{i \in D} g_i}{|D| + \lambda} \quad (3)$$

梯度提升决策树算法在加密流量识别任务上具有很强的分类能力。

2.3 系统架构

根据加密流量的特点,本系统可采用串接和旁路两种模式部署^[12]。串接部署组网简单,无须额外增加接口,并且由于防护设备可以实时监控双向流量,在个别攻击防护上优于旁路部署。但在组网结构复杂的场景下,难以使用串接模式部署,可以采用旁路部署模式,避免设备直路部署可能带来的链路短时中断。旁路部署的优点是能够保证原有组网不被破坏,同时引入了流量流向改变技术,通过引流、回注等一系列手段来控制流量的走向,从而实现对异常流量的处理。

本文根据上述加密流量检测算法,搭建了加密流量检测系统,如图 1 所示,包括数据采集、数据预处理、模型训练和评估验证 4 个模块。

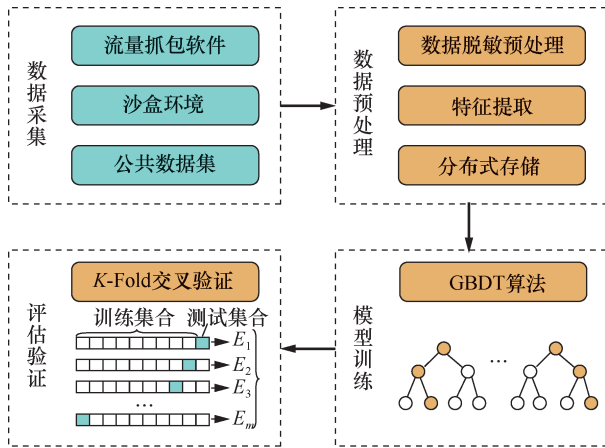


图1 加密流量检测系统架构

Figure 1 The system architecture of encrypted traffic detection

步骤 1 数据采集。加密流量包可以利用沙盒环境收集，用户向其提交可疑执行文件，每个提交的文件运行一段时间后，系统捕获、收集并存储完整的加密数据包，这些流量将被标记为恶意加密流量。正常加密流量可以通过 Wireshark 抓包软件直接从公共网络捕获，再设置黑名单过滤掉恶意 IP 地址的流量，从而得到良性流量包。为了提高可信度，可以直接采用公开数据集 ISCX、Botnet、ISCX VPN-non VPN 等。

步骤 2 数据预处理。由于完整的流量数据包较大，需要先对数据进行降维预处理，再从降维后的流量数据中分别提取 2.2 节中提到的 4 类数据特征。

步骤 3 模型训练。输入提取的流量特征，采用 2.1 节提到的 GBDT 算法对加密流量分类模型进行训练。

步骤 4 评估验证。为了防止训练过拟合，本系统采用 K-Fold 交叉验证的方法评估模型训练效果。首先将原始数据分为 K 组，其中每个子集分别做一轮验证集，同时剩余 K-1 组作为训练集，如此得到 K 个模型，经验证集验证得到 K 个均方误差，所有均方误差的平均值是最终的交叉验证误差值。

本文系统分为数据采集、数据预处理、模型训练和评估预测 4 个模块。在流量数据采集、预处理过程中，数据集将受到保护，不会被攻击者获取；在分布式存储和模型训练、评估预测时，将不受保护，攻击者可能通过训练完成的模型，

利用成员推理攻击^[13]，反向推断出训练集中所包含的数据或其特征，泄露数据集的隐私信息。但在本文系统中，数据在训练过程中引入了噪声，对训练的模型进行了保护，也对输出结果加入了拉普拉斯扰动，黑客无法通过模型准确推断原本训练集的特征，也无法挖掘训练数据集的流量信息和用户隐私。

2.4 安全模型

传统衡量数据隐私的通用标准分别是 K-anonymity、L-diversity 和 T-closeness，这 3 个标准从不同角度衡量了隐私数据泄露的风险。K-anonymity^[14]要求对于任意一行记录，其所属的等价类内的记录数量不小于 K。该标准能够保护数据的身份信息，但其缺点是无法防止属性泄露的风险，攻击者可以通过同质属性攻击和背景知识攻击两种方式来获取个体的敏感属性信息。L-diversity^[15]进一步要求，在一个等价类中所有记录对应的敏感属性的集合需要包含 L 个“良表示”。这一标准保证了属性取值的多样性，但无法衡量不同属性值的分布，因此在衡量属性泄露风险上仍有不足之处。T-closeness^[15]则要求等价类中的敏感属性分布与整体数据表中敏感属性的分布的距离小于 T，该约束减弱了间接标识符列属性与特定敏感信息的联系，进而减少了攻击者通过敏感属性的分布信息进行属性泄露攻击的可能性，但其缺点是会导致信息在一定程度上发生损失。

本文使用的隐私保护标准是由 Dwork 等^[16]于 2006 年提出的差分隐私，它给出了对个人隐私泄露的数学定义，并且具有可证明的隐私保证，可以在最大化查询结果可用性的同时，保证单个用户隐私泄露不超过预先设定的 ϵ 。

定义 1 ϵ 差分隐私。D 与 D' 是一对相邻数据集，两者只相差一个元素，给定一个随机算法 A 及其值域 $\text{Ran}(A)$ ，若算法 A 在 D 与 D' 上任意输出结果 S ($S \in \text{Ran}(A)$) 满足如下不等式，则算法 A 满足 ϵ 差分隐私。

$$\Pr[A(D) = S] \leq \exp(\epsilon) \cdot \Pr[A(D') = S] \quad (4)$$

其中， ϵ 表示隐私预算， ϵ 越小，对用户的隐私保护强度越高。差分隐私通常采用的噪声机制包括拉普拉斯机制、高斯机制和指数机制。噪声机

制需要根据函数的敏感度添加噪声。

定义 2 敏感度。对于查询函数 f ，其形式为 $f:D \rightarrow R^d$ ， R^d 是查询返回结果，函数 f 的敏感度定义为

$$S(f) = \max_{D, D'} \|f(D) - f(D')\| \quad (5)$$

其中， $\|f(D) - f(D')\|$ 是 $f(D)$ 与 $f(D')$ 之间的曼哈顿距离。

定理 1 拉普拉斯机制。添加的噪声大小是根据敏感度 $S(f)$ 来计算的，通过在返回的准确查询结果上加入一个满足 $\text{Lap}\left(0, \frac{S(f)}{\varepsilon}\right)$ 分布的噪声，使返回结果符合差分隐私标准。

定理 2 指数机制。对于效用函数 u ，若指数机制 F 以正比于 $e^{\frac{\varepsilon u(D, r)}{2S(u)}}$ 的概率输出结果 r ，其中 $S(u)$ 为效用函数的敏感度，则指数机制 F 满足 ε 差分隐私。

McSherry 等^[17]提出了差分隐私的两个重要性质：顺序合成性质和平行合成性质。

性质 1 顺序合成。对于 k 个算法组成的算法集合 $A = \{A_1, \dots, A_k\}$ ，其中 A_i 算法满足 ε_i 差分隐私，当它们作用于同一个数据集时，算法 A 可以提供 $\sum_{i=1}^k \varepsilon_i$ 差分隐私。

性质 2 平行合成。对于 k 个算法组成的算法集合 $A = \{A_1, \dots, A_k\}$ ，其中 A_i 算法满足 ε_i 差分隐私，当它们作用于一个数据集 D 的不同子集 $\{D_1, \dots, D_k\}$ 上时，算法 A 可以提供 $\max\{\varepsilon_1, \dots, \varepsilon_k\}$ 差分隐私。

3 系统设计和安全性分析

3.1 隐私保护的加密流量检测算法

本文结合差分隐私的顺序合成和平行合成，参考文献[18]的两级提升框架，设计了如图 2 所示的隐私保护的加密流量检测的系统框架。在平行合成内部，使用数据集中不相交的数据子集训练生成多棵决策树，然后依次训练生成多个这样的合成。在平行合成内部和外部，分别采用差分隐私保证隐私性，给定敏感度边界设置和隐私预算分配方案。

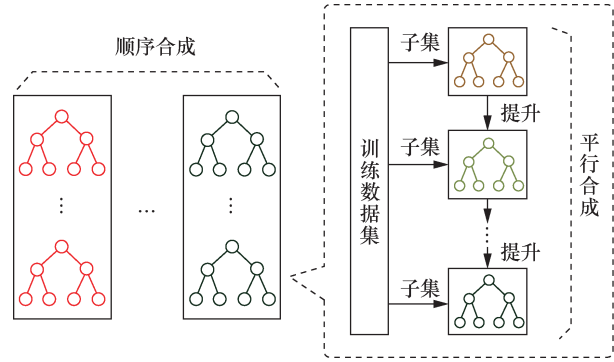


图 2 隐私保护的加密流量检测系统框架
Figure 2 The system architecture of privacy preserving encrypted traffic detection

3.2 敏感度边界

先前的研究对单棵差分隐私决策树的敏感度是通过函数输出范围来限制的，即函数的灵敏度不超过函数输出的最大值与最小值之差。为了避免训练数据集过大，导致信息增益函数 G 、节点值函数 V 的输出范围急剧增大，从而产生巨大的灵敏度，文献[18]根据严格的公式推导给出了 G 和 V 函数的灵敏度边界。

引理 1 令 $g_{\max} = \max_{i \in D} (g_i)$ ，则函数 G 的灵敏度 $S(G) \leq \frac{3\lambda + 2}{(\lambda + 1)(\lambda + 2)} g_{\max}^2$ 。

引理 2 令 $g_{\max} = \max_{i \in D} (g_i)$ ，则函数 V 的灵敏度 $S(V) \leq \frac{g_{\max}}{1 + \lambda}$ 。

由引理 1 和引理 2 可得，节点的敏感度与梯度绝对值（即 1 范数梯度）的最大值有关，因此，为了控制隐私预算，通常需要限制梯度的范围，但在 GBDT 算法中，梯度是由预测值与目标值之间的距离来计算的，限制梯度意味着间接改变了目标值，会产生巨大误差，因此，Li 等^[18]提出了一种基于梯度的数据过滤方法，通过在每次迭代中仅过滤训练数据集的一小部分来限制 1 范数梯度的最大值。

基于梯度的数据过滤 (GDF, gradient-based data filtering): 对于每一个样本 x_i ，其梯度初始值为 $g_i = \frac{\partial L(y_i, y)}{\partial y} \Big|_{y=0}$ ，令初始化过程中 1 范数梯度可能的最大值为 $g_{i \max}$ ，则 $g_{i \max} = \max_{y_i \in \{-1, 1\}} \left\| \frac{\partial L(y_i, y)}{\partial y} \Big|_{y=0} \right\|$ 。在训练过程中，当每轮迭代开始时，需要筛选删

除1范数梯度大于 g_{lmax} 的实例，即这些实例不在本轮迭代中训练，将其余实例作为输入构建新的差分隐私决策树。需要注意的是，被过滤后的实例仍然可能参与后续决策树的训练。使用这种基于梯度的数据过滤技术，可以确保使用实例的梯度不大于 g_{lmax} 。对于 GDF 算法，当输入为数据集 D 时，记输出为 $D' \leftarrow \text{GDF}(D)$ 。根据引理1和引理2，可以有如下推论。

推论1 通过在GBDT训练过程中添加GDF，可得到函数 G 和 V 灵敏度边界有如下关系。

$$\begin{cases} S(G) \leq \frac{3\lambda+2}{(\lambda+1)(\lambda+2)} g_{lmax}^2 \\ S(V) \leq \frac{g_{lmax}}{1+\lambda} \end{cases} \quad (6)$$

GDF 为所有决策树提供了相同的敏感度，由于梯度在不停迭代过程中会逐渐下降，因此随着不断迭代，可以获得更严格的敏感度范围。但实际上很难得出梯度下降的确切模式，且不合理地降低梯度阈值可能会导致巨大的精度损失。因此，需要使用一种新的算法在不同决策树训练过程中控制这种衰减效果。Li等^[18]经推导发现每棵树中的叶节点值将形成一个具有初始值为 g_{lmax} 、公比为 $1-\eta$ 的等比数列，并基于此提出了一种几何叶剪枝算法，可以限制叶节点的敏感度。

几何叶剪枝 (GLC, geometric leaf clipping): 一种自适应叶节点剪枝算法，用于实现叶节点上的敏感度衰减。当训练决策树迭代至第 k 轮时，应用拉普拉斯机制之前，需要先使用阈值 $g_{lmax}(1-\eta)^{k-1}$ 对叶子节点值进行剪枝，即 $V'_k = V_k \min(1, g_{lmax}(1-\eta)^{k-1} / |V_k|)$ ，这意味着若叶子节点值大于阈值，则将其值更新为阈值。当叶子节点输入为 V_k 时，该节点经过 GLC 剪枝后，输出值记为 $V'_k \leftarrow \text{GLC}(V_k)$ 。

推论2 结合 GDF 和 GLC，在第 k 轮中叶子节点的敏感度满足

$$S(V_k) \leq \min\left(\frac{g_{lmax}}{1+\lambda}, 2g_{lmax}(1-\eta)^{k-1}\right) \quad (7)$$

3.3 基于差分隐私的梯度提升决策树模型

本文针对加密流量检测，参考文献[18]中差分隐私的梯度提升决策树算法，提取捕获数据的8个特征。

第一步是利用算法1实现单棵差分隐私决策树的生成。

对叶节点的隐私预算 ϵ_l 和树内部节点的隐私预算 ϵ_{nl} 进行分配。

算法1 训练单棵树算法

输入 D_0 : 初始数据集; ϵ_l : 隐私预算;
max_depth: 最大深度; $A=\{a_1, a_2, \dots, a_m\}$: 属性集
输出 一棵满足 ϵ_l 差分隐私决策树 Tree

1) 函数 TreeGen(D_0, A)

2) 叶节点的隐私预算 $\epsilon_l := \frac{\epsilon_l}{2}$

3) 内部节点隐私预算 $\epsilon_{nl} := \frac{\epsilon_l}{2\text{max_depth}}$

4) 利用 GDF 算法生成新数据集 $D \leftarrow \text{GDF}(D_0)$;

5) 初始化: $d=1$, 生成根节点 root;

6) node:=root;

7) while $d < \text{max_depth}$ do

8) for 深度 d 下每个节点 node do

9) for $j = \text{to } m$ do /* a_j 作为划分属性*/

10) 增益 $G_{a_j} \leftarrow G(D_L, D_R)$ /*式(2)*/

11) $P_{a_j} \leftarrow \exp\left(\frac{\epsilon_{nl} G_{a_j}}{2S(G_{a_j})}\right)$ /*指数机制*/

12) end for

13) 选取最优分属性 $\alpha_* := \arg\max_{c \in [1, m]} P_c / \sum_j P_j$;

14) for $\forall \alpha_*^v \in \alpha_*$ do

15) 为 node 生成一个分支;

16) 令 D_0 为 D 在 α_* 上取值为 α_*^v 的样

本子集;

17) if $D_0 = \emptyset$ then

18) 该分支节点标记为叶节点 Z_l ,

其类别标记为 D 中样本最多的类;

19) else

20) 以 Tree Gen ($D_0, A \setminus \{\alpha_*\}$)

为深度 $d := d+1$ 的分支节点 node;

21) end if

22) end for

23) end for

24) end while

25) 将深度 max_depth 下的节点标记为叶节点 Z_l ;

26) 生成一棵以 root 为根节点的决策树 T_0 ;

27) for T_0 的每个叶节点 $Z_l := \{V_l\}$ do
 28) $V_l \leftarrow V(D)$ /*式(3)*/
 29) $V_l' \leftarrow V(V_l)$ /*几何叶剪枝算法*/
 30) $V_l'' \leftarrow V_l' + \text{Lap}(0, S(V) / \varepsilon_{nl})$ /*拉普拉斯机制*/
 31) end for

首先, 用 GDF 算法对初始数据集进行预处理, 然后, 从根节点开始构建决策树, 直到达到最大深度, 该过程是一个递归过程。对于深度 d 下每个节点, 进行属性测试, 选择分支节点的划分属性时, 添加差分隐私的指数机制。其中, 效用函数为信息增益 G , 具有较高增益的特征值更有可能被选作划分值。叶节点对应了决策结构, 在生成一棵决策树时, 使用 GLC 算法对叶节点剪枝, 然后用拉普拉斯机制将随机噪声添加到叶节点值上。对于不同节点的隐私预算分配, 为叶节点分配一半的隐私预算 ε_l , 然后将剩余预算平均分配给内部节点的每个深度, 每个节点可获得 ε_{nl} 的隐私预算。

定理 3^[18] 算法 1 的输出满足 ε_l 差分隐私。

第二步是利用算法 2 将生成的单棵决策树顺序合成生成一个差分隐私的梯度提升决策树。给定决策树总棵数 T 和一个合成内部决策树棵数 T_e , 可以得到合成的总数为 $N_e \leftarrow \lceil T / T_e \rceil$, 则每棵树的隐私预算为 $\varepsilon_e \leftarrow \varepsilon / N_e$ 。当建立第 t 棵差分隐私决策树时, 先计算该决策树在合成中的位置 $t_e \leftarrow t \bmod T_e$ 。

算法 2 训练差分隐私的 GBDT

输入 M : 训练数据集; ε_l : 隐私预算、max_depth: 最大深度; λ : 正则化参数; L : 损失函数; T : 决策树的总棵数; T_e : 一个合成内决策树的棵数

输出 ε 差分隐私的 GBDT

1) 合成的总棵数 $N_e: \lceil T / T_e \rceil$
 2) 每棵决策树的隐私预算 $\varepsilon_e := \frac{\varepsilon}{N_e}$
 3) for $t=1$ to T do
 4) 更新所有训练实例在损失函数 L 上的梯度
 5) $t_e := t \bmod T_e$
 6) if $t_e := 1$ then
 7) $D \leftarrow M$
 8) end if

9) $D_t \leftarrow \text{Sample}\left(D, \frac{|D| \eta (1-\eta)^{\varepsilon_e}}{1 - (1-\eta)^{\varepsilon_e}}\right)$ /*数据集采样*/
 10) $D \leftarrow D - D_t$
 11) 调用算法 1 生成决策树 Tree_t , 参数设置如下。

数据集 D_t ;

最大深度 max_depth;

隐私预算 ε_e ;

$$S(G) := \frac{3\lambda + 2}{(\lambda + 1)(\lambda + 2)} g_{\max}^2$$

$$S(V) := \min\left(\frac{g_{\max}}{1 + \lambda}, 2g_{\max}(1 - \eta)^{t-1}\right)$$

12) end for

由于每棵树的最大叶节点值在 GLC 中是不同的, 因此要利用先前生成的决策树的贡献, 分配给一棵树的实例数和其叶节点敏感度成正比。

具体来说, 利用 $\text{Sample}(D, \frac{|D| \eta (1-\eta)^{\varepsilon_e}}{1 - (1-\eta)^{\varepsilon_e}})$ 函数从数据集 D 中随机选择 $\frac{|D| \eta (1-\eta)^{\varepsilon_e}}{1 - (1-\eta)^{\varepsilon_e}}$ 个未使用的实例

生成训练数据集 D_t , D 在合成最开始被初始化为 M 。每棵决策树是由算法 1 单棵决策树生成函数生成的。所有 T 棵树被依次训练, 最终所有决策树组成了最终学习模型。

定理 4^[18] 算法 2 的输出满足 ε 差分隐私。

4 实验评估

4.1 实验设计

本文采用 LightGBM 算法库^[19]实现了差分隐私的 GBDT 算法, 对比库中不加噪声的 GBDT 算法, 对系统的有效性和可用性进行评估。实验在一台 Intel(R) Xeon(R) E5-2678 v3@2.50 GHz 为核心 CPU 的计算机上进行。

实验使用的数据集是 CICIDS2017^[20], 该数据集从真实环境中捕获, 包含良性流量和新型常见攻击的流量, 流量包经过 CICFlowMeter 特征提取工具处理后, 可以提取出时间戳、源 IP、目的 IP、使用协议等 78 个特征, 并被标记为“良性流量”和“恶性流量”, 组成 CSV 特征数据集。本文使用了该数据中与 DDoS 攻击和 port scan 相

表1 实验使用数据集的详细信息
Table 1 Detailed information of the experimental dataset

序号	数据集名称	恶意流量类型	良性流量总数	恶意流量总数
1	Friday-WorkingHours-Afternoon-DDoS.pcap	DDoS	97 717	128 027
2	Friday-WorkingHours-Afternoon-PortScan.pcap	Port scan	127 536	158 930

关的两个数据集进行实验，数据集名称和数据集详细信息如表1所示。

实验展示了本算法在二分类任务下的测试准确率和训练时间。实验中决策树最大深度根据训练情况设置为6或7，学习率设置为0.01，度量函数使用曲线下面积（AUC，area under the curve），正则化参数设置为0.01，boosting迭代次数和一个合成内部决策树总棵数均设置为 n_tree 。训练数据集和测试数据集的比例分别为80%和20%。

4.2 测试准确率

首先设置参考组，利用不加扰动的GBDT算法对流量数据进行训练拟合和预测，识别准确率为99.98%和99.97%，说明整体上GBDT算法对加密流量检测具有较好的识别效果。

然后利用本文设计的基于差分隐私的GBDT算法对数据集进行隐私保护的恶意流量检测，隐私预算 ϵ 分别设置为0.5、1、2、4、6，boosting迭代次数和一个合成内部决策树总棵数（ n_tree ）分别设置为20、50、100，实验结果如图3、图4所示。

根据实验结果可知，一个合成内部决策树总棵数过小，即 n_tree 过小时，模型拟合过程未收敛；而 n_tree 过大时，模型过拟合。因此，选取 n_tree 为50时，模型拟合情况较好，且预测准确率较高。

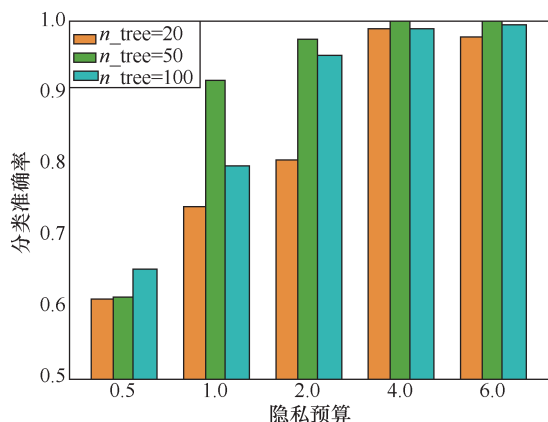


图3 数据集1：在不同隐私预算和不同 n_tree 下，DDoS攻击流量的识别准确率

Figure 3 Dataset 1: The recognition accuracy of DDoS attack traffic with different privacy budgets and n_tree

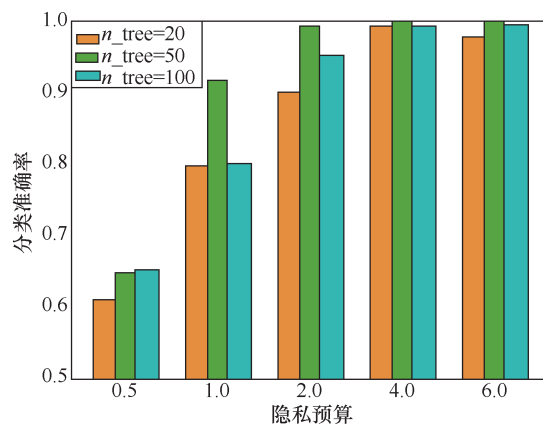


图4 数据集2：在不同隐私预算和不同 n_tree 下，PortScan流量的识别准确率

Figure 4 The accuracy of PortScan traffic recognition with different privacy budgets and n_tree

隐私预算 ϵ 越大，加密流量检测的准确率越高，但此时对训练数据的隐私保护效果越差。在图3（DDoS数据集）中，当 $\epsilon=0.5$ 时，对训练数据的隐私保护效果最好，此时在最好训练模型下，分类准确率为65.3%。结合图3、图4可得，当 ϵ 取值为1时，本文设计的系统能达到可用性与安全性的平衡，此时在各数据集模型训练最好的情况下，即 $n_tree=50$ 时，DDoS攻击、port scan分类准确率可以达到91.7%、92.4%，同时， $\epsilon=1$ 意味着隐私保护效果较好。

4.3 时间效率

本节分别用GBDT算法和差分隐私的GBDT算法（DP-GBDP）训练并预测模型，对两种算法的训练时间、预测时间进行对比。

训练模型过程中，DP-GBDP算法的计算开销主要来源于在计算每个信息增益的概率时，额外添加的指数机制，以及最后输出结果时添加的拉普拉斯扰动。指数机制的开销随着训练数据的增多而增大。

训练时间的效率分析实验中，每种分类算法参数设置与4.1节一致，隐私预算 ϵ 取1，分别运行100轮模型训练算法，对时间取平均值求得一轮模型训练所需时间， n_tree 分别取20、50、100

时, 训练时间如表 2 所示。由表 2 可知, 训练时间随着训练数据总量增大而增加, 一个合成内部决策树总棵数越多, 训练时间越大。总体上, DP-GBDT 算法训练时间比 GBDT 算法长 2~3 倍, 训练速度非常快。在训练效果最好情况下, 即 ε 取 1, n_{tree} 取 50, 在 DDoS 攻击和 port scan 攻击下两个数据集训练时间分别为 5.11 s、5.59 s, 训练效率均较高。

表 2 DP-GBDT 和 GBDT 算法的训练时间
Table 2 The Training time of DP-GBDT and GBDT algorithm

单位: s

n_{tree}	DDoS		PortScan	
	DP-GBDT	GBDT	DP-GBDT	GBDT
20	3.19	1.97	3.70	2.34
50	5.11	2.31	5.59	2.77
100	9.77	2.94	9.69	3.59

预测时间的效率分析实验中, 利用 4.2 节训练完成的模型对 DDoS 攻击和 port scan 数据集中 20% 的流量进行测试 (即 45 149 条和 57 294 条流量), 经 1 000 轮测试, 对每轮预测的时间取平均值, 测试时间结果如表 3 所示。根据实验结果可知, DP-GBDT 算法在预测效率上与 GBDT 算法近似, 这表明在 DP-GBDT 算法在预测时, 相比 GBDT 算法没有额外计算开销。

表 3 DP-GBDT 和 GBDT 算法每轮预测时间
Table 3 The prediction time of each round of DP-GBDT and GBDT algorithm

单位: s

DDoS		PortScan	
DP-GBDT	GBDT	DP-GBDT	GBDT
0.0409	0.0414	0.0522	0.0525

5 相关工作

机器学习技术不断发展优化, 被广泛应用于语音识别^[21]、图像处理^[22]、网络流量监测^[23]等领域。但随着隐私保护相关法律标准的不断规范化, 个人、机构的隐私保护意识逐渐增强, 因此, 机器学习方案的输入数据和模型参数有了一定隐私性要求, 隐私保护的机器学习技术成为研究热点, 目前有两个主流研究方向: 基

于同态加密 (homomorphic encryption)、安全多方计算 (MPC, multi-party computation)、加密布尔电路 (garbled circuits)、秘密共享 (secret sharing) 等技术的密码学方法和基于噪声扰动的差分隐私方法。

5.1 基于密码学的隐私保护机器学习

基于密码学的隐私保护机器学习研究, 通过对敏感数据加密, 保证数据传输和存储时的机密性, 利用安全协议对密文进行计算、分析, 防止恶意攻击者利用机器学习模型对训练数据集进行推测, 最后对预测结果解密, 得到明文上的正确输出。

在基于同态加密技术的隐私保护机器学习的研究上, Chen 等^[24]提出两方下的隐私保护分布式机器学习算法, 他们将数据集进行特定划分, 其中任意一方只保留特征向量的子集, 并确保任何一方的数据都不会遭到泄露。Erkin 等^[25]引入了数据打包技术, 使多个明文可以被加密为同一个密文, 提高了加法同态加密的计算效率。Bost 等^[26]结合同态加密和加密布尔电路技术, 构建了 3 种常见的分类算法 (超平面决策、朴素贝叶斯和决策树)。Dowlin 等^[27]提出了 CryptoNets, 利用加法、乘法同态加密, 实现了一种高吞吐量、高准确率、可应用于加密数据的神经网络, 并对该系统的正确性进行理论推导。Hesamifard 等^[28]提出的 CryptoDL 是一个隐私保护的卷积神经网络模型, 其中激活函数 (如 ReLU 函数、Sigmoid 函数) 利用低次多项式逼近的方法实现, 结合同态加密技术达到数据保护的目的。

在基于安全多方计算的隐私保护机器学习研究上, Mohassel 等^[29]提出了 SecureML, 针对线性回归、逻辑回归和神经网络训练等机器学习技术, 完成了两方下的安全计算任务, 同时提出了 MPC 友好的算法来替代神经网络中的非线性函数, 如 Sigmoid、Softmax 等。Konečný 等^[30]提出联邦学习的概念, 利用多个相互独立的服务器上的数据集, 共同训练出一个高质量的模型, 设计了结构化更新、草图更新两种方法, 降低系统的通信开销。Mohassel 等^[31]提出的 ABY³ 是一个三方下的隐私保护机器学习计算框架, 设计了新的向量内积、矩阵乘法、浮点数截断协议, 能在三

方服务器之间秘密共享数据, 联合训练和评估神经网络模型。Patra 等^[32]扩展了 ABY³ 框架, 改进其中的乘法协议, 提高了安全多方计算下机器学习训练和推断过程的性能表现。

5.2 基于差分隐私的隐私保护机器学习

差分隐私技术通过在模型训练过程中加入随机噪声, 使预测结果与真实值存在一定偏差, 可以防止成员推断攻击等。根据随机噪声在机器学习中添加位置的不同, 差分隐私的扰动方案可以分为输入扰动、算法扰动、输出扰动和目标扰动。

输入扰动通过直接在数据集上添加噪声, 即使公开的公开的计算过程, 也能实现输出结果的隐私保护。Dwork 等^[33]设计了差分隐私的主成分分析算法, 通过在本征分解之前, 对协方差矩阵添加高斯噪声, 使输出结果矩阵符合差分隐私。Heikkilä 等^[34]利用高斯噪声机制设计了隐私保护的差分隐私贝叶斯决策。

算法扰动是指在机器学习算法迭代过程中添加噪声, Hardt 等和 Abadi 等分别在每轮矩阵向量乘法、每轮随机梯度下降算法中引入高斯噪声机制, 实现了 DP-PAC 和 DP 深度学习算法^[35-36]。

输出扰动则是在非隐私保护算法的输出结果上添加噪声, Chaudhuri 和 Monteleoni^[37]基于该机制设计了差分隐私的逻辑回归算法, Chaudhuri 等^[38]利用指数机制, 实现了差分隐私的主成分分析算法。

目标扰动算法是在模型的目标函数上引入扰动, 经验风险最小化 (ERM, empirical risk minimization)^[39]算法则是基于此, 在目标函数表达式中加入随机噪声, 保证扰动后的训练过程满足差分隐私。Zhang 等提出利用泰勒展开多项式逼近目标函数, 在各项系数中添加拉普拉斯噪声, 从而使算法满足差分隐私, 该方法被成功应用于线性回归和逻辑回归模型^[40]。

在实际应用场景中, 数据的加密过程和密文的传输、运算过程往往需要消耗大量计算资源, 产生巨大的性能开销。而差分隐私仅通过添加噪声机制即可实现隐私保护, 不存在额外的计算开销, 但噪声会对模型的预测准确率造成一定影响。因此, 如何在保证数据集隐私性的同时, 平衡模

型准确率和可用性是当前研究重点之一。

6 结束语

随着网络安全威胁和防护问题的不断涌现, 隐私保护标准逐渐规范化, 数据隐私保护算法显得尤为重要, 尤其是在网络入侵检测领域, 网络空间中存在大量流量包、日志文件等数据, 为入侵检测算法带来大量信息的同时, 存在巨大风险与隐患, 可能导致个人隐私信息的泄露。本文从加密流量检测的原理出发, 研究了实现加密流量检测的几种机器学习算法; 同时调研了隐私保护的机器学习领域涉及的差分隐私机制、同态加密机制等。本文设计并实现了一个隐私保护的加密流量检测系统, 对其安全性进行了理论分析, 实验测试结果表明, 该方案实现对数据集隐私保护的同时, 达到较高的恶意流量识别准确率。

参考文献:

- [1] MC-GREW D, ANDERSON B. Enhanced telemetry for encrypted threat analytics[C]//2016 IEEE 24th International Conference on Network Protocols (ICNP). 2016: 1-6.
- [2] VOIGT P, VON DEM BUSSCHE A. Practical implementation of the requirements under the GDPR[M]//The EU General Data Protection Regulation (GDPR). Cham: Springer International Publishing, 2017: 245-249.
- [3] BUKATY P. The CCPA[M]//The California Consumer Privacy Act (CCPA). IT Governance Publishing, 2019: 123-169.
- [4] 王春晖. 《网络安全法》六大法律制度解析[J]. 南京邮电大学学报: 自然科学版, 2017, 37(1): 1-13.
WANG C H. Analysis of the six legal systems of Network Security Law[J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science Edition, 2017, 37(1): 1-13.
- [5] 洪延青, 葛鑫. 国家标准《信息安全技术个人信息安全规范》修订解读[J]. 保密科学技术, 2019 (8): 6.
HONG Y Q, GE X. Information Security Technology Personal Information Security Specification revision interpretation[J]. Security Science and Technology, 2019 (8): 6.
- [6] 诸葛建伟, 韩心慧, 周勇林, 等. 僵尸网络研究[J]. 软件学报, 2008, 19(3): 702-715.
ZHUGE J W, HAN X H, ZHOU Y L, et al. Research and development of Botnets[J]. Journal of Software, 2008, 19(3): 702-715.
- [7] CAO Z, XIONG G, ZHAO Y, et al. A survey on encrypted traffic-classification[M]//Applications and Techniques in Information Se-

- curity. Springer Berlin Heidelberg. 2014: 73-81.
- [8] KLEINBAUM D G, DIETZ K, GAIL M, et al. Logistic regression[M]. Springer, 2002.
- [9] HO T K. Random decision forests[C]//Proceedings of 3rd International Conference on Document Analysis and Recognition: Volume 1. 1995: 278-282.
- [10] FANG Y, QIU Y, LIU L, et al. Detecting webshell based on random forest with fasttext[C]//Proceedings of the 2018 International Conference on Computing and Artificial Intelligence. 2018: 52-56.
- [11] SI S, ZHANG H, KEERTHI S S, et al. Gradient boosted decision trees for high dimensional sparse output[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70 (ICML'17). 2017: 3182-3190.
- [12] 翟征德, 宗兆伟. 针对 DNS 服务器的抗 DDoS 安全网关系统: 中国, CN101572701B[P]. 2013-11-20.
- ZHAI Z D, ZONG Z W. Anti-DDoS security gateway system for DNS server[P]. 2013-11-20.
- [13] SHOKRI R, STRONATI M, SONG C, et al. Membership inference attacks against machine learning models[C]//2017 IEEE Symposium on Security and Privacy (SP). 2017: 3-18.
- [14] SWEENEY L. K-anonymity: a model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(5): 557-570.
- [15] MACHANAVAJJHALA A, KIFER D, GEHRKE J, et al. l-diversity: privacy beyond k-anonymity[J]. ACM Transactions on Knowledge Discovery from Data (TKDD), 2007, 1(1): 3.
- [16] DWORK C. Differential privacy: a survey of results[C]//International Conference on Theory and Applications of Models of Computation. 2008: 1-19.
- [17] FRANK M, MIRONOV I. Differentially private recommender systems: building privacy into the net[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2009: 627-636.
- [18] LI Q, WU Z, WEN Z, et al. Privacy-preserving gradient boosting decision trees[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020: 784-791.
- [19] KE G L, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Advances in Neural Information Processing Systems 30 (NIPS 2017). 2017: 3149-3157.
- [20] SHARAFALDIN I, LASHKARI A H, GHORBANI A A. Toward Generating a new intrusion detection dataset and intrusion traffic characterization[C]//4th International Conference on Information Systems Security and Privacy (ICISSP). 2018.
- [21] LIU Z, WU Z, LI T, et al. GMM and CNN hybrid method for short utterance speaker recognition[J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3244-3252.
- [22] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Advances in Communications of the ACM, 2012: 1097-1105.
- [23] MEIDAN Y, BOHADANA M, SHABTAI A, et al. ProfilloT: a machine learning approach for IoT device identification based on network traffic analysis[C]//Proceedings of the Symposium on Applied Computing. 2017: 506-509.
- [24] CHEN T, ZHONG S. Privacy-preserving backpropagation neural network learning[J]. IEEE Transactions on Neural Networks, 2009, 20(10): 1554-1564.
- [25] ERKIN Z, VEUGEN T, TOFT T, LAGENDIJK R L. Generating private recommendations efficiently using homomorphic encryption and data packing[J]. IEEE Trans Inf Forensics Security, 2012, 7(3): 1053-1066.
- [26] BOST R, POPA R, TU S, GOLDWASSER S. Machine learning classification over encrypted data[C]//NDSS. 2015: 4325.
- [27] GILAD-BACHRACH R, DOWLIN N, LAINE K, et al. Cryptonets: applying neural networks to encrypted data with high throughput and accuracy[C]//International Conference on Machine Learning. 2016: 201-210.
- [28] HESAMIFARD E, TAKABI H, GHASEMI M. CryptoDL: deep neural networks over encrypted data[J]. ArXiv preprint ArXiv:1711.05189, 2017.
- [29] MOHASSEL P, ZHANG Y. SecureML: a system for scalable privacy-preserving machine learning[C]//2017 IEEE Symposium on Security and Privacy (SP). 2017: 19-38.
- [30] KONEČNÝ J, MCMAHAN H B, YU F X, et al. Federated learning: Strategies for improving communication efficiency[J]. arXiv preprint arXiv:1610.05492, 2016.
- [31] MOHASSEL P, RINDAL P. ABY3: a mixed protocol framework for machine learning[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. 2018: 35-52.
- [32] PATRA A, SURESH A. BLAZE: Blazing Fast Privacy-Preserving Machine Learning[J]. arXiv preprint arXiv: 2005.09042, 2020.
- [33] DWORK C, TALWAR K, THAKURTA A, et al. Analyze gauss: optimal bounds for privacy-preserving principal component analysis[C]//Proceedings of the Forty-sixth Annual ACM Symposium on Theory of Computing. 2014: 11-20.
- [34] HEIKKILÄ M, LAGERSPETZ E, KASKI S, et al. Differentially private Bayesian learning on distributed data[C]//Advances in Neural Information Processing Systems. 2017: 3226-3235.
- [35] HARDT M, PRICE E. The noisy power method: a meta algorithm with applications[C]//Advances in Neural Information Processing Systems. 2014: 2861-2869.
- [36] ABADI M, CHU A, GOODFELLOW I, et al. Deep learning with differential privacy[C]//Proceedings of the 2016 ACM SIGSAC

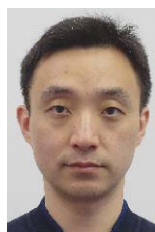
Conference on Computer and Communications Security. 2016: 308-318.

- [37] CHAUDHURI K, MONTELEONI C. Privacy-preserving logistic regression[C]//Advances in Neural Information Processing Systems. 2009: 289-296.
- [38] CHAUDHURI K, SARWATE D A, SINHA K. A near-optimal algorithm for differentially-private principal components[J]. Journal of Machine Learning Research, 2013, 14(1): 2905-2943.
- [39] CHAUDHURI K, MONTELEONI C, SARWATE A D. Differentially private empirical risk minimization[J]. Journal of Machine Learning Research, 2011, 12: 1069-1109.
- [40] ZHANG J, ZHANG Z, XIAO X, et al. Functional mechanism: regression analysis under differential privacy[J]. arXiv preprint arXiv:1208.0219, 2012.

[作者简介]



张心语（1997-），女，浙江诸暨人，浙江大学博士生，主要研究方向为人工智能安全、数据安全。



张秉晟（1984-），男，浙江杭州人，浙江大学研究员、博士生导师，主要研究方向为密码学、安全多方计算、零知识证明、区块链安全、数据安全。



孟泉润（1994-），男，河南新乡人，浙江大学硕士生，主要研究方向为数据安全。



任奎（1978-），男，安徽巢湖人，浙江大学教授、博士生导师，主要研究方向为云计算中的数据安全、计算服务外包安全、无线系统安全、隐私保护、物联网系统与安全。