

基于深度学习的加密恶意流量检测研究

翟明芳, 张兴明, 赵博

(国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘要: 随着网络安全防范意识增强, 加密通信占据主流, 加密流量快速增长。流量加密在保护隐私的同时, 也掩饰非法企图, 改变威胁形式。深度学习作为机器学习领域的重要分支, 是流量分类的有力工具。近年来, 将深度学习方法应用于入侵检测的研究不断深入, 取得良好效果。在深入调研文献的基础上, 将加密恶意流量检测的步骤总结归纳为“六步法”的一般检测框架模型, 结合模型对数据处理及检测算法进行回顾总结, 指出各类算法模型的优缺点, 并对未来研究方向进行展望, 以期为进一步研究提供帮助。

关键词: 加密流量; 恶意流量检测; 深度学习; 数据处理

中图分类号: TP391.1

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020034

Survey of encrypted malicious traffic detection based on deep learning

ZHAI Mingfang, ZHANG Xingming, ZHAO Bo

National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, China

Abstract: With the increasing awareness of network security, encrypted communication dominates and encrypted traffic grows rapidly. Traffic encryption, while protecting privacy, also masks illegal attempts and changes the form of threats. As one of the most important branch of machine learning, deep learning performs well in traffic classification. For several years, research on deep-learning based intrusion detection has been deepened and achieved good results. The steps of encrypted malicious traffic detection were introduced to be a general detection framework model named “six-step method”. Then, discussion and induction of data processing and detection algorithms were carried out combined with this model. Both advantages and disadvantages of various algorithm models were given as well. Finally, future research directions were pointed out with a view to providing assistance for further research.

Key words: encrypted traffic, malicious traffic detection, deep learning, data processing

收稿日期: 2019-11-07; 修回日期: 2020-01-21

通信作者: 翟明芳, 1353382456@qq.com

基金项目: 高安全等级网络基础设施关键装备核心芯片及软件研发项目 (2017ZX01030301)

Foundation Item: High Level Security Network Infrastructure Key Equipment Core Chip and Software Development Funding Project (2017ZX01030301)

论文引用格式: 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): 59-70.

ZHAI M F, ZHANG X M, ZHAO B. Survey of encrypted malicious traffic detection based on deep learning[J]. Chinese Journal of Network and Information Security, 2020, 6(3): 59-70.

1 引言

当今社会,信息通信技术飞速发展,网络安全问题日益突出,隐私泄露风险加剧。根据中国消费者协会测评报告^[1],有超过九成 App 涉嫌过度收集用户个人信息。2018年3月,Facebook 公司大规模数据泄露且被恶意利用;我国发生了十几亿条快递公司用户信息泄露^[2]等事件。频发的安全事件使隐私保护成为焦点,加密流量也呈飙升态势。Gartner 报告^[3]称:2019年有超过80%的流量为加密流量。图1^[4]显示了火狐网页访问中 HTTPS 流量占比的增长趋势。

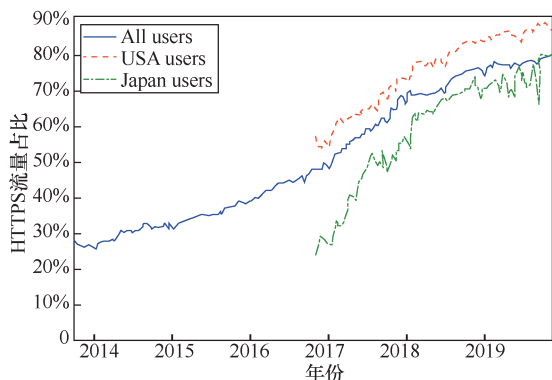


图1 火狐网页访问中 HTTPS 流量占比
Figure 1 Proportion of HTTPS traffic in Firefox Web page access

流量加密在保护用户隐私的同时,给不法分子以可乘之机。恶意移动应用普遍采用加密流量(如 HTTPS 等)传输网络数据以躲避检测^[5]。赛伦(Cyren)^[6]安全研究发现,有37%的恶意软件使用 HTTPS,每个主要的勒索软件家族都曾通过 HTTPS 进行传播。因此,对加密的恶意流量进行检测刻不容缓。

实际入侵检测过程中,加密与非加密恶意流量的差异主要表现在以下4方面。① 特征差异。二者的流量特征明显不同,部分非加密流量识别方法很难适用于加密流量,如 DPI 方法^[7]。② 复杂性差异。加密协议多样(如 SSL/TLS、SSH、P2P 等),缺少普适识别方法,常需要针对不同加密协议采取特定的识别方法,或采用多种识别策略集成的方法。③ 技术差异。恶意加密常采用流量伪装技术(如协议混淆和协议变种^[8]),把加密流量特征转换成普通流量特征。④ 精细化差异。

目前,加密恶意流量识别研究主要集中于二分类或少数几类攻击的识别,实现加密恶意流量精细化识别还存在一定的难度^[9]。

深度学习作为机器学习的一个重要分支,已经在图像处理^[10]、机器翻译^[11]、语音识别^[12]、情景分析^[13]等领域显示出优越的性能。目前,已有诸多研究^[14-18]将深度学习应用于入侵检测,并取得良好的效果。业界也投入大量人力、财力应用深度学习进行加密恶意流量检测。思科^[19-20]针对 TLS 协议,采用逻辑回归和十折交叉验证进行加密恶意流量识别,精度达到90%以上。国内观成科技^[21]分析提取了4类54个子类共计1 000多个数据特征,通过对比实验,得出随机森林精度最高,并获得了99.95%的检测率。

本文在深入查阅现有研究的基础上,对各种检测方法进行归类总结,归纳出基于“六步法”的加密恶意流量检测的一般框架。结合模型重点讨论数据处理、算法模型等加密恶意流量检测关键环节,指出各类方法的优势和不足,并对未来发展方向进行展望。

本文的贡献有以下几点:

- 1) 归纳总结“六步法”的加密恶意流量检测一般框架模型,为问题研究提供帮助;
- 2) 对现有数据集进行整理归类,指出各个数据集的优缺点和适用场景,更加方便研究者应用;
- 3) 讨论加密恶意流量检测中深度学习算法模型的应用和限制,提出算法改进方向。

2 加密恶意流量检测框架

加密恶意流量检测的本质是学习数据特征,将流量数据进行正确归类。加密成本的日益低廉使攻击者很容易利用加密协议隐藏非法企图。由于负载信息不再可见,识别加密网络流量中包含的威胁构成了一组独特的挑战。为鉴别恶意流量,研究者提出多种基于机器学习的检测方法,如文献[22-24]等。综合文献,本文将加密恶意流量检测归纳为“六步法”,模型如图2所示。

图2中,蓝色箭头表示加密恶意流量检测的步骤流程,红色箭头表示后续步骤对先前处理的影响作用。

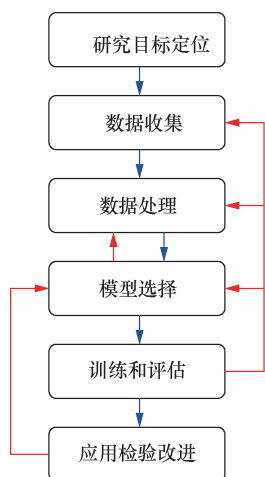


图2 “六步法”模型
Figure 2 “Six-step method” model

Rezaei 等^[25]提出了流量识别的一般框架,将流量识别分为 7 个步骤。尽管该框架适用于大多数数算法模型,但未能涵盖新颖的流量识别方法。Wang 等^[26]提出一维 CNN 分类模型,此模型未进行数据特征提取,仅对流量数据进行剪切处理,而后输入 1D-CNN 自行学习特征,进行分类。

与上述框架相比,“六步法”框架模型将数据预处理、特征提取等步骤整合为数据处理,可以很好地阐释各类不同的检测模型,因此,适用范围更广,能够涵盖绝大多数现有研究。另外,虽然“六步法”框架是针对加密恶意流量检测问题的,但对于普通的流量识别问题仍然适用,这也体现出该框架的普适性。

本文将对各个步骤逐一进行讨论,重点是数据收集、数据处理和模型选择。

3 研究目标定位

研究建立加密恶意流量分类器的第一步是定位研究目标。加密恶意流量检测是流量识别的一个子类。流量识别的目标是用相应的类来标记每条流。对于不同的研究目的,所针对的研究对象、收集的数据、提取的特征以及构建分类器算法均不相同。流量通常由 5 元组决定:源 IP、目标 IP、源端口、目标端口和协议^[27]。可按以下方面对流量进行分类。

- 1) 协议,如 SSH、SSL/TLS。
- 2) 应用,如微信、滴滴出行、百度地图。
- 3) 服务,如聊天、购物。

4) 网站,如谷歌、百度。

5) 用户行为,如浏览、下载、上传。

具体就加密流量识别而言,通常分为加密与未加密识别、加密协议识别、加密服务识别和恶意流量检测。图 3 给出了更为细致的分类。

加密与未加密识别主要将流量区分为加密流量和非加密流量。文献[27-29]对此进行了相关的研究,提出各种不同的方法进行流量识别。尽管取得不错的效果,但随着流量分析需求的提高,仅仅识别流量是否加密已经不能满足要求。加密与未加密识别通常是作为更加精细化识别的基础步骤,如在检出加密流量的基础上,进一步区分流量的应用程序或服务类型。

加密协议识别是根据协议的不同对加密流量进行归类。Jain^[30]研究了由不同优化器训练的卷积神经网络对协议识别的影响;Karagiannis 等^[31]提出了一种叫作 BLINC 的协议识别方法。目前的研究大多针对某类或者某几类协议,尚没有对所有协议进行分类的研究,且基于协议的流量识别仍然存在不够细粒度的问题。

加密服务识别是将流量按特征、服务类别或者应用程序进行分类。Gil 等^[32]针对 VPN 流量检测问题,提出了一组与时间相关的特征,采用 C4.5 与 KNN 两种算法进行实验,精度超过 80%。Wang 等^[33]提出基于 CNN 的加密流量分类方法,将数据截取为 784 byte,输入 CNN 自行学习数据特征进行分类,克服了人工提取特征困难的问题。Lotfollahi 等^[34]提出一维 CNN 与 EAV 相结合的框架对原始数据进行自动编码、分类,得到较好的结果。加密服务识别问题主要集中在对多类型识别,即同时对多种应用程序进行识别,以及提升识别的精度。

异常流量检测的目的在于检测出隐藏在加密流量之中的 DDoS、APT、Botnet 等恶意流量,是保障网络安全的重要手段。随着隐私保护重要性日益凸显,基于有效负载的传统识别方法^[35]不再适用,现在研究主要聚焦于不解密识别加密恶意流量,如基于统计特征、基于用户行为的方法。Nychis 等^[36]研究了基于熵的多个流分布相关,分析流头特征和行为特征的分布,提出使用时间序列进行异常检测更加有效。Koch 等^[37]使用基于行为的检测方法,利用命令序列以及

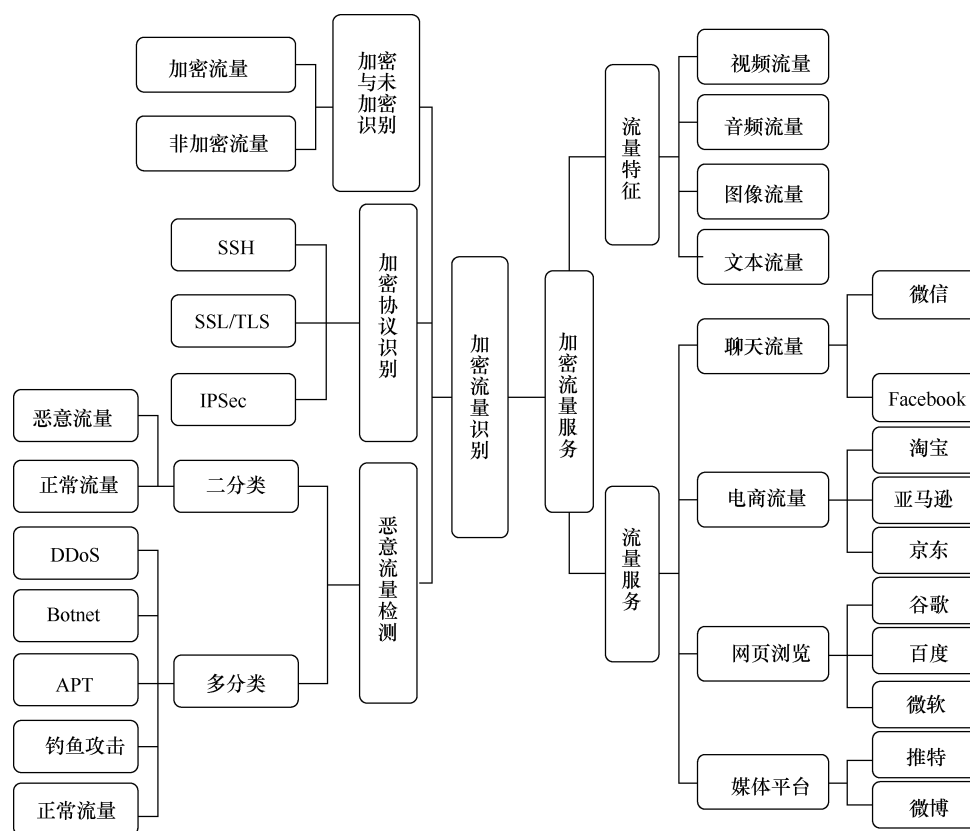


图3 加密流量识别分类结构

Figure 3 Classification structure of encrypted traffic identification

会话的关联性对恶意攻击进行检测。Ivan 等^[38]分析总结了4类40个数据特征,使用LSTM以及5折交叉验证进行检测恶意软件和钓鱼软件签名,分别得到94.87%和88.64%的正确率。Frantisek 等^[39]针对TLS加密恶意流量检测问题,比较了XGBoost、Random Forest、Neural Network、SVM这4种机器学习方法,得出XGBoost和Random Forest方法效果较好,测试精度达到95%左右。

与加密协议识别和加密服务识别相比,异常流量检测的难点有:①攻击者常常利用协议规则将攻击流量伪装成正常流量,增加检测的难度;②数据集严重不平衡,攻击流量远远小于正常流量,可能产生训练不充分的问题,从而影响检测精度;③检测错误代价大,因此对检测精度要求更高。

4 数据收集

找准了定位,就要收集相对应的数据。构建

可用的数据集是异常流量检测的关键基础环节,没有包含异常流量的加密数据集,模型建立及实验验证都无从谈起,因此数据收集对加密恶意流量检测至关重要。

4.1 数据集

无论是建立深度学习算法模型,还是验证模型性能,都需要标记数据集。Gharib 等^[40]建立一个评估框架,指出一个好的数据集应当包含以下11个特征。

- 1) 完整的网络配置。
- 2) 完整流量。
- 3) 已标记。
- 4) 完整交互。
- 5) 完整捕获。
- 6) 可用协议。
- 7) 攻击多样性。
- 8) 匿名性。
- 9) 异质性。
- 10) 特征集。

11) 元数据。

然而,目前既缺乏可以很好地符合异常加密流量识别或加密攻击流量识别的公开标签数据集,也没有普遍接受的数据收集和标记方法^[25]。研究者广泛采用私有自生成数据集进行实验,这又带来实验无法复现以及不同算法间无法对比的问题。表1中给出了部分恶意流量检测数据的比较。

从表1可以看出,大多数数据集并不针对加密流量,只有部分数据集包含加密协议流量。另外,一些数据集并未标记,或者没有特征标签类,这些都给研究带来困难。

4.2 数据收集方法

目前,常用的数据收集方法主要有直接收集法、脚本生成法以及混合方法。

4.2.1 直接收集法

直接收集法是最基本、最原始也最符合现实的收集方法。常用的数据采集软件有Wireshark、Sniffer、Fiddler等。部分学者就是利用此种方法采集私有数据集进行加密恶意流量检测。

直接收集法虽然能够采集到现实网络的真实数据,但也存在一些不足。①由于现实网络过于复杂,往往只能收集部分领域或者部门的数据,无法反映网络整体全貌。比如KDD99数据集来自模拟的美国空军局域网,UNIBS数据集是布雷西亚大学校园网的数据。②现实网络中攻击具有不确定性,且数量太少,一方面难以捕捉到所有攻击类型,另一方面可能造成数据的严重不平衡。③数据标记难度比较大。

4.2.2 脚本生成法

由于收集数据集比较困难,大多数数据集均采用脚本或者虚拟网络产生(如CICIDS2017数据集、ISCX-2012数据集)。采用脚本生成法的优点在于可以拟合大多现有的攻击模式,能够对数据集进行平衡处理,强化深度学习的训练,消除训练不充分的问题。但这种方法可能引入人为偏差,不能完美拟合现实网络流量分布。

4.2.3 混合方法

混合方法即综合利用直接收集法和脚本生成

法(如CIDDS数据集)。这种方法既能模拟出所有攻击类,又有现实网络流量背书,能够较好地符合加密恶意流量检测的需求。但该方法除需对现实网络流量数据进行标记外,还要进行数据集成,以匹配脚本生成流量和现实网络流量,因此需要大量的人力支撑。

总之,数据集的收集、标记以及特征提取是复杂的工作,如何构造公认的标记方法,捕获标记好且包含各类攻击的完美数据集仍是一个挑战。

5 数据处理

原始数据可能存在冗余、舛误以及不平衡、不匹配等问题,需要进一步处理才能应用。数据处理就是对原始收集的数据进行清洗、集成、变换以及挖掘等,使之成为符合深度学习训练、测试要求的数据集。

5.1 数据预处理

数据预处理是将未加工数据转换成适合分析的形式,包括多数据源的数据融合、数据清洗、维规约等。根据算法模型的不同,通常需要选择不同的数据预处理方法。

5.1.1 基于特征分类算法的数据预处理

目前,多数研究采用基于特征标签的深度学习分类算法进行恶意流量检测,该方法需要将原始数据处理为打上各种特征标签的数据,如统计标签、时间序列等,而后送入算法模型检测恶意流量与正常流量。基于特征分类算法的数据预处理流程如图4所示。

数据清洗是对数据进行重新审查和校验的过程,目的在于删除重复信息、纠正存在的错误。对数据中存在的无效值或错误值常采用删除的方法,包括整列删除、变量删除和成对删除等方法。对缺失值处理有均值插补、同类均值插补及高维映射等方法。

数据集成是将多个数据源收集的数据整合在一起,主要的困难在于多个数据源异构的问题,即几个数据源并不完全一致,收集的数据格式、长度等不同,相互之间存在冗余和不兼容等问题。因此,数据集成主要进行模式匹配、数据冗余处理、数据值冲突处理。

表1 恶意流量检测数据集比较
Table 1 Comparison of malicious traffic detection datasets

序号	名称	针对问题	描述	是否针对加密流量	标记	优缺点
1	DARPA 入侵检测评估数据集	入侵检测	DARPA 1999 覆盖了 Probe、DoS、R2L、U2R 和 Data 等 5 类 58 种典型攻击方式, 是目前最为全面的攻击测试数据集	否	是	应用广泛, 认可度高, 但陈旧过时
2	KDD99 数据集	入侵检测	从一个模拟的美国空军局域网采集 9 个星期的网络连接数据, 训练集中包含了 1 种正常的标识类型和 22 种攻击类型, 另外有 14 种攻击仅出现在测试数据集中	否	是	数据存在大量冗余
3	NSL-KDD 数据集	入侵检测	针对 KDD CUP 99 数据集出现的不足, NSL-KDD 数据集除去 KDD CUP 99 数据集中冗余的数据, 克服了分类器偏向于重复出现的记录, 学习方法的性能受影响等问题	否	是	缺乏入侵检测系统的公开数据, 不能完美代表现实
4	ISCX-2012 数据集	入侵检测	引入了一种系统化的方法来生成所需的数据集, 分析实际跟踪, 为 HTTP、SMTP、SSH、IMAP、POP3 和 FTP 生成实际流量的代理创建配置文件	否	是	不能完美拟合现实流量数据
5	CTU-13 数据集	恶意流量检测	CTU-13 是 2011 年在捷克共和国 CTU 大学捕获的僵尸网络流量数据集, 包含 13 个不同场景僵尸网络样本的捕获	是	是	不含特征标签
6	CAIDA 数据集	入侵检测/流量分类	包含 3 个不同的数据集, CAIDA DDoS 包含 1 h 的攻击和 5 min 的 pcap 文件	否	否	攻击类型不多, 且无标记
7	CICIDS2017 数据集	入侵检测	使用 B-Profile 系统产生良性背景流量, 基于 HTTP、HTTPS、FTP、SSH 和电子邮件协议构建了 25 个用户的抽象行为, 收集 5 天超过 50 GB 流量	可用于加密	是	正常的用户行为是通过脚本产生的
8	CIDDs 数据集	入侵检测	基于软件 Openstack, 模拟小型企业环境, 同时部署了外部服务器以获取真实流量数据, 捕获四周的时间内收集了大约 3.2×10^7 的数据流, 并分为 5 类	否	是	由 Python 脚本生成, 可能包含人为偏差
9	Kyoto 数据集	入侵检测	该数据集采用 honeypots 技术创建, 因此没有手动标记和匿名化过程	是	是	只能观察对蜜罐的攻击
10	CSIC2010 数据集	Web 流量异常检测	CSIC 2010 由西班牙国家研究委员会信息安全研究所开发, 包含 36 000 个正常请求和 25 000 多个异常请求	否	否	只针对 Web 流量, 无特征集
11	UNIBS 数据集	流量分类	布雷西亚大学校园网的边缘路由器上捕获 3 个工作日收集的 27 GB 流量, 可使用 tcptrace 提取特征	否	否	数据严重不平衡
12	ADFA IDS 数据集	入侵检测	ADFA IDS Datasets 是澳大利亚国防大学发布的一套关于 HIDS 的数据集, 包括 10 次攻击	否	否	某些攻击行为与正常行为没有很好分离
13	WIDE 流量数据	异常检测	MAWILab 是一个帮助研究人员评估他们的流量异常检测方法的数据库。它由一组在 MAWI 归档中定位流量异常的标签(采样点 B 和 F)组成。数据集每天更新, 包括即将到来的应用程序和异常的新流量	否	否	只适合做验证, 不适合做训练
14	Publicly available PCAP files	恶意流量检测	从蜜罐、沙箱或现实世界入侵捕获恶意软件流量。包含多个恶意软件数据集	否	否	数据集繁杂, 处理比较困难
15	Masquerading User Data	入侵检测	Masquerading User Data 是通过正常数据构造出来用于训练和检测 Masquerading User 攻击的数据集	否	否	用于验证比较算法
16	Honeynet 数据集	恶意流量检测	Honeynet 数据集是由 HoneyNet 组织收集的黑客攻击数据集, 能较好地反映黑客攻击模式, 数据集包括从 2000 年 4 月到 2001 年 2 月的 Snort 报警数据	否	否	只能用于验证
17	恶意流量分析数据集	恶意流量检测	各种恶意软件的原始数据	否	是	无背景流量
18	Cambridge University 计算机实验室数据	流量分类	数据采用 Weka.arff 格式。对于数据中的每个流, 有 12 个特征矢量和一个手动验证的类标签, 仅从每个流的前 5 个数据包(注意: 不是前 5 个数据包)收集	否	是	并非针对恶意流量检测
19	Moore 数据集	流量分类	用于分析 TCP、UDP、IP 的包头以及层完整性	否	是	并非针对恶意流量检测
20	SSL Certificates 数据集	流量分类	该数据集包含一组与每个研究中观察到的新 X.509 证书相关的元数据, 这些证书考虑了之前运行的所有 SSL 研究	是	否	不适合做模型训练

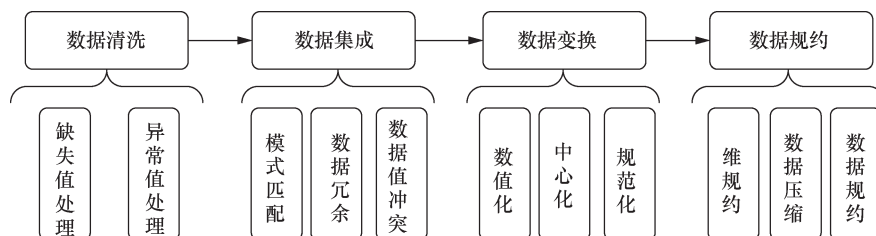


图4 基于特征分类算法的数据预处理流程

Figure 4 Data preprocessing process based on feature classification algorithm

数据变换即对数据进行规范化处理，以便于后续的信息挖掘，主要包括数值化、中心化和规范化等内容。数值化就是将非数据信息转化为数据，如网络协议信息，可用简单数值表示。中心化是指将数据减去均值或者某个指定数值的操作。规范化目的在于把数据整合到 $[0,1]$ 内以方便实验，常用最大最小值规范化方法。

数据归约指在保持数据原貌的前提下，最大限度地精简数据量。通常采取特征选择和数据采样的方法实现。

5.1.2 基于源数据分类算法的数据预处理

已有许多研究^[33-34]提出使用深度学习（如CNN、RNN等）方法自动学习数据中所隐藏的特征，从而实现加密恶意流量的检测。这类算法无须对数据进行特征提取操作，因此，数据处理相对简单，只是进行数据集成、数据切片、归一化等操作。基于源数据分类算法的数据预处理流程如图5所示。

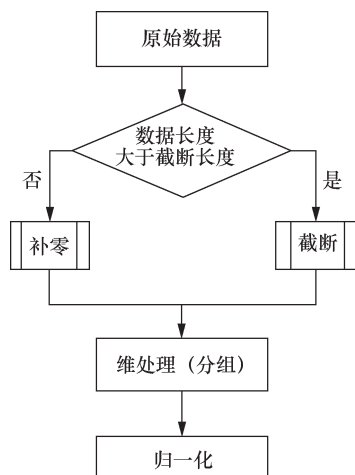


图5 基于源数据分类算法的数据预处理流程

Figure 5 Data preprocessing process based on source data classification algorithm

5.2 特征提取

数据特征提取关系检测精度，是加密恶意流量识别中的重要环节。Guyon等^[41]强调，好的特征选择可以提高分析性能，加速分析过程，并有助于理解生成数据的底层机制。学术界对数据特征的分类还没有完全统一。正如Ferreira等^[42]所讨论的，除了非常明确定义的流量类所针对的情况外，没有明确的标准方法来分析网络流量或提出一组特定特性的一般建议。

文献[43]按级别将特征分为流级、分组级、主机级和会话级4类特征。Patel等^[44]对数据收集了基础特征、内容特征、基于时间的特征和基于主机的流量特征4类的42个特征。Meghdouri等^[45]基于UNSW-NB15用随机森林方法对5个特征集的预处理性能和分类性能进行分析，得出源和流的组合效果最好。

在加密协议握手阶段，客户端与服务器往往需要通过明文协商相关加密参数，这一阶段可以得到许多与加密相关的宝贵数据，如加密采用的TLS版本、算法、证书、TLS扩展选项等。除此之外，数据包长度、流长度、报文的持续时间以及时间间隔等信息，都可以有效提升加密恶意流量检测的精度。

表2给出了本文归纳的3类数据特征。这是基于深度学习的加密恶意流量检测中应用较为广泛的一些特征。其中不包含负载特征，这主要是出于隐私保护的考虑。由于在加密条件下，负载变得不可见，负载的统计信息实际上已经包含在包或者流的统计特征之中，不再单独统计。

应当注意的是并不是特征越多越好，过量的特征选择反而可能导致检测性能的下降。另外，复杂的识别特征需要耗费大量的存储空间和计算

能力，给硬件系统带来压力。

表2 数据特征
Table 2 Data characteristics

特征类别	特征名称
基本特征	源端口、目的端口、源 IP、目的 IP、协议
	协议版本号、加密算法、证书有效性、证书自签名
	源包数、源字节数、目的包数、目的字节数
统计特征	源包长（最大值、最小值、均值、中位数、标准差）
	目的包长（最大值、最小值、均值、中位数、标准差）
	源流长度（最大值、最小值、均值、中位数、标准差）
	目的流长度（最大值、最小值、均值、中位数、标准差）
	包持续时间（最大值、最小值、均值、中位数、标准差）
时间序列特征	包到达时间（最大值、最小值、均值、中位数、标准差）
	流持续时间（最大值、最小值、均值、中位数、标准差）
	流到达时间（最大值、最小值、均值、中位数、标准差）

6 模型选择

针对加密恶意流量检测，学术界提出了各种算法模型，包括基于有效负载检测、基于熵^[36]的检测以及机器学习方法和深度学习方法。本文仅对深度学习方法进行讨论。基于深度学习的检测方法具有检测加密恶意流量的天然优势。检测模型主要分为两大类：一类是利用特征标签检测加密恶意流量；另一类是充分利用 CNN、RNN 等深度学习方法的特征学习能力，自动学习数据中隐含的特征进行恶意流量检测。

6.1 基于数据特征的深度学习检测算法

如前所述，此类方法需要首先提取数据特征，然后将有特征标签类的数据输入深度学习算法进行训练检测。

Prasse 等^[46]研究开发了基于 LSTM 的恶意软件检测模型，该模型仅使用 HTTPS 流量的握手阶段信息，能够识别网络流中的大部分恶意软件，包括以前未见过的恶意软件。

Torroledo 等^[47]专门进行特征工程，分析总结了 4 类 40 个数据特征，用于识别恶意软件和钓鱼软件签名证书。实验中采用 LSTM 以及 5 折交叉验证进行检测，结果表明具有较高的精度。Anderson 等^[19]深入分析了 18 类恶意加密软件流量的特征，并在算法模型中加入背景流量特

征，利用逻辑回归和 10 倍交叉验证进行实验，取得了 90.3% 的检测精度。

观成科技^[21]对比了线性回归、随机森林、决策树、神经网络、支持向量机、逻辑回归、卷积神经网络等算法，得出随机森林的综合效果相对较好。它们的检出率高达 99.95%，误报率可控制在 5% 以下。

基于人工打造的标签，此类算法模型的检测精度相对较高，普遍在 90% 以上，同时对算法的要求相对较低，训练比较容易。但特征提取过程复杂，且要求较高，差的特征集将对检测精度产生严重影响。另外，该类算法还可能出现过拟合的问题，泛化能力不是很强。

6.2 基于特征自学习的深度学习检测算法

为了克服数据特征提取困难的问题，学者提出了基于特征自学习的深度学习检测算法。该类算法不需要提取数据特征，只对数据进行切片操作，而后由算法自动挖掘特征进行检测。

Lotfollahi 等^[48]提出 1D-CNN 和 EAV 框架，将原始数据包长度修剪为一样，然后除以 255 将数据长度缩短，输入 EAV 进行分类，得到召回率为 98%。

Pascanu 等^[49]提出了一个基于动态分析的两层架构的恶意软件检测系统。第一层 RNN 用于学习特征；第二层是逻辑回归分类器，使用学习的特征进行分类，然而误报率较高（10%）。

Zeng 等^[50]提出一种 DFR 模型，按照 900 byte 将原始数据切片，转换为 30×30 的二维数据。而后使用 CNN、LSTM、SAE 三层结构进行流量识别。实验结果表明，该方法在 F1 分数上平均超越现有先进的方法 12.15%。

Wang 等^[26]采用基于一维 CNN 的端到端的加密流量分类方法，将流量数据截为 784 byte 的等长数据，输入 CNN 进行分类。实验证明采用 1D-CNN 效果好于 2D-CNN。

基于特征自学习的深度学习检测算法最大的特点是不需要再提取数据特征，对于数据集的要求低、依赖性小，可移植性好，使用更加灵活。但这类方法也有着天然的不足：① 数据切片取多少字节合适并没有依据，只是凭经验取舍；② 切片数据长度不能太长，切片过长可能影响计算性

能,进而影响训练的效率;③ 训练过程可解释性相对较低,训练比较困难,容易引起检测性能不稳定。

表3给出了两类检测算法模型的一个简单对比,其中I类模型代表基于数据特征的深度学习检测算法,II类模型代表基于特征自学习的深度学习检测算法。

从表3可以看出,两类模型各有优劣。实际应用中,主要关心精确度和泛化能力两项指标。随着研究的深入,II类模型也能达到比较高的精度,如Wang等^[33]在CUT数据集上取得了99.41%的精度。加之II类模型不需要提取特征,省却了大量烦琐的工作。从这些看来,II类模型较I类模型优。即便如此,应用过程中也不能简单武断地判定哪个模型更好,应当采用多种算法模型进行实验,对比实验结果,通过实际检测的数据结果具体地评定模型的优劣。

另外,Rezai等^[51]提出了一种半监督检测模型,首先在大型未标记数据集上训练,然后使用少量标记数据集重新训练模型。模型采用1D-CNN进行分类,取得较高的精度,证明了一个数据集上进行预训练,在另外一个数据集上进行再训练,可以精确分类。

需要看到这种方法除需要标记数据外,还要考虑数据集成的问题,即原始数据集与标记数据集之间匹配的问题,不同数据集特征不尽相同,可能导致模型精度降低,正如文中提到在ISCX数据集上精度不高。所以,模型的不可移植影响了此类方法在实际检测中的应用。

7 训练和评估

模型构建以后,即需要对模型进行训练和评估,训练通常采用N折交叉验证,一般为10折,不再赘述。

要真正评估一个加密恶意流量检测模型的好坏,应当将模型应用在实际的网络环境中,观察

模型的检测性能。但由于芯片研发、安全风险等代价太大,通常利用测试集上模型的性能表现来评估检测模型的优劣。评价指标主要有精确性指标和实时性指标。

7.1 精确性指标

精确性指标基于混淆矩阵生成,主要用来评估模型的检测能力和检测精度,包括准确率、召回率、精确度、误报率、F分数等。

混淆矩阵如表4所示,用于描述加密恶意流量检测中实际类别和预测类别之间的相互关系。其中,TP(真正例)表示检测模型将攻击类型正确识别为攻击类型的样本个数;TN(真负例)表示检测模型将正常类型正确识别为正常类型的样本个数;FP(假正例)表示检测模型将正常类型错误识别为攻击类型的样本个数;FN(假负例)表示检测模型将攻击类型错误识别为正常类型的样本个数。

表4 混淆矩阵
Table 4 Confusion matrix

预测	攻击	正常
攻击	TP	FN
正常	FP	TN

基于上述定义,给出各种精度性能指标的公式化定义。

$$\text{准确率} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{召回率} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{精确度} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{误报率} = \frac{FP}{FP + TN} \quad (4)$$

召回率、精确度和误报率体现了检测方法在每个攻击类别上的检测效果,特别是当数据不平衡时,这3项指标能够准确获知各类攻击的检测情况。

表3 两类检测算法模型对比
Table 3 Comparison of two types of detection algorithm models

算法模型	提取特征	精确度	训练难易	泛化能力	算法要求
I类模型	是	高	易	相对较弱	低
II类模型	否	较高	难	较强	高

由于一般情况下召回率和精确度之间存在博弈, 即召回率高, 精确度就低; 反之精确度高, 召回率就低。于是引入 F 分数 (F-score) 指标, F 分数是综合考虑召回率和精确度的精确性评价指标, 其定义如式(5)所示。当 $\beta = 1$ 时, 精确率和召回率同等重要, 权重相同, 此时 F 分数又被称为 F1 分数 (F1-score) 或 F 衡量 (F-measure)。F 分数值越高表明算法的分类性能越好。

$$F\text{分数} = (1 + \beta^2) \frac{\text{Precision} \times \text{Recall}}{\beta^2 \text{Precision} + \text{Recall}} \quad (5)$$

上述精确性指标通常基于流或基于包, Erman 等^[52]指出, 字节精度对评估流量分类算法的准确性至关重要。他们在跟踪一个为期 6 个月的数据示例中, 发现 1% 的流量占字节流量的 73%; 在阈值为 3.7 MB 的前提下, 流量的前 0.1% 占字节量的 46%。因此, 正确识别后 99.9% 的流量仍然导致数据集中 46% 的字节被错误分类。同时, 字节精度可能耗费计算资源, 提升计算复杂度。

7.2 实时性指标

实时性指标用于评估模型的检测效率, 反映加密恶意检测算法可以在线、快速地识别加密恶意流量的能力, 保证在实施恶意流量检测的过程中不影响核心网络性能。实时性主要体现在对流的前 N 个包的精确检测上。

由于深度学习方法训练数据量巨大, 常常耗费大量计算性能和时间, 因此, 关于各类深度学习算法加速的研究也正在如火如荼地展开, 并取得了不少成果, 这部分不在本文讨论之列。

关于“六步法”的最后一步应用检验改进, 是指将构建的模型应用到实际网络中, 进行实网加密恶意流量检测, 通过网络运行, 检验算法模型的实效性和健壮性, 并定期对模型进行更新, 不断完善模型以取得更高的检测精度和效能。这部分内容主要涉及系统实现, 不再深入讨论。

8 结束语

随着加密流量日益增加, 攻击者越来越多地利用加密手段掩盖攻击意图, 给检测带来相当的困难。因此, 加密恶意流量检测的重要性日益凸显。本文将加密恶意流量检测归纳为“六步法”, 并按照流程对检测的各个环节逐一进行梳理分

析。总体上讲, 加密恶意流量检测虽然已经取得了一些成果, 但还存在以下问题, 也是下一步研究的方向。

(1) 数据集问题

获得一个优良的数据集等于检测成功了 90%。如上文所述, 数据集还存在不全面、不平衡、不贴近现实等问题。公认的标记方法、生成完美代表现实的数据以及利用深度学习等方法进行自动特征挖掘等方面都需要进一步深入研究。

(2) 加密恶意流量, 虽然有多分类, 但也仅限于区分少数几类攻击, 如 DDoS、僵尸网络等, 如何将攻击精准定位, 指导防御者快速反应, 采取有效应对措施还是一个挑战。

(3) 对抗深度学习问题

已有研究^[53]指出利用 GAN 网络, 基于攻击流量生成正常流量, 同时保持数据中的攻击性。生成的数据包成功绕过了检测算法, 对系统实施了攻击。如何对对抗深度学习进行反对抗也是一个值得研究的问题。

(4) 应用问题

学术研究多存在于算法层面, 且多利用静态数据集 (不论是公开的还是私有采集的) 进行检测。但现实网络中有更多的不确定性, 将基于深度学习的恶意加密流量检测移植到现实网络上, 真正让攻击无所遁形还有很长的路要走。

参考文献:

- [1] LV Y F. In 2018, lucky star network security report and trend outlook[J]. Information Security Research, 2019, 5(3): 186-191.
- [2] CNCERT 2018 China Internet network security situation report[R]. 2018.
- [3] GURUBARAN S. Cisco ETA-provides solution for detecting malware in encrypted traffic[EB].
- [4] LET'S ENCRYPT. Linux foundation collaborative projects[EB].
- [5] HE G, XU B, ZHANG L, et al. Mobile app identification for encrypted network flows by traffic correlation[J]. International Journal of Distributed Sensor Networks, 2018, 14(12): 1-17.
- [6] ARNA M. Malware is moving heavily to HTTPS [EB].
- [7] BUJLOW T, CARELA-ESPAÑOL V, BARLET-ROS P. Independent comparison of popular DPI tools for traffic classification[J]. Computer Networks, 2015, 76: 75-89.
- [8] WRIGHT C V, COULL S E, MONROSE F. Traffic morphing: an efficient defense against statistical traffic analysis[C]//NDSS. 2009: 237-250.
- [9] VELAN P, ČERMÁK M, ČELEDA P, et al. A survey of methods

- for encrypted traffic classification and analysis[J]. *International Journal of Network Management*, 2015, 25(5): 355-374.
- [10] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks[J]. *Advances in neural information processing systems*, 2012, 25(2):84-90.
 - [11] CHO K, VAN -MERRIENBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. *arXiv Preprint arXiv:1406.1078*, 2014.
 - [12] GRAVES A, MOHAMED A, HINTON G. Speech recognition with deep recurrent neural networks[C]//2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013: 6645-6649.
 - [13] BYEONW, BREUEL T M, RAUE F, et al. Scene labeling with LSTM recurrent neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3547-3555.
 - [14] VINAYAKUMAR R, SOMAN K P, POORNACHANDRAN P. Applying convolutional neural network for network intrusion detection[C]//IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI). 2017: 1222-1228.
 - [15] WANG W, SHENG Y Q, WANGJL, et al. HAST-IDS: learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection[J]. *IEEE Access*, 2018, 6: 1792-1806.
 - [16] YUAN X Y, LI C H, LI X L. Deep defense: identifying DDoS attack via deep learning[C]//2017 IEEE International Conference on Smart Computing. 2017:18.
 - [17] LI Z P, QIN Z, HUANG K, et al. Intrusion detection using convolutional neural networks for representation learning[M]//*Neural Information Processing*. 2017: 858-866.
 - [18] YIN C L, ZHU Y F, FEI J L, et al. A deep learning approach for intrusion detection using recurrent neural networks[J]. *IEEE Access*, 2017, 5: 21954-21961.
 - [19] ANDERSON B, PAUL S, MCGREW D. Deciphering malware's use of TLS (without decryption)[J]. *Comput. Virol. Hacking Techn.* 2018, 14 (3): 195-211.
 - [20] BLAKE A, DAVID M. Identifying encrypted malware traffic with contextual flow data[C]// *ACM Workshop on Artificial Intelligence and Security (AISec)*. 2016: 35-46.
 - [21] GUANCHENG TECHNOLOGY. A report on the country's first detection engine for encrypted traffic[EB].
 - [22] HE G, XU B, ZHU H. AppFA: a novel approach to detect malicious Android applications on the network[J].*Security and Communication Networks*, 2018 :1-15.
 - [23] SHABTAI A, TENENBOIM C L, MIMRAN D, et al. Mobile malware detection through analysis of deviations in application network behavior[J]. *Computers & Security*, 2014, 43:1-18.
 - [24] GARG S, PEDDOJU S K, SARJE A K. Network-based detection of Android malicious Apps[J]. *International Journal of Information Security*, 2017, 16(4):385-400.
 - [25] REZAEI S, LIU X. Deep learning for encrypted traffic classification: an overview[J]. *IEEE Commun Mag*, 2019, 57(5):76-81.
 - [26] WANG W, ZHU M, WANG J, et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks[C]//*IEEE Int Conf Intell Security Informat*. 2017:43-48.
 - [27] OKADA Y, ATA S, NAKAMURA N, et al. Comparisons of machine learning algorithms for application identification of encrypted traffic[C]//10th IEEE International Conference on Machine Learning and Applications and Workshops. 2011: 358-361.
 - [28] ZHAO B, GUO H, LIU Q R, et al. Protocol independent identification of encrypted traffic based on weighted cumulative sum test[J]. *Journal of Software*, 2013, 24(6): 1334-1345.
 - [29] NIU W N, ZHUO Z L, ZHANG X S, et al. A heuristic statistical testing based approach for encrypted network traffic identification[J]. *IEEE Trans. Veh. Technol.*, 2019, 68 (4).
 - [30] JAIN A V. Network traffic identification with convolutional neural networks[C]//2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, Washington: IEEE Computer Society (DASC). 2018:1001-1007.
 - [31] KARAGIANNIS T, PAPAGIANNAKI K, FALOUTSOS M. BLINK: multilevel traffic classification in the dark[C]//*SIGCOMM'05*. 2005: 229-239.
 - [32] GIL G D, LASHKARI A H, MAMUN M, et al. Characterization of encrypted and VPN traffic using time-related features[C]//2nd International Conference on Information Systems Security and Privacy (ICISSP 2016). 2016: 19-21.
 - [33] WANG W, ZENG X, YE X, et al. Malware traffic classification using convolutional neural network for representation learning[C]// *The 31st International Conference on Information Networking(ICOIN)*. 2017: 712-717.
 - [34] LOTFOLLAHI M, et al , Deep packet: a novel approach for encrypted traffic classification using deep learning[J]. *arXiv: 1709.02656*, 2017.
 - [35] YUAN X, LI C, LI X. Deepdefense: identifying DDoS attack via deep learning[C]//2017 IEEE International Conference Smart Computing (SMARTCOMP). 2017: 1-8.
 - [36] NYCHIS G, SEKAR V, ANDERSEN D, et al. An empirical evaluation of entropy-based traffic anomaly detection[J]. *The 8th ACM Computing Surveys*, 2007, 39(1): 1-42.
 - [37] KOCH R, GOLLING M, RODOSEK G D. Behavior based intrusion detection in encrypted environments[J]. *Communications Magazine*, 2014, 52(7): 124-131.
 - [38] IVAN T, LUIS D C, ALEJANDRO C B. Hunting malicious TLS certificates with deep neural networks[C]//*The 11th ACM Workshop on Artificial Intelligence and Security*. 2018: 64-73.
 - [39] FRANTISEK S, SEBASTIAN G. Machine learning for network HTTPS analysis[C]//*BRUCON Conference*. 2017:5.
 - [40] GHARIB A, SHARAFALDIN I, HABIBI L A, et al. An evaluation framework for intrusion detection dataset[C]//2016 International Conference on Information Science and Security (ICISS). 2016:1-6.
 - [41] GUYON I, ELISSEEFF A. An introduction to variable and feature selection.[J]. *Mach Learn Res* 2003(3): 1157-1182.
 - [42] FERREIRA D C, VAZQUEZ F I, VORMAYR G, et al. A meta-analysis approach for feature selection in network traffic research[C]//*Reproducibility Workshop (Reproducibility'17)*. 2017: 17-20.
 - [43] PAN W B, CHENG G, GUO X J, et al. Review and perspective on encrypted traffic identification research[J]. *Journal on Communications*, 2016, 37(9):154-167.
 - [44] PATEL S, SONDHI J. A review of intrusion detection technique

- using various technique of machine learning and feature optimization technique[J]. International Journal of Computer Applications. 2014, 93(14): 43-47.
- [45] MEGHDOURI F, ZSEBY T, IGLESIAS F. Analysis of lightweight feature vectors for attack detection in network traffic[J]. Applied Sciences. 2018, 8(11): 219
- [46] PRASSE P, MACHLICA L, PEVNY T, et al. Malware detection by analysing encrypted network traffic with neural networks[C]// European Conference on Machine Learning and Knowledge Discovery in Databases. 2017: 73-88.
- [47] TORROLEDO I, CAMACHO L D, BAHNSEN A C. Hunting malicious TLS certificates with deep neural networks[C]//11th ACM Workshop Artif. Intell. Secur. 2018: 64-73.
- [48] LOTFOLLAHI M, ZADE R S H, SIAVOSHANI M J, et al. Deep packet: a novel approach for encrypted traffic classification using deep learning[J]. eprint arXiv: 1709.02656. 2017.
- [49] PASCANU R, STOKES J W, SANOSSIAN H, et al. Malware classification with recurrent networks[C]//40th Int Conf on Acoustics, Speech and Signal Processing. 2015: 1916-1920.
- [50] ZENG Y, GU H, WEI W, et al. Deep-full-range: a deep learning based network encrypted traffic classification and intrusion detection framework[J]. IEEE Access, 2019: 1.
- [51] REZAEI S, LIU X. How to achieve high classification accuracy with just a few labels: a semi-supervised approach using sampled packets[J]. arXiv preprint arXiv:1812.09761, 2018.
- [52] ERMAN J, MAHANTI A, ARLITT M. Byte me: a case for byte accuracy in traffic classification[C]//3rd Annual ACM Workshop on Mining Network Data(MineNet '07). 2007: 35-38.
- [53] HE Y. Research on confrontational learning for industrial control

system intrusion detection[D]. Zhejiang: Zhejiang University, 2019.

[作者简介]



翟明芳（1986—），男，河南巩义人，国家数字交换系统工程技术研究中心硕士生，主要研究方向为网络空间安全及深度学习算法。



张兴明（1963—），男，河南新乡人，博士，国家数字交换系统工程技术研究中心教授，主要研究方向为网络空间安全、高性能计算、软硬件协同设计和深度学习。



赵博（1981—），男，吉林公主岭人，博士，国家数字交换系统工程技术研究中心副研究员，主要研究方向为网络空间安全和拟态防御体系结构。