

doi:10.19665/j.issn1001-2400.2021.03.022

加密流量中的恶意流量识别技术

曾 勇¹, 吴正远¹, 董丽华², 刘志宏¹, 马建峰¹, 李 赞²

(1. 西安电子科技大学 网络与信息安全学院, 陕西 西安 710071;

2. 西安电子科技大学 综合业务网理论及关键技术国家重点实验室, 陕西 西安 710071)

摘要: 网络流量的加密传输是互联网的发展趋势之一, 而加密流量中的恶意流量识别是维护网络空间安全的重要手段。识别恶意流量需要将加密流量进行密/非密、应用程序以及加密算法的细粒度区分以提高识别效率, 再将不同精细度区分后的流量经过预处理后转化为图像、矩阵和 N-gram 等形式导入机器学习训练模型中进行训练, 实现良性/恶意流量的二分类以及多分类。基于机器学习的识别效果严重依赖于样本数量和质量, 同时无法有效地应对整形和混淆后的流量, 而基于密码学的恶意流量识别技术通过深度融合可搜索加密技术、流量审查机制和可证明安全模型, 在加密流量上检索恶意关键词以避免样本数目不足和流量整形的问题, 同时实现对数据和规则的隐私保护。对加密流量中的恶意流量识别所涉及到的上述技术进行了总结, 指出存在的问题并展望未来发展的方向。

关键词: 加密流量; 恶意流量; 机器学习; 密码学

中图分类号: TP393 **文献标识码:** A **文章编号:** 1001-2400(2021)03-0170-17

Research on malicious traffic identification technology in encrypted traffic

ZENG Yong¹, WU Zhengyuan¹, DONG Lihua², LIU Zhihong¹, MA Jianfeng¹, LI Zan²

(1. School of Cyber Engineering, Xidian University, Xi'an 710071, China;

2. State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China)

Abstract: The encrypted transmission of network traffic is one of the development trends of the Internet. The identification of malicious traffic in encrypted traffic is an important way to maintain the security of cyberspace. One of the prior tasks of identifying malicious traffic is to classify encrypted traffic into the encrypted/unencrypted, different kinds of the application programs and encryption algorithms in order to improve the efficiency of identification. Then they are transformed into the image, matrix, n-gram or other forms which will be sent into the machine learning training model, so as to realize the binary classification and multi classification of benign malicious traffic. However, the machine learning based way relies seriously on the number and quality of samples, and can not effectively deal with the data after traffic shaping or confusion. Fortunately, cryptography based malicious traffic identification can search malicious keywords over encrypted traffic to avoid such problems, which must integrate searchable encryption technology, deep packet inspection and a provable security model to protect both data and rules. Finally, some unsolved

收稿日期: 2020-12-18

网络出版时间: 2021-04-12

基金项目: 国家自然科学基金(619411105)

作者简介: 曾 勇(1978—), 男, 副教授, 博士, E-mail: yzeng@mail.xidian.edu.cn

吴正远(1997—), 男, 西安电子科技大学硕士研究生, E-mail: 18066746790@163.com

董丽华(1977—), 女, 副教授, 博士, E-mail: lih_dong@mail.xidian.edu.cn

刘志宏(1968—), 男, 副教授, 博士, E-mail: liuzhihong@mail.xidian.edu.cn

马建峰(1963—), 男, 教授, 博士, E-mail: jfma@mail.xidian.edu.cn

李 赞(1975—), 女, 教授, 博士, E-mail: zanli@xidian.edu.cn

网络出版地址: <https://kns.cnki.net/kcms/detail/61.1076.TN.20210412.1528.002.html>

problems of malicious traffic identification technology in encrypted traffic are presented.

Key Words: encrypted traffic; malicious traffic; machine learning; cryptography

在互联网技术日益健全的今天,网络流量识别技术对网络管理、服务质量保障和网络安全等具有重大的意义。伴随着加密技术的不断发展,加密流量在互联网流量中的数量和比例也不断上升。根据最近的互联网研究趋势报告^[1],如今87%的web流量是加密的,文献[2]预计在2020年超过70%的恶意软件活动将使用某种类型的加密来隐藏恶意软件的传输、控制命令活动和数据渗漏。由于加密后流量的特征发生了改变,因而传统流量检测方式在加密环境下难以复现。如何在加密流量上进行有效的恶意流量识别,已成为了网络安全领域的重要挑战。

已有文献[3-7]综述了当前加密流量识别技术的研究进展。其中,文献[3]总结了网络加密流量的基本概念、研究进展和评价指标等概念。文献[4]总结了用于加密流量识别的方法以及影响加密流量识别的因素,文献[5]总结了流量识别的四种常用方法:基于端口号、深度报文检测、机器学习和深度学习,着重介绍了常用于深度学习训练的几种特征和深度学习模型。但是这些综述并未涉及加密流量中的恶意流量识别。文献[6]对加密流量数据集和加密流量中的恶意流量检测步骤进行了小结,并介绍了基于机器学习的识别方法。文献[7]介绍了通过将流量转化为图像后采用计算机视觉的方法进行恶意流量识别的模型,这类方法目前仅在三大类恶意流量上有比较好的识别结果,且仅部分识别方法面向加密流量。

上述恶意流量检测工作均为基于机器学习的方法,由于基于机器学习方法的识别效果严重依赖于样本数量和质量,而当前在加密流量领域尚且缺乏类似ImageNet^[8]的经典数据集,不仅如此,基于机器学习的方法无法有效应对混淆和整形后的流量^[9-10]。近年来研究人员发现,基于密码学的方法可以通过密文检索技术与深度报文检测技术的融合来避免此类问题。为了更好地理解这些方法,笔者对已有的相关技术进行了如下总结:

(1) 加密算法区分。当前流量中采用的加密算法种类繁多,不同的加密方式会产生不同的特征,需要不同的识别方法。当前尚无通用方法能够应用于所有类型的加密流量,而恶意流量识别的任务之一需要先将加密流量按照不同精细度区分开,例如区分AES、DES、3DES、Grain等密码体制,为后续恶意流量区分提供支撑^[11-55]。

(2) 恶意特征识别。基于机器学习的恶意流量识别效率和准确率随着机器学习技术的发展而提高。其最新进展是通过改进加密流量中恶意特征提取方式,构建不同的带标签恶意特征集,并将其输入各种机器学习模型进行训练;通过模型设计与参数调优等方法来保证加密流量中对恶意流量识别的准确度^[56-81]。

(3) 恶意密文检索。基于密码学的恶意流量识别技术,结合流量审查机制、可搜索加密技术以及可证明安全模型,通过在加密流量上检索恶意关键词,从而达到在加密流量中识别恶意流量的效果,并对用户数据与检测方的检测规则同时提供保护^[82-113]。

1 加密算法区分

加密算法区分是识别加密流量中恶意流量的第一步。将加密流量按照不同加密算法区分开,可以有效精简数据集大小,提高识别效率,并为后续加密流量中的恶意流量识别做先验准备。不同密码算法所产生的密文在统计特性上存在一定的差异。这些差异是识别密码算法的重要依据。基于统计学和机器学习的方法可以较有效地区分加密算法。

1.1 加密非加密流量区分

将加密与非加密流量区分开是进行恶意流量识别的基础。笔者总结了近三年在公开的ISCXVPN/Non-VPN^[11]流量数据集上的相关工作,详细工作见表1。该数据集包含七种VPN流量和七种非VPN流量,每种类型内又包含多个应用的流量。其中浏览器类型的流量可能包含了其他类型的部分流量,因此实际识别中常见的区分方法有二分法(区分VPN与非VPN流量)和六分法(区分流量来源于何种类型)。LOTFOLLAHI等^[12]提出了深度数据包框架Deep packet。主要提取加密数据包的有效负载特征,并采用了栈式自动编码器(SAE)和卷积神经网络(CNN)两种深度神经网络结构,从而对VPN和非VPN网络流量进行二分类。并可对VPN或者非VPN中的流量进行大类区分。例如可区分出VPN或者非VPN的聊天、

<http://journal.xidian.edu.cn/xdxb>

邮件、视频等类别。BAGUI 等^[13]使用与时间相关的特征进行了类似的研究。WANG 等^[14]提出了一种基于 CNN 的端到端深度学习模型。通过将流量转化为图像的方式自动选取特征,从而实现 VPN 与非 VPN 的二分类与大类区分。GUO 等^[15]用收缩自动编码器(CAE)和卷积神经网络(CNN)两种方法进行比较,对 WANG 等的工作进行了改进。唐舒烨等^[16]引入基于分段熵分布的随机性检测方法,实现了对加密 VPN 流量与非加密 VPN 流量的细粒度分离。ZHOU 等^[17]采用了熵估计和人工神经网络的方法进行二分类。CASINO 等^[18]提出了基于数据流随机性评估的流量识别方法 HEDGE,可识别压缩流量和加密流量。该方法可以应用于单个数据包,而无需访问整个数据流。ACETO 等^[19]使用深度学习技术,基于自动提取的特征建立识别器,能够对手机的加密流量进行分类操作。上述方法均基于对有效载荷的统计。NIU 等^[20]提出了一种结合统计和机器学习的启发式统计测试方法。该方法优于仅统计或机器学习的方法。

表 1 密/非密流量区分及应用程序分类方法小结

文献	识别特征	识别模型	数据集	识别目的	详细描述	评价指标
[12]	头部特征/ 负载特征	CNN/ SAE	ISCXVPN/ Non-VPN	二分类:vpn/非vpn 六分类:流量类型 分类	对包的应用层有效载荷中可用信息的分析	召回率 98% 召回率 94%
[13]	时序特征	CNN	ISCXVPN/ Non-VPN	六分类:流量类型 分类	使用 CNN 对时间相关特征的数据集进行分类	最高准确率 94.6%
[14]	头部特征	CNN	ISCXVPN/ Non-VPN	二分类:vpn/非vpn 六分类:流量类型 分类	流量头部特征可视化图像	准确率 99.9% 准确率 94.9%
[15]	头部特征	CAE/ CNN	ISCXVPN/ Non-VPN	二分类:vpn/非vpn 六分类:流量类型 分类	流量头部特征可视化图像	准确率 98.77% 准确率 92.92%
[16]	负载特征	胶囊神经网络	ISCXVPN/ Non-VPN	二分类:vpn/非vpn	针对 VPN 加密报文序列进行高低熵划分	最高准确率 99.87%
[17]	负载特征/ 统计特征	人工神经网络	ISCXVPN/ Non-VPN	二分类:vpn/非vpn	使用熵估计和人工神经网络相结合	最高准确率 92.9%
[18]	负载特征	神经网络	IMG/ COCO ^[115]	二分类:加密/压缩 流量分类	可以应用于单个数据包,而无需访问整个流	最高准确率 94.72%
[19]	头部特征/ 负载特征	深度学习 技术	FB/私有 数据集	多分类: 加密手机流量分类	基于自动提取的特征建立分类器	优于已有方法
[20]	统计特征	C4.5	私有数据集	多分类: 加密网络流量分类	同时使用统计特征和机器学习的方式	优于单一方法
[21]	统计特征	机器学习	私有数据集	四分类: 应用程序分类	消除了非高斯分布的特征,实现高精度	最高准确率 97.4%
[24]	头部特征/ 负载特征	有监督机器学习	私有数据集	五分类: 应用程序分类	离流流量通过指纹识别应用程序	最高准确率 99.64%
[26]	统计特征	联合建模	私有数据集	多分类: 应用程序分类	联合建模用户行为模式等特征进行分类	最高准确率 97%
[27]	头部特征/ 负载特征	属性图 分类	Campus/ appScanner ^[79]	多分类: 应用程序分类	基于二阶马尔可夫链的属性感知加密流量分类	最高准确率 93.49%
[28]	头部特征/ 统计特征	随机森林 RF	私有数据集	识别个人敏感信息	自动从 IoT 网络流量中推断出个人敏感信息	最高准确率 99.8%

1.2 加密流量中的应用识别

将加密非加密流量区分后,另一个重要的工作是将加密流量所属的应用程序进行分类,详细内容见表 1。OKADA 等^[21]使用统计特征的最佳组合进行识别。由于消除了非高斯分布的特征,因此能够以较少的计算量实现高精度的识别。为了提高现有方法在鉴别准确度方面的性能,HE 等^[22-23]指出攻击者精心选择一些流量特征,并利用一些有效的机器学习算法对不同类型的应用程序进行建模。这些模型可用于对目标的洋葱路由(Tor)流量进行识别并推断其应用程序类型。类似地,ALMUBAYED 等^[24]指出,通过使用监

<http://journal.xidian.edu.cn/xdxb>

督学习方式,Tor 流量仍可以在网络中的其他 HTTPS 流量中识别。考虑到用户使用同一应用程序做出不同的行为时会产生不同的流量,CONTI 等^[25]指出,即便是在加密的条件下,攻击者依然可以通过特定的方法识别用户在网络中的行为,这些行为导致了隐私泄露的风险。FU 等^[26]开发了一个名为 CUMMA 的系统,通过联合建模用户行为模式、网络流量特征和时间依赖性,可对应用程序内的使用进行识别。SHEN 等^[27]指出应用程序长度二元组有助于应用程序识别。该二元组由证书包长度和 SSL/TLS 会话中的首个应用程序数据长度组成。SUBAHI 等^[28]开发了 IoT-app 作为隐私检查器的工具。该工具可以通过其应用程序的数据包,自动从 IoT 网络流量中推断出敏感个人身份信息(例如用户的位置等)。

1.3 加密算法识别

当流量中的明文文区分开后,密文所采用的加密算法也可以进行区分。早在上世纪 80 年代,密码学者已开始关注机器学习与密码学的联系,并提出一些相关的概念和结论^[29]。RIVEST 等发表《密码学与机器学习》^[30],探讨了机器学习应用于密码学的可行性。不同密码算法所产生的密文在统计特性上存在一定的差异,这些差异可以作为识别密码算法的重要依据^[31]。基于这些差异,机器学习在古典密码体制识别^[24]、分组密码体制识别^[32-33]、布尔函数设计^[34-36]等方向取得了初步的研究成果。机器学习方法亦可用于密码攻击,包括基于神经网络的明文恢复^[37-39]、基于机器学习的侧信道攻击^[40-42]等。表 2 总结了用于加密算法识别的详细方法。

表 2 加密算法识别小结

文献	识别算法	识别目的	详细描述	识别结果
[43]	SVM	5 种密码体制	在 ECB、CBC 模式下,使用 SVM 识别相同密钥和不同密钥加密生成的密文	ECB 模式优于 CBC 模式
[45]	SVM	5 种密码体制	DES、AES、TDES、RC5、Blowfish,共五种密码算法加密后的密文的直方图信息	平均识别准确率约 25%
[46]	Adaboost	5 类分组密码	对生成的密文进行学习,将识别错误的样本数据再次训练	平均识别准确率约 55%
[47-48]	8 种不同分类器	分组密码	使用 8 种不同分类器模型对分组密码进行识别	RF 效果的识别最佳
[49]	神经网络	AES 5 种候选算法	使用神经网络进行密码体制识别	神经网络配置得当时,可以正确分类
[50]	MLP	RC4 密钥流和随机密钥流	利用多层感知器学习特征并区分 RC4 密钥流和随机密钥流	平均识别准确率约 69%
[51]	K-means	5 种分组密码	分析 5 种分组密码构成的密码体制,使用 K-Means 算法识别加密后的密文	若参数合理,密文的识别率约 90%
[52-54]	RF	密码体制分层识别	提出密码体制分层识别方案,选择机器学习的 RF 算法进行识别	多分类任务下准确率约 60%~70%
[55]	RF	二分类任务识别	涉及的序列密码算法包括 Grain-128、RC4、Salsa	两两识别的平均识别准确率约 64%

DILEEP 等^[43]采用文档分类技术,在分组密码的 ECB、CBC 模式下,使用机器学习中的支持向量机(SVM)算法对 AES、DES、3DES、Blowfish、RC5 这 5 种密码体制进行识别。CHOU 等^[44]采取类似的识别方案表明在 CBC 模式下的分组密码的密文不能很好地区分。NAGIREDDY^[45]使用 SVM 分析同样 5 种密码算法加密后的密文的直方图信息。SONI 等^[46]提出使用 Adaboost 算法对 5 种分组密码算法进行分类学习。MISHRA 等^[33]提出包含 11 种密码算法的密码体制,采用 C 4.5 算法生成的决策树研究密码体制识别。在多分类任务下,从密文中提取 8 个特征进行识别。文献[47-48]分别使用 8 种不同分类器模型对分组密码进行识别。结果表明组合分类器模型随机森林(RF)效果的识别准确率最佳。SOUZA 等^[49]提出选取机器学习方法中的神经网络进行 RC6、Rijndael 等 5 种 AES 最终轮候选密码算法。BHATEJA 等^[50]提出了一种基于反向传播网络的方法来区分 RC4 密钥流和随机密钥流。WU 等^[51]使用机器学习方法中的 K 均值聚类(K-Means)算法识别加密后的密文,并分析 AES、Camellia、DES、3DES、SMS4 共 5 种分组密码构成的密码体制。文献[52-54]提出密码体制分层识别方案,选择机器学习的随机森林算法进行识别。随后,ZHAO 等^[55]研究 Grain-128 算法与其他 11 种密码算法的识别。主要进行了二分类任务识别(即密码算法两两识别),其中涉及的序列密码算法包括 Grain-128、RC4 与 Salsa。

<http://journal.xidian.edu.cn/xdxb>

2 基于机器学习的恶意流量识别

不同类型的流量具备不同的网络行为模式,这些信息直观地体现在其数据包上。例如,恶意流量与良性流量在进行握手协议时会产生不同的数据包头部信息,而不同类型的恶意流量在其平均包长、包间间隔等方面也存在差异;这些行为模式的差异是识别恶意流量的重要依据。因此可从采集到的加密流量中提取恶意流量的行为模式,将其进一步归纳为恶意流量的特征,并利用机器学习模型进行识别。

基于机器学习的恶意流量识别是将加密流量进行恶意特征提取,从而构建恶意特征集,并作为训练集输入训练模型,通过模型设计与参数调优等方法得到理想的准确度。该方法具有效率高、适用性广的优点。因此,基于机器学习的加密流量识别成为了当前研究热点。基于机器学习的恶意流量识别体系如图 1 所示。首先需要采集所需数据集,通常有使用公开流量数据集和私有流量数据集两种方法;然后对采集到的流量数据进行预处理工作,包括流量清洗、流量分割、特征集构建和流量转换;最后将数据集作为输入,利用机器学习模型学习恶意流量的恶意特征,通过迭代训练识别出恶意流量。

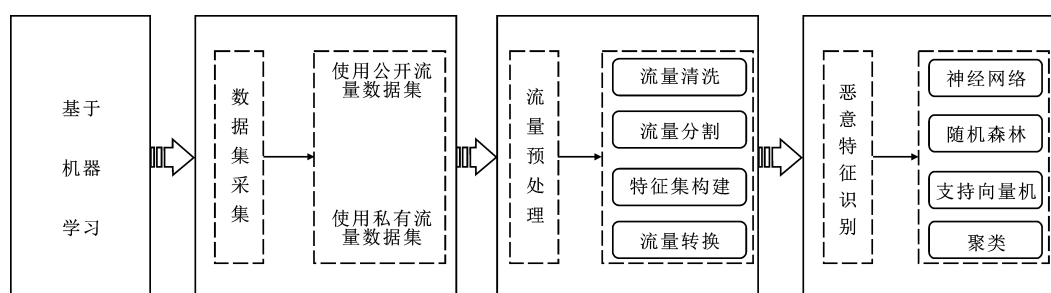


图 1 基于机器学习的恶意流量识别体系

2.1 数据集采集

使用机器学习模型进行恶意流量识别首先需要一个有代表性的数据集。尽管当前已有一些公开的加密流量数据集(例如 ISCX2012^[120], CTU-13^[121] 和 CICIDS2017^[123]),但目前加密流量领域仍然缺乏一个普遍被认可的加密流量数据集。其原因在于:通过不同方式加密后的流量需要不同的收集方法和场景,而一个数据集几乎不可能包含所有的流量类型。

因此,研究人员更倾向于首先使用私有流量数据集进行识别,利用脚本或者沙箱生成特定类型的加密流量,然后采集这些特定流量并打上标签。通过这种方法采集到的流量比直接从真实网络环境中进行采集更为精纯,同时易于贴上标签。但使用私有流量数据集的恶意流量识别方法往往难以复现,同时不方便与已有方法进行比较。

2.2 流量预处理

通常情况下,收集到的流量数据集并不能直接作为机器学习模型的输入,需要对其进行预处理工作。预处理通常包括流量清洗、流量分割、特征集构建和流量转换。流量清洗需要将收集到的流量中重复和无效部分清除。流量分割需要将过长的影响识别效率的流量分割为片段。将收集到的流量进行流量清洗和分割后,下一步是构建特征集,机器学习常用的流量特征有时空特征、头部特征、负载特征和统计特征 4 种。4 种特征在流量包上的表现如图 2 所示。

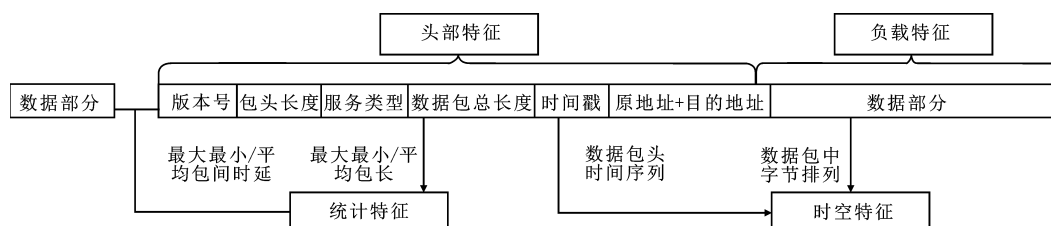


图 2 机器学习四种特征示意图

流量转换则将构建后的流量转化为图像^[62-67]、矩阵^[56-61,75-79]或者 N-gram^[68-74],以便于机器学习模型识

<http://journal.xidian.edu.cn/xdxb>

别。这一过程的一个关键步骤是对数据的标准化和归一化。二者的目的是将原始数据限定在一定的范围内,从而降低奇异样本数据产生的负面影响。归一化是对原始数据进行变换,并按照不同的处理方式固定到某个区间中(如图像区间为 $[0, 255]$ 等)。归一化数学表达如下:

$$\frac{X_i - X_{\min}}{X_{\max} - X_{\min}}, \quad (1)$$

其中, X_i 表示当前样本的值, X_{\max} 表示样本最大值, X_{\min} 表示样本最小值。

标准化对数据进行变换,使其符合均值为0、标准差为1的分布。标准化数学表达如下:

$$\frac{X_i - \mu}{\sigma}, \quad (2)$$

其中, μ 表示样本数据的均值, σ 表示样本数据的标准差。

2.3 恶意特征识别

将转化后的流量集导入 CNN、RF、SVM、聚类等机器学习模型进行训练以识别恶意特征;识别结果反馈信息给训练模型,通过模型设计与参数调优等方法得到理想的准确度,最终实现将流量进行良性和恶意的二分类,并进一步地对恶意流量进行细粒度的分类。图3总结了这一过程,表3归纳了基于机器学习的加密流量中的恶意流量识别相关工作。下文将对相关工作按照不同特征集构建方式进行详细介绍。

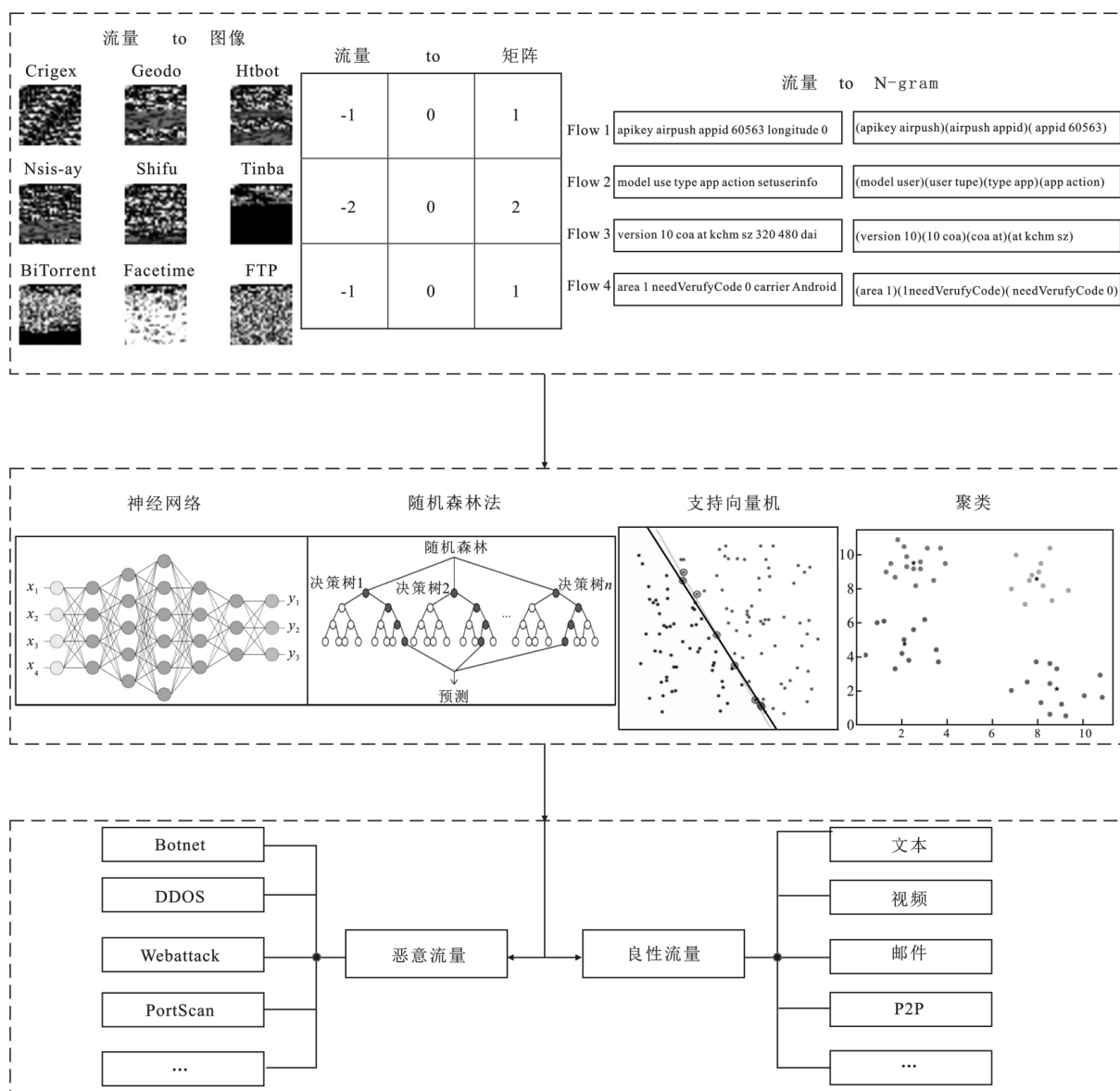


图3 机器学习流程图

<http://journal.xidian.edu.cn/xdxb>

表 3 基于机器学习的恶意流量识别小结

文献	识别特征	识别方法	数据集	识别目的	详细描述	准确率
[56]	时空特征	CNN	DARPA1998 ^[119] / ISCX2012 ^[120]	二分类:识别良性/ 恶意流量	利用深层神经网络学习原始流量数据的时空特征	最高 99.96%
[59]	头部特征	随机森林	CTU-13 ^[121] / MCFP	二分类:识别良性/ 恶意流量	只从流量开始的大约 8 个数据包中提取特征,可以提前检测到恶意应用流量	未提及
[60]	头部特征	聚类	私有数据集 和 Drebin ^[122]	多分类:识别恶意 应用流量	利用网络流量信息中多维应用层数据的恶意应用分类和检测	最高 90%
[61]	头部特征/ 统计特征	无监督学习 算法	CICIDS 2017 ^[123]	多分类:识别恶意 应用流量	用高斯混合模型和排序点识别聚类结构,计算恶意应用之间的距离	平均 91.74%
[62]	头部特征/ 负载特征	CNN	USTC-TFC 2016	多分类:识别恶意 应用流量	使用原始流量数据转化为图像,利用 CNN 进行图像分类	平均 99.41%
[63]	头部特征/ 负载特征	CNN	CTU-13	二分类:识别良性/ 恶意流量	从连接元数据中提取出上下文特征,使用 Perlin 噪声将特征编码到图像中	最高 97%
[68]	头部特征/ 负载特征	Svm	私有数据集	二分类:识别良性/ 恶意流量	将移动流量视为文档,使用 NLP 执行恶意应用检测	最高 99.15%
[69]	头部特征/ 负载特征	多视图神经网络	私有数据集	二分类:识别良性/ 恶意流量	设计了一种利用应用程序访问的 url 来识别恶意应用程序的方法。	最高 98.75%
[76]	统计特征/ 头部特征	C 4.5	私有数据集	多分类:识别恶意 应用流量	网络流量分析与 C 4.5 相结合识别 Android 恶意应用	最高 99.65%
[77]	统计特征	自组织特征 映射	VirusTotal API	多分类:识别恶意 应用流量	使用难以混淆的机器数据对恶意应用进行分类	最高 89%
[78]	统计特征	模块相似性	私有数据集	二分类:识别良性/ 恶意流量	基于行为的检测体系结构,使用相似性度量来检测入侵	未提及
[79]	统计特征	随机森林	私有数据集	多分类:识别恶意 应用流量	在加密的 Android 应用程序流量中指纹识别和实时识别	最高 99%

时空特征包括流量的时间特征和空间特征。WANG 等^[56]提出了基于分层时空特征的入侵检测系统(HAST-IDS)。首先使用深度卷积神经网络(CNN)学习网络流量的浅层空间特征,然后使用长短期记忆网络(LSTM)学习深层时间特征。实验表明,使用时空特征相结合方式的识别结果要优于只使用其中一种方法的识别结果。类似地,KIM 等^[57]通过同时查看时间和空间中的多个计算元素来检测恶意流量,并建立了一个随机图模型来表示网络攻击行动在时间和空间上的组合。该系统能够有效地检测到未知攻击。

头部特征包括流量包头包含的用户信息相关特征。ANDERSON 等^[58]通过学习加密流量的数据特征,提出了基于 TLS 握手数据包中暴露的版本号、密钥长度等非加密信息来检测恶意流量的方法。为了避免机器学习中手动选择特征的繁琐,LIU 等^[59]提出了 MalDetect,仅在流量开始时从大约 8 个数据包中提取特征,这使其能够在恶意应用行为产生实际影响之前而不是流量传输完成后检测恶意应用流量。LI 等^[60]引入了 DroidClassifier 系统框架,通过从多个 HTTP 标头字段中提取通用标识符来构建模型,并利用监督学习的方法进行识别。LIU 等^[61]计算恶意软件之间的距离并利用该距离的聚类结构定义新的恶意软件类,并利用无监督学习算法高斯混合模型(GMM)和排序点进行辨识。

负载特征包括流量包中的有效载荷部分。WANG 等^[62]提出了一种新的流量分类方法。将流量数据可

<http://journal.xidian.edu.cn/xdxb>

视为图像,再通过图像使用 CNN 进行分类。通过这种方法,可以实现端到端的恶意流量识别,并且能够满足实际应用的精度。BAZUHAIR 等^[63]从连接元数据中提取上下文特征,然后使用 Perlin 噪声将给定的连接特征编码为图像,最后训练深度学习模型进行二进制识别。WANG 等^[68]认为移动应用程序生成的每个 HTTP 流都可视为文本文档。因此处理自然语言的方式同样可以用于处理网络流量的语义,并提出了一种使用自然语言处理对网络流量文本语义检测的恶意应用检测模型。基于同样的思路,WANG 等^[69]设计了一种利用应用程序访问时 HTTP 流中的统一资源定位器(URL)来识别恶意应用程序的方法。

统计特征包括流量包平均包长、平均包间时延等特征。早在 2008 年,WRIGHT 等^[75]就证明了利用音素与 VoIP(Voice over Internet Protocol)编解码器在呈现这些音素时输出的数据包长度之间存在相关性。加密的 VoIP 包的长度可以用来识别通话中所说的短语。WANG 等^[76]通过收集 TLS 的 6 个统计特征(上传字节、下载字节大小等)和 HTTPS 流中的 4 个统计特征(用户代理、请求 URL 等),用 C 4.5 算法识别出恶意应用流量。BURNAP 等^[77]使用通过创建无监督聚类的方式建立应用的活动度标准来识别正常应用和恶意应用。类似地,NEU 等^[78]提出了一种新的基于相似度来检测恶意流量的检测体系结构。TAYLOR 等^[79]介绍了称为 AppScanner 的框架,用于自动加密和实时识别 Android 应用程序的加密网络流量。在物理设备上自动运行应用程序以收集其网络跟踪,通过网络跟踪生成指纹并用于流量识别。

加密电子设备在运行过程中产生的侧信道信息也可用于恶意应用识别。文献^[80]中提出了 HoMonit,在物联网环境下通过侧信道信息识别恶意智能应用。文章利用嗅探器从集线器和云平台之间收集泄露的侧信道信息(包大小和包间时序)。通过对加密流量进行流量分析来推断出智能设备和集线器之间的事件序列,然后从智能应用的源代码或 UI 接口中提取预期程序逻辑,二者通过 DFA(有穷自动机)算法进行匹配,从而识别出越权或是行为异常的恶意智能家居应用。

以上恶意行为主要表现在加密流量上,文献^[81]提出了一种基于机器学习的安全分析模型来识别出认证与密钥协商协议过程中对协议的攻击。相比于传统形式化协议中分析方案与分析精度往往取决于分析人员所掌握的先验知识和对协议的主观理解,将机器学习与协议形式化分析相结合是一种可行的尝试。但该方案目前尚缺乏足够的样本对每一种攻击进行细分类。

3 基于密码学的恶意流量识别

基于密码学的恶意流量识别的本质在于,检索流量中是否存在加密后的恶意关键字,即在不解密所有数据包的前提下,实现在一段加密过的信息上实现恶意关键词的搜索,这也是可搜索加密技术的一种应用。然而,检测的中间盒设备往往没有解密流量的权限或者密钥,不仅如此,可搜索加密无法对检测规则提供保护。因此,在网络流量中的恶意流量识别不能直接应用可搜索加密,而是需要深度融合可搜索加密技术、深度报文检测技术、流量审查机制和可证明安全模型,进行综合设计,使其可以在保护用户数据隐私以及检测方检测规则的前提下检索加密流量上的恶意关键词以识别恶意流量。基于密码学的恶意流量识别的关键技术如图 4 所示。笔者将从可搜索加密技术出发,介绍公钥可搜索加密和对称可搜索加密的技术难点,包括密文检索和密文计算。其中密文检索可依次区分为单关键词检索、多关键词检索、模糊关键词检索和区间检索。密文计算可分为同态加密和函数(属性)加密。最后介绍如何结合可搜索加密技术、深度报文检测以及流量审查等机制进行加密流量中的恶意流量检索。

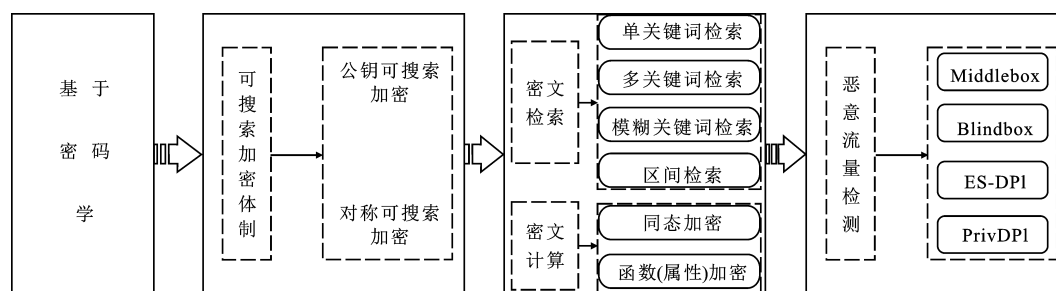


图4 基于密码学的恶意流量识别体系

<http://journal.xidian.edu.cn/xdxb>

3.1 可搜索加密体制

可搜索加密体制分为对称可搜索加密体制和公钥可搜索加密体制。SONG 等^[82]提出对称可搜索加密, 密钥拥有者可以查询检索密文, 但因为对称密码本身存在的密钥管理和分发问题, 导致这类方案在密钥管理的开销过大。因此, BONEH 等^[83]提出基于公钥的可搜索加密受到更为广泛的关注, 该体制中发送者通过接收者公钥进行关键字加密, 拥有对应私钥者可生成陷门进行密文搜索^[84]。

3.2 密文检索和密文计算

基于公钥可搜索加密的技术难点主要集中在密文检索和密文计算技术方向上。密文检索可以通过检索关键词的方式直接对密文数据进行访问^[85]。密文检索技术可以通过单个关键词、多个关键词、模糊关键词到区间检索恶意流量的特征。

单关键字检索复杂度与数据库的大小呈线性关系。为了提高搜索效率, GOH^[86]为每个文档构建了一个 Bloom 过滤器, 服务器可以使用 Bloom 过滤器来测试文档中是否有特定的关键字。CURTMOLA 等^[87]设计了一个基于关键字的倒排检索。在这种结构中, 服务器可以直接找到所查询关键字的所有搜索结果。但服务器可能会返回不完整或不正确的结果。因此, KUROSAWA 等^[88]在每个关键字的搜索结果上使用消息验证码, 确保搜索结果的完整性。

多关键词检索最早由 GOLLE 等^[89]提出, 该方案的检索复杂度与查询关键字数呈线性关系。为了解决检索效率低下的问题, CASH 等^[90]提出了不经意交叉标记方案, 搜索检索由 Tset 和 Xset 组成。服务器首先根据 Tset 检索匹配文档数相对最少的搜索结果, 然后根据 Xset 过滤包含其他查询关键字的搜索结果。PAPPAS 等^[91]提出 blind seer 方案, 树状图上的节点对应 Bloom 过滤器, 节点包括所有子节点中的所有记录, 通过遍历找到所需的结果。

模糊关键词检索可以检索不够精确的关键词。LI 等^[92]通过距离来估计两个关键词的相似度, 并利用通配符技术构造了模糊关键词搜索方案。但该方案必须包含所有可能的错误关键字的检索, 这将导致检索冗余。GIONIS^[93]构造了另一种模糊关键字方案, 其主要思想是局部敏感散列技术可以以很高的概率将相似的项映射到相同的哈希值。然而, 一些正确的搜索结果可能无法检索到。考虑恶意服务器可能返回错误, SUN 等^[94]提出了一种基于符号树的可验证模糊关键字搜索方案。该方案支持模糊关键字搜索, 同时也实现了搜索者对排名关键字搜索的可验证性。

区间检索^[95]是数据检索的重要方式之一。被授权用户利用私钥生成陷门, 服务器检索陷门后返回对应密文, 再由用户解密得到明文。为了确保搜索关键词的可用性, AGRAWAL 等^[96]提出了基于保序加密的区间检索, 但此方案没有隐藏关键词的信息, 因而存在隐私泄露的风险。CAI 等^[97]提出了一种单断言的区间检索方案, 该方案被证明是安全的, 不会泄露敏感数据的特征。

全同态加密最早由 GENTRY^[98]设计完成。全同态加密公钥方案含有 4 个算法: 密钥生成算法 (KeyGen)、加密算法 (Encrypt)、解密算法 (Decrypt) 和密文计算算法 (Evaluate)。全同态加密的含义是对明文加密后, 在密文上进行任意计算, 其结果等同于对应明文计算结果, 即

$$\epsilon = (\text{KeyGen}, \text{Encrypt}, \text{Decrypt}, \text{Evaluate}) \quad (3)$$

$$f(\text{Enc}(\mu)) = \text{Enc}(f(\mu)) \quad (4)$$

其中, ϵ 表示全同态加密算法, f 表示进行任意计算, Enc 表示加密过程, μ 代表明文数据。

由于全同态加密的特性, 使其可以应用在密文检索技术之中, 这样既可以保证用户数据的安全, 也可以在一定程度上提高密文检索的效率。

密文计算主要有两类: 一类是同态加密, RIVEST 等^[99]首次提出同态加密的概念, 同态加密可以直接应用于密文操作, 并在解密后得到对应的明文; 另一类是函数(属性)加密, SAHAI 等^[100]提出利用属性基加密 (ABE) 方案, 用户信息由用户属性组成, 从而实现密文访问控制。LI 等^[101]提出了一个新的安全外包 ABE 系统, 该系统同时支持安全外包密钥发布和解密。该构造以有效的方式提供了外包计算结果的可检查性。SHAMIR^[102]提出利用身份基加密 (IBE) 方案, 用户身份即为公钥, 解决公钥认证繁琐的问题。王贇玲^[103]提出支持服务器端解密匹配的匿名属性基加密方法, 在不泄露用户属性的前提下进行高效检索。

<http://journal.xidian.edu.cn/xdxb>

3.3 恶意流量检测

目前的可搜索加密技术主要应用于数据库上,在保证数据机密性的同时实现密文数据的高效检索。虽然这项技术与在加密流量上检索特征具备一定的相似性,但是可搜索加密技术不能直接应用于流量检测,其原因是加密流量检测工作应用层或者 TCP/IP 层中,需要在用户数据保护、加密规则保护、密钥管理、检索算法效率等方面进行深度融合。首先,可搜索加密技术虽然可以防止用户数据泄露,但是无法保证用户掌握检测规则后通过混淆来逃避检测。将可搜索加密技术应用于流量检测的同时还需要保护加密规则不泄露。此外检测效率的问题也制约了可搜索加密在恶意流量检测上的应用。其次,可搜索需要密钥,规则保护也需要密钥,这两种密钥生成与管理不是简单的协议组合,需要严格的密码学意义上的安全协议来保证,同时还需要和网络中间盒(Middlebox)上的流量检测协议深度融合。最后,恶意流量识别需要高效的算法,否则过高的设置延迟和开销大小在实际应用中是不现实的。因而学者开始探索如何行之有效地在 Middlebox 上检索恶意流量的模糊特征以识别恶意流量。

网络中间盒(Middlebox)是大型网络的一部分,实现与安全性(例如防火墙和入侵检测)和性能(例如缓存和负载平衡)相关的多种功能。Middlebox 通过执行深度包检测(DPI)以检测网络流量中的异常和可疑活动,然而一旦数据包以加密的方式发送,Middlebox 即面临失效,因为深度包检测需要对有效负载进行分析,而 Middlebox 没有权限解密有效负载。文献[106-113]通过采用密文检索的相关技术,以 Middlebox 技术为基础,提出了对加密数据包上进行模糊关键词检索,以检测加密流量中的恶意流量。

Middlebox 需要以不解密负载的方式执行深度包检测以保护用户数据,但迫于硬件设备不足的压力,Middlebox 需要在不可信的外包设备上运行。为了保护用户数据不被恶意泄露,GOLTZSCHE 等^[104]设计了 Endbox 系统,Endbox 允许在不可信的客户端机器上安全地部署并确保 Middlebox 功能。该系统允许将部分 Middlebox 功能外包给用户,利用用户的 CPU 资源进行计算。HUNT 等^[105]设计了 Ryoan,通过沙箱保护系统免受潜在的恶意攻击和防止用户数据泄露,并允许各方在不信任的基础架构上以分布式方式处理敏感数据,但目前沙箱的安全性暂未得到有效证明。文献[104-105]虽然使用分布式和沙箱的方法保护了用户隐私,但是没有实现对加密规则的保护,因此无法识别出用户的恶意行为。

保护用户数据隐私的同时也需要对加密规则进行保护,因为加密规则一旦泄露,用户就可以绕过规则检测,使得 Middlebox 失效。为了能够同时实现 Middlebox 的功能性和对加密规则的保护,YUAN 等^[106]设计了一个安全的 DPI 系统,使外包的 Middlebox 能够对加密流量进行深度数据包检查,而不需要揭示包的有效载荷或检查规则。该系统建立了一个高性能的加密规则过滤器,过滤器安全地存储从规则中提取的加密字符串对,并对从数据包有效负载中提取的随机令牌进行加密检查。该系统实现了同时对数据包有效负载和检查规则的保护。SHERRY 等^[107]提出 BlindBox,这是第一个同时提供 Middlebox 的功能性和加密的隐私性这两个属性的系统。具体而言,BlindBox 利用乱码电路生成加密规则,直接对加密流量进行深度包检测而无需对底层流量进行解密。通过这样的方法,BlindBox 既可以进行深度包检测,又不需要获取过多的负载信息,保护了用户的隐私。文献[106-107]保护了加密数据和加密规则,但在实际应用中,每个会话都需要设置加密规则,这导致了设置延迟和开销大小都很高,并且由于通信只能在生成加密规则之后才能开始,所以这种延迟在许多实时应用中是不可容忍的。

为了最大限度地减少设置阶段的计算和通信开销,许多研究者开始探索更好的加密规则生成算法和密钥管理方法以同时保持 BlindBox 相同的属性和隐私要求。REN 等^[108]提出了一种高效安全的深度包检测方案 ES-DPI。该方案中网关对数据包中的位串进行令牌化,同时混淆加密令牌的位置顺序以防止令牌位置顺序泄漏导致的问题,并对令牌进行加密以保护隐私。中间盒部分利用两个非共谋云服务器进行规则匹配:令牌过滤服务器过滤掉大多数不匹配的令牌,规则匹配服务器执行单个关键字搜索,以确定是否有任何单个规则的关键字匹配,在保护包负载和中间盒规则的同时实现数据效用。该方案中加密规则集的安全性建立在文献[109]的基础之上,同时使用了文献[110]中的协议以防止中间人攻击。NING 等^[111]提出基于模糊规则的深度包检测技术 PrivDPI,通过模糊规则为每个会话产生新的加密规则。该方法规则生成快而且会话开销小,并且同样满足深度包检测和隐私要求。另外,PrivDPI 中采用了高效的密钥管理方案:用户和服务端通过 TLS 握手协议生成会话密钥 S_k , S_k 基于伪随机数发生器生成另外 3 个密钥。其中, k_{TLS} 用于加密流

量, k 用来生成可重用的模糊规则和会话规则, k_{rand} 用做生成随机数的种子。该方案在随机预言机(Random Oracle)模型下给出了严格的密码学安全证明。

PrivDPI 的优势在于: MB 无需学习客户 C 和服务器 S 的密钥, 而 C 和 S 也无需学习 MB 的规则, 保证了数据通信与规则的安全和隐私。MB 只需要生成一次模糊规则, 就可以通过重用为每个新会话生成加密规则, 有效地减少了规则准备中的计算和通信开销。该方法与 BlindBox 相比, 计算规则生成快而且会话开销小, 并且同样满足深度包检测和隐私要求。PrivDPI 流程如图 5 所示。

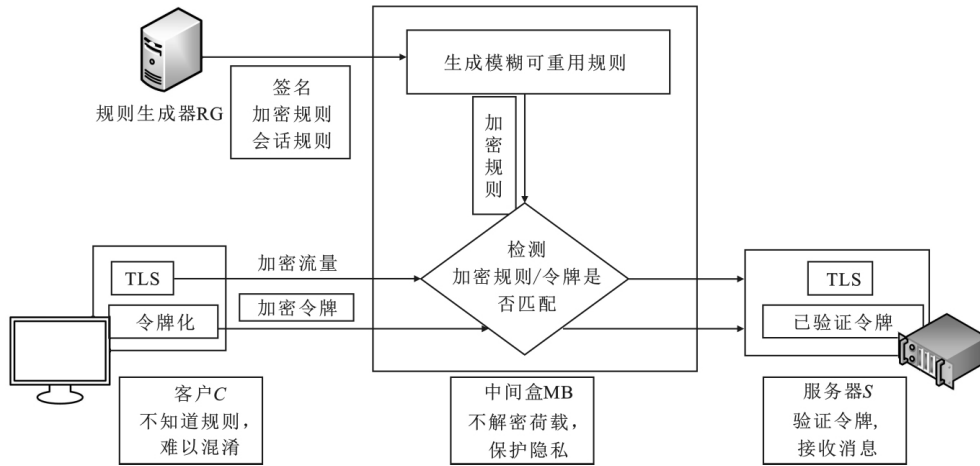


图 5 PrivDPI 示意图

PrivDPI 有两个数据流, 一个是 TLS 会话, 客户 C 和服务器 S 通过 TLS 握手协议建立会话密钥。另一个基于滑动窗口将数据令牌化, 生成加密令牌 C_{t_i} :

$$C_{t_i} = H(\text{salt} + c_{t_i}, T_t) \quad (5)$$

其中, C_{t_i} 表示哈希后的加密令牌; H 表示哈希函数; c_{t_i} 表示计数器; salt 表示随机数, 其通过在密文任意固定位置插入随机字符串, 改变密文散列结果以防止频率分析攻击; T_t 表示加密令牌, 其基于以下公式计算出:

$$T_t = g^{\text{kat} + k^2} \quad (6)$$

其中, g 表示规则生成器 RG 生成的参数, α 表示随机数, k 表示会话密钥, t 表示准备阶段的令牌。

随后 MB 生成加密规则 C_{r_i} :

$$C_{r_i} = H(\text{salt} + c_{r_i}, I_i) \quad (7)$$

其中, C_{r_i} 表示哈希后的加密规则, I_i 表示可重用模糊规则。

中间盒通过 RG 生成的规则元组, 建立一组可重用的模糊规则 I_i :

$$I_i = g^{\text{kar}_i + k^2} \quad (8)$$

其中, r_i 表示规则生成器 RG 提供的初始规则。

随后 MB 检查 C_{r_i} 是否匹配 C_{t_i} ; 如果二者匹配, 则认为流量是恶意的。

PrivDPI 不能直接应用于物联网的场景, 因为物联网环境大多采用一对多或者多对多的通信模式。如果采用 PrivDPI 中一对一的握手方案, 则产生的通信开销是巨大的; 此外, 相应的密钥生成和管理同样会带来巨大开销。为此, 文献[112]提出了一种基于群组密钥协商的物联网加密流量检测方案(GKA_DPI)。GKA_DPI 使用 BlindBox 框架进行深度流量检测, 并通过动态组密钥协议降低功耗, 在不解密消息的情况下检测恶意流量, 保证传感器网络通信的安全性。该文同时在物联网中广泛使用的协议消息队列遥测传输协议(MQTT)上证明了 GKA_DPI 更适用于物联网群组通信, 并能有效保证传感器网络通信的安全性和前向/后向保密性。

然而上述方案本质上是静态的, 因为检测规则一经确定就无法随意更改, 添加任何规则都需要大量的预处理步骤和协议的重新实例化。为此, NING 等[113]提出了改进方案 Pine, 在没有任何客户端参与的情况下, 通过网关设备在本地执行规则添加过程。Pine 允许动态添加新的规则而不影响连接, 同时减小了用户与服务器之间建立 TLS 连接的计算时间和通信开销, 并通过属性规则隐藏保证了规则的隐私性不受中间盒设备

的影响。从目前公开发表的文献看,Pine 在效率上达到了最高。

4 总 结

加密流量中的恶意流量识别问题是当前网络安全领域的热点和难点。笔者综述了最新的相关工作,将其分类为基于机器学习和基于密码学的两类识别框架,并对两大类的相关工作进行了总结。目前的工作已经在分类效率上与可证明安全取得了较大进展,但是还存在以下问题。

(1) 基于机器学习的恶意流量识别核心在于正确的数据集,正如开源数据集 ImageNet^[114]和 COCO^[115]为计算机视觉领域的突飞猛进提供了重要支持。加密流量也需要一个开源、有正确标签并且在恶意流量上有详细分类的数据集。另外,QUIC^[116]/HTTP 3^[117]等新型协议的出现使得流量数据包中头部的明文占比进一步下降,这对基于机器学习的恶意流量识别带来了进一步的挑战。

(2) 机器学习领域对抗样本的飞速进展使得流量混淆与伪装的难度进一步降低,攻击者可通过学习流量数据,添加合适的噪声将攻击流量伪装成正常流量,甚至误导模型将正常流量识别为攻击流量;另外,基于机器学习的恶意流量识别能力会随着时间流逝而逐渐下降,同时在训练/测试集中的恶意流量分布必须符合现实分布^[118]。如何应对流量混淆和时间衰减等问题将是下一步的研究重点。

(3) 目前的基于密码学的恶意流量识别建立在关键词搜索之上,其本质上类似黑/白名单机制,攻击者可通过混淆技术改变自己恶意特征的关键词,从而混淆与正常流量的边界,达到逃避检测的目的。因此如何行之有效地应对加密流量混淆技术,将成为下一步的研究重点之一。

(4) 基于中间盒的深度包检测技术在加密规则生成、数据包令牌化和加密规则匹配上都需要高效安全的算法支撑,否则过高的设置延迟和运算开销在实际应用中是不现实的。而全同态加密算法将为基于密码学的恶意流量识别技术带来新的曙光。

参考文献:

- [1] MEEKER, M. Internet Trends (2019)[R/OL]. [2020-12-01]. <https://www.bondcap.com/report/itr19/>.
- [2] CISCO. Encrypted Traffic Analytics (2018)[R/OL]. [2020-12-01]. <https://www.cisco.com/c/dam/en/us/solutions/collateral/enterprise-networks/enterprise-network-security/nb-09-encrytd-traf-anlytcs-wp-cte-en.pdf>.
- [3] 陈良臣,高曙,刘宝旭,等.网络加密流量识别研究进展及发展趋势[J].信息安全,2019(3):19-25.
CHEN Liangchen,GAO Shu,LIU Baoxu,et al. Research Status and Development Trends on Network Encrypted Traffic Identification[J]. Netinfo Security,2019(3):19-25.
- [4] 潘吴斌,程光,郭晓军,等.网络加密流量识别研究综述及展望[J].通信学报,2016,37(9):154-167.
PAN Wubin,CHENG Guang,GUO Xiaojun,et al. Review and Perspective on Encrypted Traffic Identification Research [J]. Journal on Communications,2016,37(9):154-167.
- [5] REZAEI S, LIU X. Deep Learning for Encrypted Traffic Classification: An Overview [J]. IEEE Communications Magazine,2019,57(5):76-81.
- [6] 翟明芳,张兴明,赵博.基于深度学习的加密恶意流量检测研究[J].网络与信息安全学报,2020,6(3):66-77.
ZHAI Mingfang,ZHANG Xingming,ZHAO Bo. Survey of Encrypted Malicious Traffic Detection Based on Deep Learning [J]. Chinese Journal of Network and Information Security,2020,6(3):66-77.
- [7] ZHAO J, MASOOD R, SENEVIRATNE S. A Review of Computer Vision Methods in Network Security(2020)[J/OL]. [2020-05-07]. <https://arxiv.org/abs/2005.03318v1>.
- [8] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009:248-255.
- [9] BAHRAMALI A, HOUMANSADR A, SOLTANI R, et al. Practical Traffic Analysis Attacks on Secure Messaging Applications[C]//Proceedings of the 2020 Network and Distributed System Security Symposium (NDSS). Piscataway: IEEE, 2020:508.
- [10] GRUBBS P. Pancake: Frequency Smoothing for Encrypted Data Stores[C]//Proceedings of the 29th USENIX Security Symposium. Piscataway: IEEE, 2020:2451-2468.

<http://journal.xidian.edu.cn/xdxb>

- [11] DRAPER-GIL G, LASHKARI A H, MAMUN M S I, et al. Characterization of Encrypted and VPN Traffic using Time-Related Features [C]//Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP). Piscataway: IEEE, 2016, 2: 407-414.
- [12] LOTFOLLAHI M, ZADE R S H, SIAVOSHANI M J, et al. Deep Packet: A Novel Approach for Encrypted Traffic Classification Using Deep Learning[J]. Soft Computing, 2017, 24(3): 1999-2012.
- [13] BAGUI S, FANG X, KALAIMANNAN E, et al. Comparison of Machine-Learning Algorithms for Classification of VPN Network Traffic Flow Using Time-Related Features[J]. Journal of Cyber Security Technology, 2017, 1(2): 108-126.
- [14] Wang W, ZHU M, WANG J L, et al. End-To-End Encrypted Traffic Classification with One-Dimensional Convolution Neural Networks[C]//Proceedings of the 2017 IEEE International Conference on Intelligence and Security Informatics. Piscataway: IEEE, 2017: 43-48.
- [15] GUO L, WU Q, LIU S, et al. Deep Learning-Based Real-Time VPN Encrypted Traffic Identification Methods[J]. Journal of Real-Time Image Processing, 2020, 17(1): 103-114.
- [16] 唐舒烨, 程光, 蒋泊森, 等. 基于分段熵分布的 VPN 加密流量检测与识别方法[J]. 网络空间安全, 2020, 11(8): 23-27.
TANG Shuye, CHENG Guang, JIANG Bomiao, et al. Detection and Recognition of VPN Encrypted Traffic Based on Segmented Entropy Distribution[J]. Information Security and Technology, 2020, 11(8): 23-27.
- [17] ZHOU K, WANG W, WU C, et al. Practical Evaluation of Encrypted Traffic Classification Based on A Combined Method of Entropy Estimation and Neural Networks[J]. ETRI Journal, 2020, 42(3): 311-323.
- [18] CASINO F, CHOO K K R, PATSAKIS C. HEDGE: Efficient Traffic Classification of Encrypted and Compressed Packets [J]. IEEE Transactions on Information Forensics and Security, 2019, 14(1): 2916-2926.
- [19] ACETO G, CIUONZO D, MONTIERI A, et al. Toward Effective Mobile Encrypted Traffic Classification through Deep Learning[J]. Neurocomputing, 2020, 409: 306-315.
- [20] NIU W, ZHUO Z, ZHANG X, et al. A Heuristic Statistical Testing Based Approach for Encrypted Network Traffic Identification[J]. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3843-3853.
- [21] OKADA Y, ATA S, NAKAMURA N, et al. Application Identification from Encrypted Traffic Based on Characteristic Changes by Encryption [C]//Proceedings of the 2011 IEEE International Workshop Technical Committee on Communications Quality and Reliability(CQR). Piscataway: IEEE, 2011: 1-6.
- [22] HE G, YANG M, LUO J, et al. Inferring Application Type Information from Tor Encrypted Traffic[C]//Proceedings of the Second International Conference on Advanced Cloud & Big Data. Piscataway: IEEE, 2014: 220-227.
- [23] HE G F, YANG M, LUO J Z, et al. A Novel Application Classification Attack Against Tor [J]. Concurrency and Computation Practice and Experience, 2015, 27(18): 5640-5661.
- [24] ALMUBAYED A, HADI A, ATOUM J. A Model for Detecting Tor Encrypted Traffic using Supervised Machine Learning[J]. International Journal of Computer Network and Information Security, 2015, 7(7): 10-23.
- [25] CONTI M, MANCINI L V, SPOLAOR R, et al. Analyzing Android Encrypted Network Traffic to Identify User Actions [J]. Information Forensics and Security IEEE Transactions on, 2016, 11(1): 114-125.
- [26] FU Y, XIONG H, LU X, et al. Service Usage Classification with Encrypted Internet Traffic in Mobile Messaging Apps [J]. IEEE Transactions on Mobile Computing, 2016, 15(11): 2851-2864.
- [27] SHEN M, WEI M, ZHU L, et al. Classification of Encrypted Traffic with Second-Order Markov Chains and Application Attribute Bigrams[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(8): 1830-1843.
- [28] SUBAHI A, THEODORAKOPOULOS G. Detecting IoT User Behavior and Sensitive Information in Encrypted IoT-App Traffic[J]. Sensors, 2019, 19(21): 4777.
- [29] KEARANS M J. The Computational Complexity of Machine Learning[M]. Massachusetts: The MIT Press, 1990.
- [30] RIVEST R L. Cryptography and Machine Learning[C]//Proceedings of the International Conference on the Theory and Application of Cryptology. Heidelberg: Springer, 1991: 427-439.
- [31] 张经纬, 舒辉, 蒋烈辉, 等. 公钥密码算法识别技术研究[J]. 计算机工程与设计, 2011, 32(10): 3243-3246.
ZHANG Jingwei, SHU Hui, JIANG Liehui, et al. Research on Public Key's Cryptography Algorithm Recognition Technology[J]. Computer Engineering and Design, 2011, 32(10): 3243-3246.
- [32] POORJA M. Classification of Ciphers[D]. Varanasi: Indian Institute of Technology, 2001.
- [33] MISHRA S, BHATTACHARJYA A. Pattern Analysis of Cipher Text: A Combined Approach[C]//Proceedings of the

- 2013 International Conference on Recent Trends in Information Technology(ICRTIT). Piscataway:IEEE,2013:393-398.
- [34] DIMOVSKI A, GLIGOROSKI D. Generating Highly Nonlinear Boolean Functions Using A Genetic Algorithm [C]// Proceedings of the 6th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service. Piscataway:IEEE,2003:604-607.
- [35] MILLAN W, CLARK A, DAWSON E. Heuristic Design of Cryptographically Strong Balanced Boolean Functions [C]// Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques. Heidelberg: Springer,1998:489-499.
- [36] SEGHER A, LI J. ParallelSteepest Ascent Hill-Climbing for High Nonlinear Boolean and Vectorial Boolean Functions (S-Boxes) [C]// Proceedings of the International Conference on Information and Communications Security. Heidelberg: Springer,2019:413-429.
- [37] ALANI M M. Neuro-Cryptanalysis of DES [C]// Proceedings of the World Congress on Internet Security (WorldCIS-2012). Piscataway:IEEE,2012:23-27.
- [38] HETTER B, GEHRER S, GUNEYSU T. DeepNeural Network Attribution Methods for Leakage Analysis and Symmetric Key Recovery [C]// Proceedings of the International Conference on Selected Areas in Cryptography. Heidelberg: Springer,2019:645-666.
- [39] CHERVYAKOV N I. The EC Sequences on Points of an Elliptic Curve Realization Using Neural Networks [J]. Advances in Intelligent Systems and Computing,2016,427:147-154.
- [40] AGOSTA G, BARENGHI A, PELOSI G. Compiler-Based Techniques to Secure Cryptographic Embedded Software Against Side Channel Attacks [J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems,2020 39(8):1550-1554.
- [41] BATINA L, BHASIN S, JAP D, et al. Poster: Recovering the Input of Neural Networks Via Single Shot Side-Channel Attacks [C]// Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM,2019:2657-2659.
- [42] LIU F F, YAROM Y, GE Q, et al. Last-Level Cache Side-Channel Attacks Are Practical [C]// Proceedings of the 2015 IEEE Symposium on Security and Privacy (SP). Piscataway:IEEE,2015:605-622.
- [43] DILEEP A D, SEKHAR C C. Identification of Block Ciphers Using Support Vector Machines [C]// Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings. Piscataway:IEEE,2006:2696-2701.
- [44] CHOU J, LIN S, CHENG C M. On the Effectiveness of Using State-Of-The-Art Machine Learning Techniques to Launch Cryptographic Distinguishing Attacks [C]// Proceedings of the 5th ACM Workshop on Security and Artificial Intelligence. Piscataway:IEEE,2012:105-109.
- [45] NAGIREDDY S. A pattern recognition approach to block cipher identification [D]. Madras: Master of Science Dissertation- Indian Institute of Technology,2008.
- [46] SONI A, KARNICK H, AGARWAL M. Learning encryption algorithms from ciphertext [R]. Kanpur: Dept of Computer Science and Engineering, IIT Kanpur,2009.
- [47] SHARIF S O, MANSOOR S P. Performance Evaluation of Classifiers Used for Identification of Encryption Algorithms [C]// Proceedings of the International Conference on Advances in Information & Communication Technologies. Piscataway:IEEE,2011:42-45.
- [48] SUHAILA O S, KUNCHEVA L I, MANSOOR S P. Classifying Encryption Algorithms Using Pattern Recognition Techniques [C]// Proceedings of the 2010 IEEE International Conference on Information Theory and Information Security. Piscataway:IEEE,2011:1168-1172.
- [49] SOUZA W A R D, ALLAN T. A Distinguishing Attack with A Neural Network [C]// Proceedings of the 2013 IEEE 13th International Conference on Data Mining Workshops. Piscataway:IEEE,2013:154-161.
- [50] BHATEJA A K, DIN M. ANN Based Distinguishing Attack on RC4 Stream Cipher [C]// Proceedings of the 7th International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Piscataway:IEEE,2013: 101-109.
- [51] WU Y, WANG T, XING M, et al. BlockCiphers Identification Scheme Based on The Distribution Character of Randomness Test Values of Ciphertext [J]. Journal on Communications,2016,36(4):147.
- [52] 黄良韬,赵志诚,赵亚群. 基于随机森林的密码体制分层识别方案 [J]. 计算机学报,2018,41(2):382-399.

- HUANG Liangtao, ZHAO Zhicheng, ZHAO Yaquin. A Two-Stage Cryptosystem Recognition Scheme Based on Random Forest[J]. Chinese Journal of Computers, 2018, 41(2): 382-399.
- [53] 赵志诚. 基于机器学习的密码体制识别研究[D]. 郑州: 战略支援部队信息工程大学, 2018.
- [54] ZHAO Z, ZHAO Y, LIU F. The Research of Cryptosystem Recognition Based on Randomness Test's Return Value[C]// Proceedings of the International Conference on Cloud Computing and Security. Heidelberg: Springer, 2018: 3-15.
- [55] ZHAO Z, ZHAO Y, LIU F. Research on Grain-128's Cryptosystem Recognition[C]// Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). Piscataway: IEEE, 2018: 2013-2017.
- [56] WANG W, SHENG Y, WANG J, et al. HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection[J]. IEEE Access, 2018, 6: 1792-1806.
- [57] KIM J, KIM H S. Intrusion Detection Based on Spatiotemporal Characterization of Cyberattacks[J]. Electronics, 2020, 9(3): 460.
- [58] ANDERSON B, DAVID M G. Identifying Encrypted Malware Traffic with Contextual Flow Data[C]// Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security. New York: ACM, 2016: 35-46.
- [59] LIU J, ZENG Y, SHI J, et al. MalDetect: A Structure of Encrypted Malware Traffic Detection[J]. Computers, Materials and Continua, 2019, 60(2): 721-739.
- [60] LI Z, SUN L, YAN Q, et al. DroidClassifier: Efficient Adaptive Mining of Application-Layer Header for Classifying Android Malware[C]// Proceedings of the International Conference on Security & Privacy in Communication Systems. Heidelberg: Springer, 2016: 597-616.
- [61] LIU J, TIAN Z, ZHENG R, et al. A Distance-Based Method for Building an Encrypted Malware Traffic Identification Framework[J]. IEEE Access, 2019, 7: 100014-100028.
- [62] WANG W, ZHU M, ZENG X, et al. Malware Traffic Classification Using Convolutional Neural Network for Representation Learning[C]// Proceedings of the 2017 International Conference on Information Networking (ICOIN). Piscataway: IEEE, 2017: 712-717.
- [63] BAZUHAIR W, LEE W. Detecting Malign Encrypted Network Traffic Using Perlin Noise and Convolutional Neural Network[C]// Proceedings of the 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). Piscataway: IEEE, 2020: 200-206.
- [64] NI S, QIAN Q, ZHANG R. Malware Identification Using Visualization Images and Deep Learning[J]. Computers & Security, 2018, 77: 871-885.
- [65] SU J, VARGAS D V, PRASAD S, et al. Lightweight Classification of IoT Malware based on Image Recognition[C]// Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). Piscataway: IEEE, 2018: 664-669.
- [66] VENKATRAMAN S, ALAZAB M, VINAYAKUMAR R. A Hybrid Deep Learning Image-Based Analysis for Effective Malware Detection[J]. Journal of Information Security and Applications, 2019, 47: 377-389.
- [67] DARSHAN S L S, CD J. Windows Malware Detector Using Convolutional Neural Network Based on Visualization Images[J]. IEEE Transactions on Emerging Topics in Computing, 2019, 99: 1.
- [68] WANG S, YAN Q, CHEN Z, et al. Detecting Android Malware Leveraging Text Semantics of Network Flows[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(5): 1096-1109.
- [69] WANG S, CHEN Z, YAN Q, et al. Deep and Broad Learning Based Detection of Android Malware via Network Traffic[C]// Proceedings of the 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). New York: ACM, 2018: 1-6.
- [70] WANG S, CHEN Z, YAN Q, et al. Deep and Broad URL Feature Mining for Android Malware Detection[J]. Information Sciences, 2019, 513: 600-613.
- [71] WANG W, WEI J, ZHANG S, et al. LSCDroid: Malware Detection Based on Local Sensitive API Invocation Sequences[J]. IEEE Transactions on Reliability, 2019, 69: 174-187.
- [72] KATO H. Android Malware Detection Scheme Based on Level of SSL Server Certificate[J]. IEICE Transactions on Information and Systems, 2020, 103(2): 379-389.
- [73] REN B, LIU C, CHENG B, et al. MobiSentry: Towards Easy and Effective Detection of Android Malware on Smartphones

- [J]. Mobile Information Systems, 2018, 2018: 1-14.
- [74] YOUSEFI-AZAR M, HAMEY L, VARADHARAJAN V, et al. Malytics: A Malware Detection Scheme[J]. IEEE Access, 2018, 6: 49418-49431.
- [75] WRIGHT C V, BALLARD L, COULL S E, et al. Spot Me if You Can: Uncovering Spoken Phrases in Encrypted VoIP Conversations[C]//Proceedings of the 2008 IEEE Symposium on Security and Privacy (Sp 2008). Piscataway: IEEE, 2008: 35-49.
- [76] WANG S, CHEN Z, ZHANG L, et al. TrafficAV: An Effective and Explainable Detection of Mobile Malware Behavior Using Network Traffic[C]//Proceedings of the 2016 IEEE/ACM 24th International Symposium on Quality of Service (IWQoS). New York: ACM, 2016: 1-6.
- [77] BURNAP P, FRENCH R, TURNER F, et al. Malware Classification Using Self Organising Feature Maps and Machine Activity Data[J]. Computers & Security, 2018, 73: 399-410.
- [78] NEU C V, ZORZO A F, OROZCO A M S, et al. An Approach for Detecting Encrypted Insider Attacks on OpenFlow SDN Networks[C]//Proceedings of the 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST). Piscataway: IEEE, 2016: 210-215.
- [79] TAYLOR V F, SPOLAOR R, CONTI M, et al. AppScanner: Automatic Fingerprinting of Smartphone Apps from Encrypted Network Traffic[C]//Proceedings of the 2016 IEEE European Symposium on Security and Privacy (EuroS&P). Piscataway: IEEE, 2016: 439-454.
- [80] ZHANG W, MENG Y, LIU Y, et al. HoMonit: Monitoring Smart Home Apps from Encrypted Traffic[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2018: 1074-1088.
- [81] MA Z, LIU Y, WANG Z, et al. A Machine Learning-Based Scheme for The Security Analysis of Authentication and Key Agreement Protocols[J]. Neural Computing and Applications, 2018, 32(22): 16819-16831.
- [82] SONG D, WAGNER D, PERRIG A. Practical Techniques for Searches on Encrypted Data[C]//Proceeding of the 2000 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2000: 44-55.
- [83] BONEH D. Public Key Encryption with Keyword Search[C]//Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2004). Heidelberg: Springer, 2004: 506-522.
- [84] 陈晓峰, 王育民. 公钥密码体制研究与进展[J]. 通信学报, 2004(8): 109-118.
CHEN Xiaofeng, WANG Yumin. A Survey of Public Key Cryptography[J]. Journal of China Institute of Communications, 2004(8): 109-118.
- [85] WANG Y, WANG J, CHEN X. Secure Searchable Encryption: A Survey[J]. Journal of Communications and Information Networks, 2016, 1(4): 52-65.
- [86] GOH E. Secure Indexes (2003)[J/OL]. [2003-10-07]. <https://eprint.iacr.org/2003/216.pdf>.
- [87] CURTMOLA R, GARAY J, KAMARA S, et al. Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions[J]. Journal of Computer Security, 2011, 19(5): 895-934.
- [88] KUROSAWA K, OHTAKI Y. UC-Secure Searchable Symmetric Encryption [C]//Proceedings of the International Conference on Financial Cryptography and Data Security. Heidelberg: Springer, 2012: 285-298.
- [89] GOLLE P, STADDON J, WATERS B R. Secure Conjunctive Keyword Search over Encrypted Data[C]//Proceedings of the International Conference on Applied Cryptography and Network Security. Piscataway: IEEE, 2004: 31-45.
- [90] CASH D, JARECKI S, JUTLA C, et al. Highly-Scalable Searchable Symmetric Encryption with Support for Boolean Queries[J]. Advances in Cryptology-CRYPTO 2013, 2013, 8042: 353-373.
- [91] PAPPAS V, KRELL F, Vo B, et al. Blind Seer: A Scalable Private DBMS[C]//Proceedings of the 2014 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2014: 359-374.
- [92] LI J, WANG Q, WANG C, et al. Fuzzy Keyword Search over Encrypted Data in Cloud Computing[C]//Proceedings of the 2010 IEEE INFOCOM. Piscataway: IEEE, 2010: 441-445.
- [93] GIONIS A. Similarity Search in High Dimensions via Hashing[C]//Proceedings of the 25th International Conference on Very Large Data Bases. Edinburgh: Morgan Kaufmann, 1999: 518-529.
- [94] SUN W, WANG B, CAO N, et al. Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(11): 3025-3035.
- [95] SHI E, BETHENCOURT J, CHAN T, et al. Multi-Dimensional Range Query over Encrypted Data[C]//Proceedings of


- the 2007 IEEE Symposium on Security and Privacy(SP'07). Piscataway: IEEE, 2007: 350-364.
- [96] AGRAWAL R, KIERNAN J, SRIKANT R, et al. Order Preserving Encryption for Numeric Data[C]//Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2004: 563-574.
- [97] CAI K, ZHANG M, FENG D G. Secure Range Query with Single Assertion on Encrypted Data[J]. Chinese Journal of Computers, 2011, 34(11): 2093-2103.
- [98] GENTRY C. Fully Homomorphic Encryption Using Ideal Lattices[C]//Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing. New York: ACM, 2009: 169-178.
- [99] RIVEST R, ADLEMAN L, DERTOUZOS M. On Databanks and Privacy Homomorphism[J]. Foundations of Secure Computation, 1978, 4(11): 169-180.
- [100] SAHAI A, WATERS B. Fuzzy Identity-Based Encryption[C]//Proceedings of the 24th Annual International Conference on Theory and Applications of Cryptographic Techniques. Piscataway: IEEE, 2005: 457-473.
- [101] LI J, HUANG X, LI J W, et al. Securely Outsourcing Attribute-Based Encryption with Checkability[J]. IEEE Transactions on Parallel and Distributed Systems, 2014, 25(8): 2201-2210.
- [102] SHAMIR A. Identity-Based Cryptosystems and Signature Schemes[C]//Proceedings of the CRYPTO 84 on Advances in Cryptology. Heidelberg: Springer, 1984: 47-53.
- [103] 王冀玲. 云环境下密文数据的连接关键词检索技术研究[D]. 西安: 西安电子科技大学, 2019.
- [104] GOLTZSCHE D, RUSCH S, NIEKE M, et al. EndBox: Scalable Middlebox Functions Using Client-Side Trusted Execution[C]//Proceedings of the 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks. Piscataway: IEEE, 2018: 386-397.
- [105] HUNT T, ZHU Z, XU Y, et al. Ryoan: A Distributed Sandbox for Untrusted Computation on Secret Data[J]. ACM Transactions on Computer Systems, 2018, 35(4): 32.
- [106] YUAN X, WANG X, LIN J, et al. Privacy-Preserving Deep Packet Inspection in Outsourced Middleboxes[C]//Proceedings of the IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. Piscataway: IEEE, 2016: 1-9.
- [107] SHERRY J, LAN C, POPA R A, et al. BlindBox: Deep Packet Inspection over Encrypted Traffic[J]. ACM SIGCOMM Computer Communication Review, 2015, 45(5): 213-226.
- [108] REN H, LI H, LIU D, et al. Toward Efficient and Secure Deep Packet Inspection for Outsourced Middlebox[C]//Proceedings of the ICC 2019-2019 IEEE International Conference on Communications (ICC). Piscataway: IEEE, 2019: 1-6.
- [109] KAMARA S, MOATAZ T. Boolean Searchable Symmetric Encryption with Worst-Case Sub-Linear Complexity[C]//Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques. Heidelberg: Springer, 2017: 94-124.
- [110] FAN J, GUAN C, REN K, et al. SPABox: Safeguarding Privacy During Deep Packet Inspection at A MiddleBox[J]. IEEE/ACM Transactions on Networking, 2017, 25(6): 3753-3766.
- [111] NING J T, POH G S, CHIA J, et al. PrivDPI: Privacy-Preserving Encrypted Traffic Inspection with Reusable Obfuscated Rules[C]//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. Piscataway: IEEE, 2019: 1657-1670.
- [112] FAN Z Q, ZENG Y, ZHU X Y, et al. A Group Key Agreement Based Encrypted Traffic Detection Scheme for Internet of Things[C]//Proceedings of the 1st ACM International Workshop on Security and Safety for Intelligent Cyber-Physical Systems. New York: ACM, 2020: 19-26.
- [113] NING J T, HUANG X Y, POH G E, et al. Pine: Enabling Privacy-Preserving Deep Packet Inspection on TLS with Rule-Hiding and Fast Connection Establishment[C]//Proceedings of the European Symposium on Research in Computer Security. Heidelberg: Springer, 2020: 3-22.
- [114] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 248-255.
- [115] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context[C]//Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2014: 740-755.
- [116] IYENGAR J, THOMSON M. QUIC: A UDP-Based Multiplexed and Secure Transport[S/OL]. [2019-09-21]. [https://](https://http://journal.xidian.edu.cn/xdxb)

<http://journal.xidian.edu.cn/xdxb>

tools.ietf.org/html/draft-ietf-quic-transport-23.

- [117] SINHA G, KANAGARATHINAM M R, JAYASEELAN S R, et al. CQUIC: Cross-Layer QUIC for Next Generation Mobile Networks[C]//Proceedings of the 2020 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE, 2020: 1-8.
- [118] PENDLEBURY F, PIERAZZI F, JORDANEY R, et al. TESSERACT: Eliminating Experimental Bias in Malware Classification across Space and Time[C]//Proceedings of the 28th USENIX Security Symposium. Piscataway: IEEE, 2019: 729-746.
- [119] LIPPMAN R, CUNNINGHAM R, FRIED D, et al. Results of the DARPA 1998 Offline Intrusion Detection Evaluation(1998) [R/OL]. [1998-02-01]. <https://www.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-dataset>.
- [120] SHIRAVI A, SHIRAVI H, TAVALLAE M, et al. Toward Developing A Systematic Approach to Generate Benchmark Datasets for Intrusion Detection[J]. Computers and Security, 2012, 31(3): 357-374.
- [121] Stratosphere Research Laboratory. Index of/public Datasets/CTU-Malware-Capture-Botnet-42 (2020)[DS/OL]. [2020-01-05]. <https://mcfp.felk.cvut.cz/public Datasets/CTU-Malware-Capture-Botnet-42>.
- [122] ARP D, SPREITZENBARTH M, HUBNER M, et al. DREBIN: Effective and Explainable Detection of Android Malware in Your Pocket[C]//Proceedings of the Network and Distributed System Security Symposium. Piscataway: IEEE, 2014, 23-26.
- [123] CANADIAN INSTITUTE FOR CYBERSECURITY. Intrusion Detection Evaluation Dataset (CIC-IDS 2017)(2017)[R/OL]. [2017-12-31]. <http://www.unb.ca/cic/datasets/ids-2017.html>.

(编辑: 牛姗姗)




西安电子科技大学
XIDIAN UNIVERSITY

欢迎订阅

西安电子科技大学学报

JOURNAL OF XIDIAN UNIVERSITY



投稿网址: <http://journal.xidian.edu.cn/xdxb/CN/1001-2400/home.shtml>