计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目： TLS 协议恶意加密流量识别研究综述
作者： 康鹏，杨文忠，马红桥
网络首发日期： 2022-03-09
引用格式： 康鹏，杨文忠，马红桥．TLS 协议恶意加密流量识别研究综述[J/OL]．计算机工程与应用. https://kns.cnki.net/kcms/detail/11.2127.TP.20220308.0853.002.html

网络首发时间：2022-03-09 13:17:32
网络首发地址：https://kns.cnki.net/kcms/detail/11.2127.TP.20220308.0853.002.html

TLS 协议恶意加密流量识别研究综述

康鹏[1]，杨文忠[1,2]，马红桥[1]

1. 新疆大学 信息科学与工程学院，乌鲁木齐 830046
2. 新疆大学 信息科学与工程学院新疆维吾尔自治区多语种信息技术重点实验室，乌鲁木齐 830046

摘　要：随着 5 G 时代的来临，以及公众对互联网的认识日益加深，公众对个人隐私的保护也越来越重视。由于数据加密过程中存在着恶意通信，为确保数据安全，维护社会国家利益，加密流量识别的研究工作尤为重要。为此，本文针对 TLS 流量详细的阐述并分析了早期识别方法的改进技术，包括常见的流量检测技术，DPI 检测技术，代理技术以及证书检测技术。还介绍了选取不同 TLS 加密流量特征的机器学习模型，以及无需特征选择的深度学习模型等诸多最新研究成果。对相关研究工作的不足进行总结，并对未来技术的研究工作和发展趋势进行了展望。

关键词：5G 时代；个人隐私；恶意流量；数据安全；TLS 加密流量识别

文献标志码:A　　中图分类号:TP309.7　　doi：10.3778/j.issn.1002-8331.2110-0029

TLS Malicious Encrypted Traffic Identification Research

KANG Peng[1], YANG Wenzhong[1,2], MA Hongqiao[1]

1. College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
2. College of Information Science and Engineering, Key Laboratory of Multilingual Information Technology in Xinjiang Uygur Autonomous Region, Xinjiang University, Urumqi 830046, China

Abstract: With the advent of the 5G era and the increasing public awareness of the Internet, the public has paid more and more attention to the protection of personal privacy. Due to malicious communication in the process of data encryption, to ensure data security and safeguard social and national interests, the research work on encrypted traffic identification is particularly important. Therefore, this paper describes the TLS traffic in detail and analyses the improved technology of early identification method, including common traffic detection technology, DPI detection technology, proxy technology, and certificate detection technology. It also introduces machine learning models for selecting different TLS encrypted traffic characteristics, as well as many recent research results of deep learning models without feature selection. The deficiencies of the related research work are summarized, and the future research work and development trend of the technology have been prospected.

Key words: 5G era; personal privacy; malicious traffic; data security; TLS encrypted traffic identification

基金项目：国家自然科学基金项目（U1603115）；新疆维吾尔自治区重点科技专项（2020A02001-1）；国家重点研发计划项目子课题(2017YFC0820702-3)；中国电子科学研究院，社会安全风险感知与防控大数据应用国家工程实验室主任基金项目：面向公共安全的多语种网络舆情监测系统。

作者简介：康鹏(1996–),男,硕士研究生,研究方向为网络安全、区块链安全;杨文忠(1971–),通讯作者,男,博士,教授,CCF 会员,研究方向为图像处理、网络舆情、情报分析、信息安全、无线传感器,E-mail:ywz_xy@163.com;马红桥（1997-）,男,硕士研究生,研究方向为网络安全，区块链安全。

最新互联网研究趋势报告指出[1]，大约有 87%的 Web 流量是加密的，在 2020 年更是有超过 70%的恶意活动通过加密来传输恶意软件，攻击者也通过加密绕过未授权活动的检测，来隐藏恶意程序和服务器的交互。目前大多数网络应用程序和服务只支持由传输层安全(Transport Layer Security, TLS)封装的加密通信[2]。这样会导致两个方面的问题。一方面由于保密数据的价值性，使得降级攻击、Lucky Thirteen 攻击等针对协议本身的攻击递增；另一方面恶意流量也通过各种手段达到加密传输。因此加密恶意流量的识别工作就显得更加重要。为此，本文从网络协议发展现状出发，分析了 TLS 加密恶意流量的发展历程及各个阶段存在的问题；将常见的流量解密技术和 TLS 流量解密技术进行了总结，分析了各自的方法特征，优缺点以及数据集的选择。同时针对以上现存技术存在的问题和此领域未来的技术发展做了合理的展望。

1 TLS 加密协议现状分析

网络加密也称为网络层或者网络级加密，是在网络传输层选用加密服务。网络加密只在传输中加密，对于终端用户是透明的。常见网络加密协议有网络认证协议 Kerberos；安全外壳协议 SSH；安全电子交易协议 SET；网络层安全协议 IPsec 以及我们主要研究的传输层安全协议 TLS。

SSL 3.0 协议是由 P. Kocher，P. Karlton 和 A. Freier 一起设计实现的，我们将 TLS 的发展与特点总结为表 1 和表 2。协议发展历程如表 1 所示。TLS1.0 协议和 SSL 协议的主要不同点如表 2 所示。从表中可以清楚知道，相较于 TLS1.0，TLS1.1 以及 TLS1.2 来说，由于 TLS1.1 对于 TLS1.0 来说被视为一个微量升级，都存在类似于降级攻击的大量缺陷，所以大部分软件都将它弃用，而直接使用 TLS1.2 版本。这里为了让读者更好得了解 TLS 协议的发展历程，我们依旧将 TLS1.0，TLS1.1 以及 TLS1.2 做出对比分析，分析结果如表 3 所示。

TLS1.2 协议主要分为两层，底层是 TLS 记录协议，主要负责使用对称密码对消息进行加密；上层握手协议负责在客户端和服务器端商定密码算法和共享密钥；密码规格变更协议负责向通信对象传达变更密码方式的信号；警告协议负责在发生错误时将错误传达给对方；应用数据协议负责将 TLS 承载的应用数据传达给通信对象。由于 SSL2.0,SSL3.0,TLS1.0,TLS1.1 分别于 2011 年，2015 年和 2020 年弃用，同时文章主要针对 TLS 协议进行研究，SSL 协议和久远协议版本不再进行过多赘述。

表 1　TLS 协议发展总结

Table 1　Summary of TLS Agreement Development

协议	年份	RFC	描述
SSL1.0	1994		NetScape 公司设计 1.0 版，未发布
SSL2.0	1995		NetScape 公司发布 SSL 2.0 版
SSL3.0	1996	RFC 6101	NetScape 公司发布 SSL 3.0 版
TLS1.0	1999	RFC 2246[3]	IETF 将 SSL 标准化改名为 TLS 发布 1.0 版
TLS1.1	2006	RFC 4346	发布 TLS1.1 版
TLS1.2	2008	RFC 5246	发布 TLS1.2 版
TLS1.3	2018	RFC 8446[4]	发布 TLS1.3 版

表 2　TLS1.0 与 SSL 对比分析

Table 2　Comparative Analysis of TLS1.0 and SSL

差异项	SSL	TLS1.0
报文鉴别码	填充字节与密钥间采用连接计算	采用 HMAC 算法的异或运算
伪随机函数	未将秘钥拓展为数据块	使用 PRF 伪随机函数
报警代码	TLS 在继承 SSL3.0 之上加入了如解密失败，记录溢出等	
密文族和客户证书	TLS 不支持 Fortezza 秘钥交换、加密算法和客户证书	
填充	填充后数据长度达到密文块长度的最小整数倍	填充后数据长度达到密文块长度的任意整数倍

表 3　TLS 三个版本对比分析

Table 3　Comparative Analysis of The Three Versions of TLS

差异项	TLS1.0	TLS1.1	TLS1.2
对 Finish 报文影响	计算 finish 时使用 MD5+SHA1 组合运算		单次 SHA256 运算
对 PRF 算法影响	两次 P_HASH，第一次使用 MD5 和 Secert 的前半部，第二次使用 SHA1 和 Secret 的后半部		单次 P_HASH，并使用 SHA256 或 SHA384
对 Certificate verify 影响	使用 MD5+SHA1 的形式对握手信息进行摘要运算		Certificate Verify 报文格式不同，多出 HASH_ALG 和 SIGN_ALG 两个字节，再根据 HASH_ALG 单次计算具体握手摘要
对 Server key exchange 影响	报文格式延续之前没有变化		报文格式多了 2 个字节表示 HASH 算法和签名算法
对加密影响	在加密数据前填充 IV_SIZE 长度的随机数并作为数据一起加解密，最后解密后丢弃	CBC 加密使用包含在每个 TLS 记录中的显式Ⅳ，不对这个 IV_SIZE 进行加密，在下一数据块的 WRITE_IV 改成这个随机数	

2 TLS1.3 问题与归纳

TLS1.3 协议在通信过程中，服务端 Hello 报文之后的所有信息都做了加密处理。由于 TLS1.3 相较之前的协议有较大的差异[5][6]，表 4 将 TLS1.3 与其他版本做了综合对比分析。文献[7]提出构建最全面、最可靠、最模块化的 TLS 1.3 draft 21 候选版本的符号模型来推动 1.3 版本发展，但发展还是受到了阻碍，表 5 列出了部分 TLS1.3 存在的问题,并进行了分析。

表 4　其他版本与 TLS1.3 对比分析

Table 4　Comparing other versions with TLS1.3

参数	其他版本协议	TLS1.3
使用程度	其他版本的弃用情况在之前章节已有总结	TLSv1.3 正在被大力推广和使用
加密性算法	存在无法确保前向安全的静态 MD5，SHA-1 等算法	全面使用 ECC 密码算法，删除了一些加密性能较弱的哈希算法[5]和加密组件使安全性得到了提升
组成协议	存在更改密码规范协议	删除更改密码规范协议；增加 0-RTT 握手协议[6]
握手过程	3 次握手	2 次握手，因此具有更快的访问速度，也促使更多的握手过程被加密
证书支持	支持 DSA 证书	不再支持 DSA 证书
重新协商	支持通过重新协商回退到更早的不安全版本	不再支持重新协商机制

表 5　TLS1.3 存在部分问题总结

Table 5　Summary of some problems with TLS1.3

文献	问题发现	结果
[8]	发现针对密钥交换协议的新型攻击	禁用 MD5 和 SHA-1 等弱散列函数
[9]	易受旁路攻击的 RSA 默认的填充方式	不支持 PKCS#1v1.5，但仍然无法避开攻击
[10]	现有防护措施针对特定攻击，有时甚至相互矛盾	对记录协议进行分析并提出一般性防护对策
[11]	使用重放攻击对 0-RTT 握手进行分析	阐明重放攻击下 0-RTT 安全性限制
[12]	进行服务器欺骗等降级攻击	发现可重置握手计时器的漏洞
[13][14][15][16][17][18]	版本发展可追溯	导致实际攻击累积[19][20][21][22][23][24][25][26][27]
[28]	协议设计	设计者依据公开论文指出协议漏洞从而改变协议实现细节，整体改善不明显
[29]	协议实现	TLS1.3 实现工具单一，版本间难实现互操作
[30]	多个实体拥有相同 PSK，在对应模式会出现漏洞	提出缓解措施防止这种脆弱性

3 TLS 协议加密网络流量识别技术

由于以上协议本身存在的安全漏洞和恶意加密流量的指数增长,保证数据和网络空间的安全刻不容缓。在网络加密流量占比不多,同时 TLS 协议未被提出时,常见流量识别技术基本可以保证网络安全;TLS 被提出后流量加密技术逐渐成熟,目前存在的检测技术问题如表 6 总结。

表 6　现存检测技术问题总结

Table 6　Summary of existing detection technology problems

检测技术	存在缺陷
常见流量识别技术	无法有效检测恶意加密数据并造成巨大的网络开销
DPI	仍然无法完全解决上述问题并进行信息解密,带来隐私泄露的问题
证书检测,JA3/JA3S,代理技术	解决了部分问题也带来新问题,如通过代理绕过筛选器,从而访问加密恶意网站等
机器学习	由于近几年恶意软件变种;公开数据集不完善等问题的存在,随时间迁移使机器学习识别准确率和精度下降以及过度依赖专家经验等问题

为了解决上述问题,基于深度学习的检测方法在 2020 年逐渐进入研究正轨。为此本章节针对加密流量检测以及特定加密协议 TLS 的流量检测技术展开了探索并做出归纳。图 1 显示了近四年的研究成果数量和技术壁垒,在 2019 年(2018 年 TLS1.3 版本正式发布)达到了研究顶峰后,机器学习已到达不错的精度和准确率,此方法研究上限逐渐饱和。

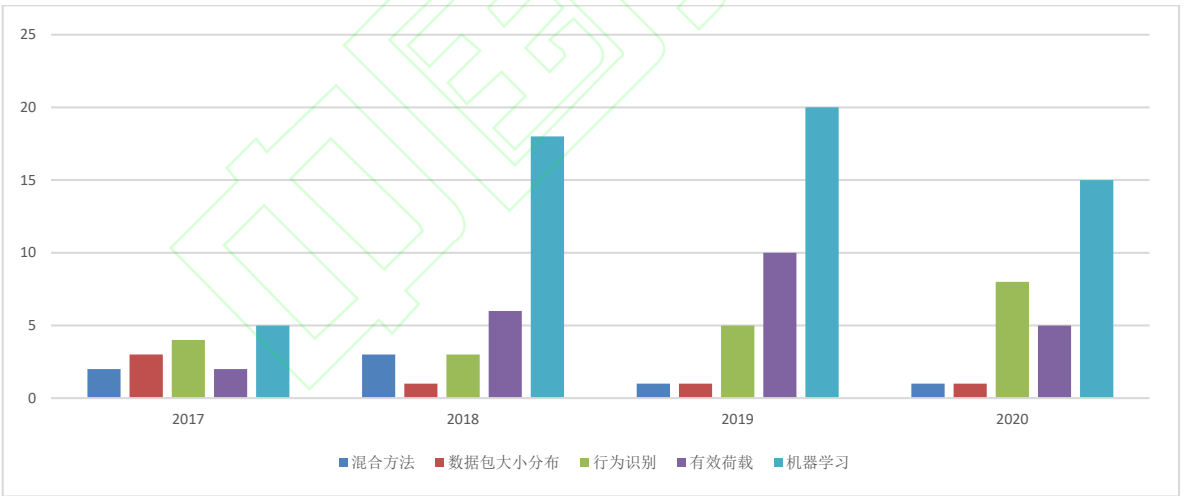

图 1　近四年流量识别研究

Fig.1　Research on Traffic Identification in Recent Four Years

3.1 早期流量识别技术

文献[31]根据常见加密流量分类识别方法的不同进行了分类,随着 TLS 协议普及也存在少量文献将这些技术融入进行研究,如表 7 所示,但这些流量识别技术普遍存在网络开销大;网络延迟高;数据存储不安全;数据信息易泄露;检测效率低;对 TLS 加密信息识别有限等问题。因此,为了解决常见流量识别方法应用到 TLS 流量检测时存在的以上弊端,出现了新的检测技术,如基于传统 IP 数据包检测的改良技术 DPI 检测,证书检测,代理检测等。本节主要针对近几年 TLS 加密流量的检测方法进行了介绍。

表 7　常见流量识别方法总结分析

Table 7　Summary and analysis of common flow identification methods

识别方式	定义	方法利用	问题总结
有效荷载识别	从未加密部分检测出少量信息后结合统计方法对应用或服务进行识别	文献[32]使用上下文流量中的有效负载（DNS 响应和 HTTP 头）帮助识别加密流量中的威胁	数据激增；系统处理困难；存储开销庞大；涉及用户隐私和数据安全
端口映射识别	检查数据包源目的端口号，根据网络协议或应用在通信时使用的端口号规则并与之映射，进而识别不同应用	针对不同唯一端口映射一一对应的网络应用	网络端口号与单个应用无法总端到端实时对应导致识别准确率和可靠性不断下降
行为特征识别	通过考虑主机的行为(如 P2P)或用户行为来划分不同的应用	文献[33]提出了基于主机行为,粗粒度对 P2P 流量实时识别	从宏观角度分析过于模糊，系统时空开销较大，识别实时性较差
数据包大小分布识别	通过分析实际网络环境中的流量，不同业务类型产生的数据包大小分布不同	文献[34]提出减少数据包处理量的同时实现对 P2P；VoIP 应用准确识别	识别准确性有欠缺，且对加密数据的识别甚微

3.1.1 DPI 解密技术

DPI（深度包检测技术）是在传统 IP 数据包检测技术之上增加了对应用层数据的应用协议识别，数据包内容检测与深度解码的功能。文献[35]将 DPI 技术分类为：基于特征字的识别技术，应用层网关识别技术和行为模式识别技术。

为了缓解解密 TLS 数据消耗大量服务器性能的问题，较为流行的方法是安装加速卡，但此方法在服务器主机之上处理数据，没有完全消除系统负荷；系统兼容性不佳、对主机的依赖过大，无法满足大型应用的需求；其他网络设备无法复用解密后的明文流量，需要重复解密，造成资源的严重浪费。

应用 DPI 技术识别恶意加密流量弥补了加速卡的不足。但 DPI 技术依旧存在解析加密流量不完全，网络性能延迟，设备迭代困难，可视化不足等问题。文献[36]和文献[37]提出对加密流量进行深度包检测(DPI)而无需解密的技术，但在设置阶段需要大量的计算和较长的检测时间。文献[38]提出基于 DPI 检查负载随机性的加密流量识别算法，但涉及的流量仅仅包含 TCP,SFTP,HTTP,SMTP 和 SSL 协议。为解决 DPI 方法存在的缺陷，表 8 对比分析了部分方法。

文献[1]从隐私的角度分析了网络中 TLS 截获对用户的影响,为了渗透到加密连接中经常使用解密技术，如何保证解密数据的安全性，合法性，有效准确性仍是不可忽略的重点。

表 8　部分解密方法总结

Table 8　Summary of Partial Decryption Methods

文献	方法	问题
[39]	对服务器本身执行检查	增加网络负担，造成网络延迟
[40]	连接 Fiddler 等工具由 IDS 自动解密	造成加密信息泄露，隐私无法得到保障
[41]	生成会话密钥文件与拦截的 TLS 流量一起使用	假设内部人员可访问计算机或网络并设置访问书签
[42]	提出 IA2-TLS(随时随地检查 TLS)方法	解决了网络负担，中间代理证书管理等问题，但无法保证秘钥保密性。

3.1.2 Proxy 技术

文献[43]显示了一份用户对 TLS 流量使用 Proxy 技术进行保护检测的调查如图 2。超过 65%的用户同意接受代理的使用，但同时也要求浏览器能通知代理的相关信息，并建议有解决代理的法律存在。可事实上浏览器既不能保障代理的安全合法性，用户也无法有效识别安全代理。可将 Proxy 技术可分类为：正向代理(Forward Proxy)，反向代理（Reverse Proxy）和公开代理（Open Proxy）技术。

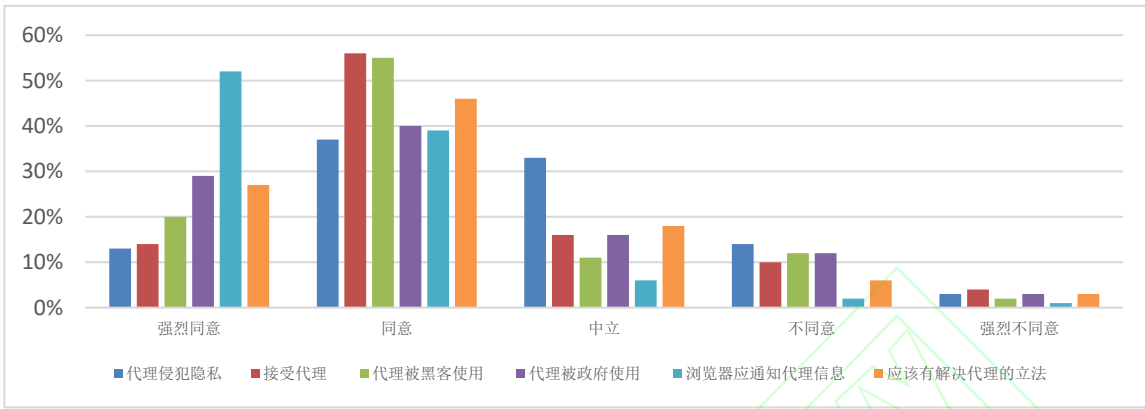

图 2　TLS 流量使用代理意愿

Fig.2　TLS Traffic Use Proxy Intention

图 3 描述了代理技术的原理。代理可以解密、监视或修改所有用户流量，通过加密通道将请求传递到所需网站。浏览器和用户没有区分善意和恶意 TLS 代理的能力，用户甚至完全不知道一个组织或攻击者正在拦截加密流量。即使存在 TLS 代理，浏览器也会显示一个令人安心地锁定图标，这可能会误导用户认为正在安全的与网站进行通讯。目前的研究主要采用三种方法在加密流量上启用代理功能：MITM 方法解密或修改 TLS 流量[44][45][46][47]；握手期间显式地包含中间件[48]；允许直接检查加密流量[49]。现存有关代理的研究已经很多，部分总结如表 9 所示。

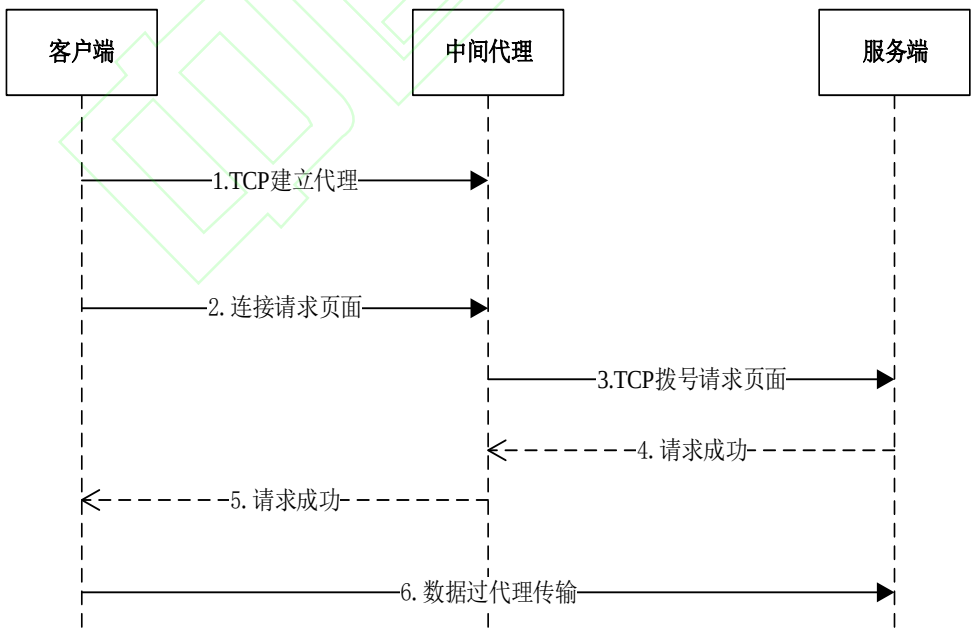

图 3　代理技术原理

Fig.3　Principle of Agent Technology

表 9　代理技术部分总结

Table 9　Agent Technology Part Summary

文献	研究	分析
[50]	客户端到服务器是否存在透明 HTTP 代理	代理存在客户端和服务端之间的网络路径上的任意地方
[51]	中间人拦截 HTTPS 的程度	拦截会显著降低安全性，并将客户端暴露于攻击中
[52]	关注报头操作和 HTTPS 拦截	揭露了开放代理所表现出的多种形式的可疑行为
[53]	安全模型下的 TLS 流量解密技术	利用 SGX 技术实现了加密流量的可见性
[54]	网络监控和 TLS 指纹下的轻量级加密客户端实时识别	分析 TLS 握手，估计出通信中客户端用户代理，以此建立密码套件列表和用户代理字典进而对加密流量分类
[55]	拦截代理使用过程的变化	使恶意软件改变客户端行为
[56]	开放代理	修改代理页面注入广告，收集用户信息并重定向到恶意软件页面

目前代理技术存在的问题总结为：

（1）在实时解密和重新加密流量中，造成计算和通信方面的性能下降。

（2）绕过审查，并充当各种攻击的垫脚石访问无法访问的（如：黄赌毒）网站。

（3）为运营商提供了一个扩展的网络流量视图来窃听通信，执行中间人攻击盈利。

（4）连同 DPI 加解密技术，造成用户隐私泄露和篡改，用户不信任和不允许此类技术。

3.2 JA3 技术

2015 年提出了一种称为 JA3 指纹识别的 TLS 指纹识别实现方法，此方法被整合到多个网络监控和入侵检测系统(IDS)中。用它进行恶意软件检测[57]或识别网络应用等。JA3(S)为特定客户端与服务器之间的加密通信提供具有更高识别度的指纹。为了启动 TLS 会话，客户端将在 TCP 三次握手后发送 TLS Client-hello 数据包。如果接受 TLS 连接，服务器将使用基于服务器端的库配置以及基于 Client-hello 详细信息创建的 TLS Server-hello 数据包进行响应。由于 TLS 协商[58]是以明文的方式传输的，所以可以使用 TLS Client-hello 数据包中的详细信息对客户端应用程序进行指纹识别。

通过 Wireshark 工具我们收集 Client-hello 数据包的协议版本，可接受密码套件，扩展列表，椭圆曲线密码和椭圆曲线密码格式 5 个字段值，用‘字段来分隔各个字段,用‘‘隔来分隔各个字段中的十进制值，再将这些值串联在一起并计算出 MD5,就是一个 JA3 。图 4 为百度抓取数据包。

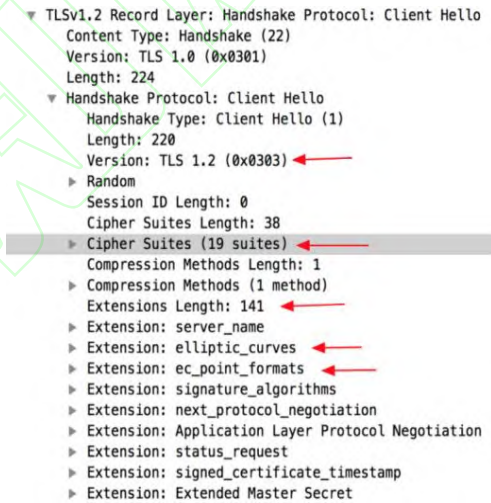

图 4　JA3 方法选择参数

Fig.4　JA3 Method Selection Parameters

JA3S 与 JA3 原理类似，提取了 Server-hello 数据包的相同值进行同样的操作。文献[59]中发现为取证目的创建唯一且稳定的 TLS 指纹并不容易，并在移动应用程序上进行了使用 JA3 散列的实验。JA3 指纹能够指示客户端应用程序通过 TLS 通信的方式，而 JA3S 指纹能够指示服务器响应。将两者结合起来，实质上就生成了客户端和服务器之间的加密协商的指纹。但这种方法不一定能保证映射到客户端应用程序，恶意用户可以通过恶意行为改变 TLS 版本，密码套件等信息，从而躲避检测。因此它的发展受到了限制。

3.3 证书技术

数字证书被定义为附加在公共加密密钥上的未加密文件，它包含关于证书和加密密钥所有者的组织细节。TLS 所使用的加密系统是基于对称加密的 RSA 标

准。通常在现实场景中服务器用 RSA 生成公钥和私钥后把公钥放在证书里发送给客户端，同时自己保存私钥，客户端收到消息后首先向一个权威的服务器（CA）检查证书的合法性（存在域验证，组织验证和扩展验证），如果证书合法，客户端产生一段随机数，这个随机数就作为通信的密钥，再用公钥加密这段随机数，然后发送到服务器，服务器用密钥解密获取对称密钥，然后，双方就可以进行加密通信了。但是证书可以由任何人创建并允许任何人加密和保护任何通信通道，这就滋生了潜在的风险。

因此，证书检测就十分有必要了，这一技术是通过证书链检测证书颁发者，证书使用者，证书序列号，证书指纹以及证书有效期这五项指标与原始的一致性，进而达到分类检测的目的，使用者完全可以自己生成证书并替换掉默认证书，或者采取修改证书的办法；在 TLS1.3 后对 Server Hello 报文后的所有信息采取加密处理，使可见明文大幅减少，证书信息也变为不可见，最终导致此类检测方法的检测能力大幅削弱。

3.4 基于机器学习的方法

虽然 TLS 加密方式对认证过程的大部分内容进行了加密，但是仍然可以得到一些非加密内容数据作为训练数据，使用人工智能算法仍然可以发现其中的规律，基于机器学习的加密恶意流量识别技术将加密流量进行恶意特征提取，构建一个恶意特征数据集，作为训练/测试集输入训练模型，通过模型设计与参数调优等方法得到理想的准确度。方法识别体系如图 5 所示。

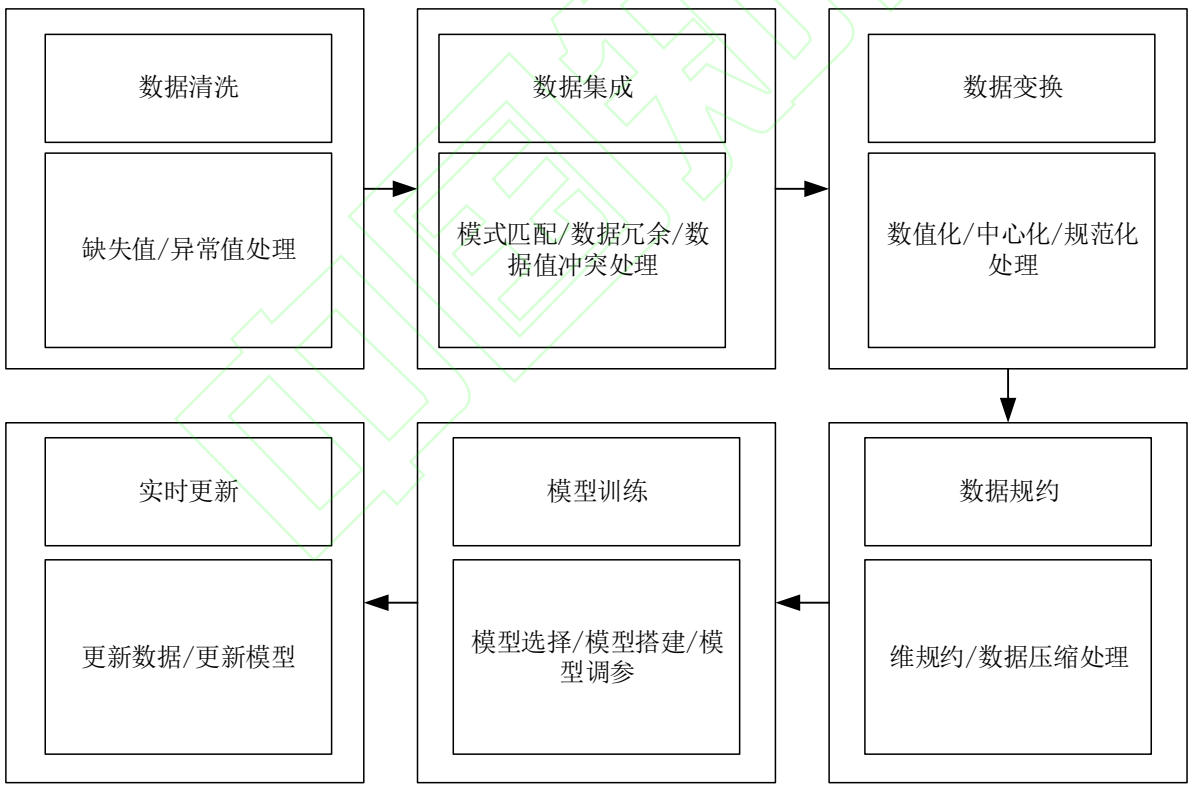

图 5　恶意流量识别方法体系

Fig.5　Malicious Traffic Identification Method System

文献[60]对机器学习训练过程所需实验各项评估参数做出了归纳，文献[61]介绍了机器学习的详细过程。但对于高度伪装的指挥控制(C&C)通信，单纯基于统计特征或 TLS 握手特征的传统分类器检测能力逐渐下降。在这种情况下，探索其他维度的特征进行多维特征融合的方式更具有针对性。在选择数据特征时常选用以下有效特征如表 10 所示。

表 10　常见有效特征总结

Table 10　Summary of common effective features

特征类别	特征名称
时间序列特征	包持续时间，包到达时间，流到达时间，流持续时间
统计特征	正/逆向包数量，正/逆向字节数，正/逆向包长度，正/逆向流长度
基本特征	源端口，目的端口，源 IP，目的 IP，协议，协议版本号
报文负载多协议特征	密码套件，证书自签名，证书有效性，公钥长度，证书数量，TTL 值，DNS 响应返回 IP 地址数，目标域名在前 100 万，HTTP 特征，以及其他协议相关联特征…

目前最新的一些关于机器学习对TLS恶意加密流　　量检测的研究结果如表 11，12 所示。

表 11　研究模型方法总结

Table 11　Summary of Research Model Methods

文献	识别特征	年份	数据集	方法描述	准确率
[62]	流量统计、SSL 握手字段和证书	2019	CTU13[63]	多视图特征的恶意流量检测	XGBOOST 模型 ACC:97.71%
[64]	TLS 握手元数据、DNS 上下文流，HTTP 上下文流	2016	个人真实数据	通过数据 Omnia 方法开发有监督机器学习模型	联合特征严格过滤后 ACC：99.993%
[65]	TLS 流量特征	2020	个人沙箱数据	使用扩展可观察元数据的半监督分类管道的增强方法，基于聚类方法的自动数据包数据标记管道生成数据	Mean-Shift ACC:87.56%
[66]	TLS 通信信道分布特征、一致性特征和统计特征	2020	良性个人真实数据，恶意来自 MT[67],CIC[68],SIPS[69]	以信道为检测单元的恶意 TLS 流量检测	ACC:92.57%
[70]	统计特征, TLS 握手特征，加密套件等	2021	个人真实和沙箱数据	结合多特征的恶意加密流量检测	ACC 达到 98%以上
[71]	报文负载特征，流指纹特征	2021	CTU13[63]	不依赖流量五元组特征的 TLS 加密恶意流量检测	ACC: 97.86%
[72]	TLS 特征，数据元特征，上下文数据特征	2020	个人真实和沙箱数据，部分公开数据	基于机器学习的分布式自动化恶意加密流量检测	随机森林 RECALL:97%
[73]	加密流量多尺度局部特征，双层全局特征	2021	CICAndMal2017[74]	基于 HST-MHSA 模型的端到端恶意加密流量识别	RECALL:93.49%
[75]	HELLO 报文与 TCP 协议交互信息	2021	CICIDS2017[74], CTU13[63],Malware-Traffic-Analysis[76]	用 HNNIM 模型来进行识别与分类	关于恶意样本 RECALL:98.9%
[77]	更独特的数据包长度和数据包到达时间的表示特征	2018	个人真实数据	自动编码器，卷积神经网络	CNN 获得 ACC:97.9%
[78]	无人工干预的原始流量特征	2019	ISCX VPN-non-VPN[79] ISCX2012IDS[80]	深度全范围(DFR)轻量级框架	最优的 DFR 分类器:一维 CDD-Based ACC:99.85%
[81]	TLS 证书相关特征	2018	个人真实数据	利用深度神经网络识别 TLS 证书恶意使用	识别恶意证书 ACC:94.87%
[82]	将原始包蒸馏为相同长度	2020	ISCX VPN-Non-VPN[79]	1D-CNN 和 EAV 框架	ACC：98%

表 12　研究方法优缺点总结

Table 12　A comparative summary of advantages and disadvantages of research methods

文献	优点	缺点
[62]	采用多视图特征融合方法提高了准确率	使用数据集过时，不具有代表性，实用性和真实性
[64]	融合其他协议特征加入模型提高了准确率同时个人真实数据具有真实性	样本不均衡，使用 DNS 和 http 一个重要假设是 DGA 和 HTTP C&C
[65]	通过聚类方法弥补了研究领域标记数据及缺乏的不足	数据集并没有被证实全部聚类方法，不具有说服力；改进增加管道方法并丰富数据多样性是进一步研究
[66]	融合多种通信信道特征提高检测准确率	特征不具有实时性，需要不断更新和探索，同时通道中包含流种类会影响特征选取和准确率
[70]	全面地提取了 TLS 流和统计流特征，数据集实时有效	实验只针对 TLS1.2 有效，缺乏对其他维度特征考虑
[71]	摆脱传统基于五元组特征的检测，提高了 30%多准确率	限于在单一网络环境流量下训练模型，无法保证未知标签加密流检测准确率
[72]	从三个特征维度，多个模型实现了检测对比	维度与思科研究相同，特征少于思科研究，同时采用公开数据不具有带代表性
[73]	创新性的引入多头注意力机制，利用深度学习改进了特征选择中表征效果不佳等问题	仅从算法层面改进不平衡数据分类效果，未考虑数据层面处理不平衡问题
[75]	利用深度学习模型解决了传统机器学习方法受专家经验的影响，识别与分类效果不理想的问题	实验只针对 TLS1.2 及以下版本做了分析研究
[77]	对比多个机器学习模型后，融合两种深度学习模型分类用于自动提取具有代表性特征和训练	无法保证对不同的特征集有稳定检测率，深度学习算法优化是下一步研究重点
[78]	实现了一个轻量级框架并使用深度学习模型在没有人工干预和私人信息的情况下从原始流量中学习	使用的数据不够多样，有效，欠缺说服力
[81]	将 TLS 证书与深度学习模型和循环神经网络结合，发现新的特征，更有效地分析文本数据	方法局限于网络钓鱼和恶意软件的检测
[82]	将深度包检测与深度学习模型融合提高检测准确率	只在实验数据集上的进行了验证工作，目前此方法的技术还不成熟，或将成为下一步重点研究

4 未来发展

4.1 恶意分类多样化

TLS 恶意加密流量识别研究主要集中于二分类或少数几类攻击的识别，由于应用程序和版本的多样性，实现加密恶意流量精细化识别还存在一定的难度[83]。目前已存在不少研究，类似于 1D-CNN 的诸多深度学习模型致力于拓宽加密流量的种类。在接下来的研究中，存在的问题可以总结为：首先，由于不同类型的流量有不同类型的数据包，选取更适合的字节数需要进一步研究；其次，目前存在的公开数据集不够丰富，种类不够齐全，个人数据集不够均衡，会导致模型训练不真实，对实验性能造成巨大影响，因此如何获取并公开种类丰富，数据量庞大的数据集就显得尤为关键。

4.2 领域技术迁移

在前面发展的基础上，在解决了数据集的问题后，不乏尝试将用在文本/图像处理，甚至语音识别，情景分析等深度学习模型应用到加密流量的检测领域。这些模型在本身的领域已有相当成熟的研究，且取得了不错的研究成果。如目前较为新颖的 BERT 模型，在解决 Transformer 模型需要训练大量的参数基础上，通过上下文全向实现自然语言文本的更精准识别处理。想要将 BERT 模型应用到本领域，还存在着下面的问题：如何高效准确地将 TLS 加密流量转换成如图像，自然语言处理文本，甚至语音进行处理。将胶囊神经网络（Capsule Network），对抗神经网络（Generative Adversarial Networks, GAN）等模型应用到加密恶意流

量识别中，如：在胶囊网络中可以通过将获取的 TLS 数据集（.PCAP 数据包等）转化为图像特征，并作为模型的原始数据输入进行训练，这些低层胶囊对其输入执行一些相当复杂的内部计算，然后将这些计算的结果封装成一个包含丰富信息的小向量；再如设计动机为自动化特征提取的 GAN 网络，利用 GAN 网络生成器，可以初步解决因为恶意流量少而导致的数据不平衡问题，并利用判别器迭代优化数据，以此有效提高自学习特征的可解释性和检测效率。

5 不足与展望

本文针对 TLS 协议中恶意流量检测的特点，将 TLS 协议中恶意流量检测技术分为传统的解密技术、代理技术、证书检测技术、JA3 (S)技术和机器学习技术等几个方面，并对最新研究成果进行了阐述。当前工作在分类效率和安全保障方面已取得较大进展，但仍存在一些问题：

（1）深度包检测技术：存在无法完全解析加密流量，降低网络性能，设备迭代困难，可视化不足，加密规则匹配上无法获得高效安全的算法支撑等问题。因此基于硬件的 DPI 技术迫切需要存储高效特征的匹配算法，根据不同的加密算法形成不同特征，进行规则匹配和加密算法特征探索或将成为进一步研究重点。如文献[82]开创性的将深度学习模型与 DPI 技术融合。

（2）代理检测技术：存在将客户端暴露于各种攻击中，导致用户信息被持续性跟踪；安全性显著降低；计算和通信方面的性能也会下降等问题。文献[84]系统地研究了相关技术并比较了它们的优缺点，借鉴设置二级代理的思想，如何更好的保证代理对数据的保密性和安全性以及提高实时网络性能会是进一步研究的重点。

（3）证书检测技术：存在自生成证书，替换默认证书或者修改证书的问题。目前已有将证书检测与机器学习相融合的方法，但尚未形成成熟体系，如何发现更有效证书特征和探索准确率更高的学习模型或成为进一步研究方向。

（4）数据集：存在缺乏带标签符合研究的公开数据集；缺乏被普遍接受的数据收集和标记方法；恶意软件变种和骨干网络统计特征的流量信息收集也存在困难等问题。基于机器学习的恶意流量识别核心在于正确的数据集，文献[85]提出了良好数据集的评估框架，文献[60]也对现有的公开数据集做了归纳总结。但 TLS 加密流量识别工作迫切需要一个开源、有正确标签、在恶意流量中有详细分类,并且能持续更新的数据集。同时在训练/测试集中的恶意流量分布必须符合现实分布[86]。因此，数据集的实时性、丰富多样性、有效性是进一步研究的重点。

（5）机器学习技术：存在模型的辨识性特征密度降低；模型拟合过慢；识别能力会随着时间迁移而整体下降等问题。恶意样本或僵尸网络主机往往会混淆或随机端口，而这些无规律和快速变化造成了五元组中端口特征的不稳定，不适合作为机器学习模型的学习特征。现阶段依然存在把五元组特征作为检测 TLS 加密流量的主要特征的研究。在样本数较少，采集环境相似的情况下，加密流量五元组特征高度相似；而样本数量复杂，采集环境不同的情况下，加密流量五元组特征又毫无规律。对不同元组特征的探索以及新型有效多维特征融合的方法；QUIC[87]，HTTP/3[88]协议使得流量数据包头部的明文占比进一步下降，这对基于机器学习多维特征的流量识别带来了进一步挑战，如何应对流量混淆，时间衰减，明文占比下降导致特征减少的问题将是下一步的研究重点。

6 总结

机器学习的检测方式给加密流量的识别工作带来了希望，但由于数据集种类不够丰富；实时公开数据集的数量匮乏以及恶意加密流量的激增，导致模型训练的正确性和有效性无法保证。因此，在确保数据集被有效丰富的前提下，可以预见，融合其他领域技术以及运用深度学习方法将会打破目前的桎梏，使检测模型在对抗中自学习发现隐藏的未知恶意加密流量，并解决流量检测的准确性能实时保持稳定的上升。

参考文献：

[1] Mary Meeker. Internet trends Online[EB/OL], (2019-06-11)https://www.bondcap.com/report/itr19/.

[2] Eckersley, Peter. How unique is your web browser?. Privacy Enhancing Technologies[C]. Berlin, German: Springer, 2010. 1 – 18.

[3] RFC 2246, The TLS Protocol Version 1.0[EB/OL]. (1999).https://tools.ietf.org/html/rfc2246.

[4] RFC 8446, The Transport Layer Security (TLS) Protocol Version 1.3[EB/OL]. [2018]. https://tools.ietf.org/html/rfc8446.

[5] 沈若愚,卢盛祺,赵运磊.TLS1.3 协议更新发展及其攻击与防御研究[J].计算机应用与软件，2017，34(11):264-269+329.
SHEN Ruo-yu, LU Sheng-qi, ZHAO Yun-lei. Research on the update and development of TLS1.3 protocol and its attack and defense[J]. Computer Applications and Software, 2017, 34(11):264-269+329.

[6] N Aviram, K Gellert, T Jager. Session Resumption Protocols and Efficient Forward Security for TLS 1.3 0-RTT. Advances in Cryptology – EUROCRYPT 2019[C]. Berlin, German: Springer, 2019. 117-150.

[7] C Cremers, M Horvat, S Hoyland, et al. Comprehensive Symbolic Analysis of TLS 1.3. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security[C]. New York, NY: ACM, 2017. 1773 – 1788.

[8] K Bhargavan, G Leurent. Transcript Collision Attacks: Breaking Authentication in TLS, IKE and SSH. NDSS[C]. San Diego, CA, USA: NDSS, 2016. 21-24.

[9] J Sherry, C Lan, P A Popa, et al. BlindBox: Deep Packet Inspection over Encrypted Traffic[J]. Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, 2015, 45(4): 213 – 226.

[10] O Levillain, B Gourdin, H Debar. TLS Record Protocol: Security Analysis and Defense-in-depth Countermeasures for HTTPS. In Proceedings of the 10th ACM Symposium on Information Computer and Communications Security[C]. New York, NY: ACM, 2015. 225-236.

[11] Fischlinm, Gntherf. Replay attacks on zero round-trip time: The case of the TLS1.3 handshake candidates. IEEE European Symposium on Security and Privacy[C]. New York, NY: IEEE Communications Societ, 2017. 82-113.

[12] The Many Flaws of Dual_EC_DRBG[EB/OL]. [2013-09-18].https://blog.cryptographyengineering.com/2013/09/18/the-many-flaws-of-dualecdrbg/ 2013.

[13] The Vulnerability of SSL to Chosen Plaintext Attack[EB/OL],[2004].http://eprint.iacr.org/2004/111 2004.

[14] G Bard. A Challenging but Feasible Blockwise-Adaptive Chosen-Plaintext Attack on SSL[J]. Proceedings of the International Conference on Security and Cryptography, 2006, 2006: 99-109.

[15] D Bleichenbacher. Chosen Ciphertext Attacks Against Protocols Based on the RSA Encryption Standard PKCS #1. 18th Annual International Cryptology Conference[C]. Santa Barbara, California, USA: CRYPTO, 1998. 1-12.

[16] B Canvel, A Hiltgen, S Vaudenay, et al. Password Interception in a SSL/TLS Channel. In Advances in Cryptology-CRYPTO 2003, 23rd Annual International Cryptology Conference[C]. Santa Barbara, California, USA: CRYPTO, 2003. 583-599.

[17] Security of CBC Ciphersuites in SSL/TLS: Problems and Countermeasures[EB/OL]. [2004]. http://www.openssl.org/~bodo/tls-cbc.txt 2004.

[18] S Vaudenay. Security Flaws Induced by CBC Padding - Applications to SSL, IPSEC, WTLS. International Conference on the Theory and Applications of Cryptographic Techniques[C]. Amsterdam, The Netherlands: EUROCRYPT, 2002. 534-546.

[19] J Nadhem, AlFardan, G Kenneth. Lucky Thirteen: Breaking the TLS and DTLS Record Protocols. In 2013 IEEE Symposium on Security and Privacy[C]. Berkeley, CA, USA: IEEE, 2013. 526-540.

[20] N Aviram, S Schinzel, J Somorovsky, et al. DROWN: Breaking TLS with SSLv2. In 5th USENIX Security Symposium[C]. Austin, TX: USENIX, 2016. 689-706.

[21] K Bhargavan, G Leurent. On the Practical (In-)Security of 64-bit Block Ciphers: Collision Attacks on HTTP over TLS and Open VPN. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security[C]. New York, NY, USA: ACM, 2016. 456 – 467.

[22] K Bhargavan, G Leurent. Transcript collision attacks: Breaking authentication in TLS, IKE, and SSH. In Network and Distributed System Security Symposium – NDSS 2016[C]. San Diego, CA, USA: NDSS, 2016. 1-17.

[23] C Garman, K Paterson and T V D Merwe. Attacks Only Get Better: Password Recovery Attacks Against RC4 in TLS. 24th USENIX Security Symposium[C]. Washington, D.C.: USENIX, 2015. 113 – 128.

[24] T Jager, J Schwenk, J Somorovsky. On the Security of TLS 1.3 and QUIC Against Weaknesses in PKCS#1 v1.5 Encryption. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communi-

cations Security[C]. CO, USA: ACM, 2015. 1185－1196.

[25] Attacking SSL when using RC4[EB/OL].[2015]. https://www.imperva.com/docs/HII_Attacking_SSL_when_using_RC4.pdf 2015.

[26] N Mavrogiannopoulos, F Vercauteren, V Velichkov, et al. A cross-protocol attack on the TLS protocol. In the ACM Conference on Computer and Communications Security[C]. Raleigh, NC, USA: ACM, 2012. 62－72.

[27] This POODLE bites: Exploiting The SSL 3.0 Fallback[EB/OL].[2014].https://www.openssl.org/~bodo/ssl-poodle.pdf 2014.

[28] 张兴隆，程庆丰，马建峰. TLS 1.3 协议研究进展[J]. 武汉大学学报(理学版), 2018, 64(6):471-484.

ZHANG Xing-long, CHENG Qing-feng, MA Jian-feng. Research Progress of TLS 1.3 Protocol[J]. Journal of Wuhan University (Science Edition), 2018, 64(6): 471-484.

[29] O Levillain. Implementation Flaws in TLS Stacks: Lessons Learned and Study of TLS 1.3 Benefits. Risks and Security of Internet and Systems[C]. Berlin, German: Springer, 2020. 87-104.

[30] L Akhmetzyanova, E Alekseev, E Smyshlyaeva, et al. On post-handshake authentication and external PSKs in TLS 1.3. Journal of Computer Virology and Hacking Techniques[C]. Berlin, German: Springer, 2020. 1-6.

[31] R Liu, YU Xiang-zhan. A Survey on Encrypted Traffic Identification. Proceedings of the 2020 International Conference on Cyberspace Innovation of Advanced Technologies[C]. New York, NY: ACM, 2020. 159-163.

[32] B Anderson, D McGrew. Identifying Encrypted Malware Traffic with Contextual Flow Data. Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security[C]. New York, NY: ACM, 2016. 36-46.

[33] J Hurley, E G Palacios, S Sezer. Host-Based P2P Flow Identification and Use in Real-Time[J]. ACM Transactions on the Web, 2011, 5(2):1-7.

[34] T Qin, L Wang, Z Liu, et al. Robust application identification methods for P2P and VoIP traffic classification in backbone networks[J]. Knowledge Based Systems, 2015, 82(jul.):152-162.

[35] 饶瑾. 深度包检测(DPI)技术浅谈及应用[J]. 信息通信,2014，11(182):245-246.

RAO Jin. Discussion and Application of Deep Packet Inspection (DPI) Technology[J]. Information and Communication, 2014，11(182):245-246.

[36] S Canard, A Diop, N Kheir, et al. Blindids: Market-compliant and privacy-friendly intrusion detection system over encrypted traffic. In Asia CCS 2017[C]. New York, NY, USA: ACM, 2017. 561－574.

[37] J Sherry, C Lan, P A Popa, et al. BlindBox: Deep Packet Inspection over Encrypted Traffic[J]. Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication, 2015, 45(4): 213－226.

[38] 孙中军，翟江涛，戴跃伟. 一种基于DPI和负载随机性的加密流量识别方法[J]. 应用科学学报. 2019. 37(5):711-720.

SUN Zhong-jun, ZHAI Jiang-tao, DAI Yue-wei. An encryption flow identification method based on DPI and load randomness[J].Journal of Applied Sciences. 2019. 37(5):711-720.

[39] J. Michael Butler, David Wells.

Finding Hidden Threats by Decrypting SSL/TLS[EB/OL].[2013-11-08].https://www.sans.org/webcasts/finding-hidden-threats-decrypting-ssl-tls-97315.

[40] Snort and SSL/TLS Inspection[EB/OL]. [2017]. https://www.sans.org/reading-room/whitepapers/detection/paper/37735/.

[41] T Radivilova, L Kirichenko, D Ageyev, et al. DecryptingSSL/TLS traffic for hidden threats detection. 2018 IEEE 9th International Conference on Dependable Systems[C]. New York, NY, USA: ACM, 2018. 143-146.

[42] J Baek, J Kim, W Susilo. Inspecting TLS Anytime Anywhere: A New Approach to TLS Interception. Proceedings of the 15th ACM Asia Conference on Computer and Communications Security[C]. New York, NY, USA: ACM, 2020. 116－126

[43] S Ruoti, M O Neill, D Zappala, et al. User Attitudes Toward the Inspection of Encrypted Traffic[J]. Arxiv: 1510.04921. 2016.

[44] L S Huang, A Rice, E Ellingsen, et al. Analyzing Forged SSL Certificates in the Wild[J]. 2014 IEEE Symposium on Security and Privacy.2014. 1: 83-97.

[45] T Hunt, Z Zhu, Y Xu, et al. Ryoan: A Distributed Sandbox for Untrusted Computation on Secret Data[J]. Association for Computing Machinery. 2018. 35(4):1-34.

[46] M A Jamshed, Y Moon, D Kim, et al. MOS: A Reusable Networking Stack for Flow Monitoring Middleboxes. 14th USENIX Symposium on Networked Systems Design and Implementation[C]. Boston, MA: USENIX, 2017. 113-129.

[47] SSL/TLS Interception Proxies and Transitive Trust[EB/

OL]. [2012]. https://media.blackhat.com/bh-eu-12/Jarmoc/bh-eu-12-Jarmoc-SSL_TLS_Interception-WP.pdf.

[48] D Naylor, K Schomp, M V arvello, et al. Multi-Context TLS (mcTLS): Enabling Secure In-Network Functionality in TLS[J]. SIGCOMM Computer. 2015. 45(4):199–212.

[49] J Sherry, C Lan, R A Popa. BlindBox: Deep Packet Inspection over Encrypted Traffic[J]. SIGCOMM Computer. 2015. 45(4). 213–226.

[50] N Weaver, C Kreibich, M Dam, et al. Here be web proxies, Passive and Active Measurement[C]. Berlin, German: Springer, 2014. 183-192.

[51] Z Durumeric, Z Ma, D Springall, et al. The Security Impact of HTTPS Interception. In Network and Distributed Systems Symposium[C]. San Diego, CA, USA: NDSS, 2017.

[52] A Mani, T Vaidya, D Dworken, et al. An Extensive Evaluation of the InternetptiOpen Proxies. Proceedings of the 34th Annual Computer Security Applications Conference[C]. New York, NY, USA: ACM, 2018. 252–265.

[53] J Han, S Kim, J Ha, et al. SGX-Box: Enabling Visibility on Encrypted Traffic using a Secure Middlebox Module. Proceedings of the First Asia-Pacific Workshop on Networking[C]. New York, NY, USA: ACM, 2017. 99–105. 2016.

[54] M Hus2016.cure Middlebox Module. Proceedings ffic analysis and client identification using passive SSL/TLS fingerprinting[J]. EURASIP Journal on Information Security. 2016:1-14.

[55] M J Erquiaga, S Garc6:1-14. SIP Journal on Inforc analysis and client identififfic Forces Malware to Change Their Behavior. Communications in Computer and Information Science[C]. Berlin, German: Springer, 2017. 272-281.

[56] G Tsirantonakis, P Ilia, S Ioannidis, et al. A large-scale analysis of content Modifcation By Open Http Proxies. Network and Distributed Systems Security[C]. San Diego, CA, USA:NDSS, 2018.1-15.

[57] B Anderson, S Paul, D McGrew. Deciphering malwarerge-scale analysis of content Modifcat Journal of Computer Virology and Hacking Techniques. 2018. 14: 195-211.

[58] Kotzias, P Razaghpanah, A Amann, et al. Coming of age: a longitudinal study of TLS deployment. Proceedings of the Internet Measurement Conference[C]. New York, NY, USA : ACM, 2018. 415–428.

[59] P Matousek, I Burgetova, O Rysavý, et al. On Reliability of JA3 Hashes for Fingerprinting Mobile Applications. Digital Forensics and Cyber Crime[C]. Berlin, German: Springer, 2021. 1-22.

[60] 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(03): 66-77.
ZHAI Ming-fang, ZHANG Xing-ming, ZHAO Bo. Research on Encrypted Malicious Traffic Detection Based on Deep Learning[J]. Journal of Network and Information Security, 2020, 6(03):66-77.

[61] 曾勇，吴正远，董丽华，刘志宏，马建峰，李赞. 加密流量中的恶意流量识别技术[J]. 西安电子科技大学报, 2021:1-18.
ZENG Yong, WU Zheng-yuan, DONG Li-hua, LIU Zhi-hong, MA Jian-feng, LI Zan. Malicious traffic identification technology in encrypted traffic [J]. Journal of Xidian University, 2021:1-18.

[62] DAI Rui, GAO Chuan, LANG Bo. SSL Malicious Traffic Detection Based On Multi-view Features. Proceedings of the 2019 the 9th International Conference on Communication and Network Security[C]. New York, NY, USA: ACM, 2019. 40–46.

[63] CTU Malware Capture Facility Project[EB/OL]. [2019].https://mcfp.weebly.com/the-ctu-13-dataset-a-labeled-dataset-with-botnet-normal-and-background-traffic.html.

[64] B Anderson, D McGrew. Identifying Encrypted Malware Traffic with Contextual Flow Data. Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security[C]. New York, NY, USA: ACM, 2016. 35-46.

[65] S M Raza, J Caballero. Malware Traffic Classification: Evaluation of Algorithms and an Automated Ground-truth Generation Pipeline[J]. arXiv: 2010.11627. 2020.

[66] R Zheng, J Liu, K Li, et al. Detecting Malicious TLS Network Traffic Based on Communication Channel Features. 2020 IEEE 8th International Conference on Information Communication and Networks[C]. New York, NY: IEEE Communications Society. 2020. 14-19.

[67] A Source For Pcap Files And Malware Samples[EB/OL]. [2020].https://www.malware-traffic-analysis.net/.

[68] Malware Capture Facility Project[EB/OL]. [2020]. https://www.stratosphereips.org/datasets-malware.

[69] Canadian Institute For Cybersecurity[EB/OL]. [2020]. https://www.unb.ca/cic/datasets/index.html, 2020.

[70] 李慧慧，张士庚，宋虹，王伟平. 结合多特征识别的恶意加密流量检测方法[J]. 信息安全学报，2021, 6(2): 129-142.
LI Hui-hui, ZHANG Shi-geng, SONG Hong, WANG Wei-ping. Malicious encrypted traffic detection method combined with multi-feature recognition[J].

Journal of Information Security, 2021,6(2):129-142.

[71] 胡斌,周志洪,姚立红,李建华.结合报文负载与流指纹特征的恶意流量检测[J].计算机工程, 2020, 46(11): 157-163.
HU Bin, ZHOU Zhi-hong, YAO Li-hong, LI Jian-hua. Malicious traffic detection combining packet load and flow fingerprint characteristics[J/OL]. Computer Engineering, 2020, 46(11):157-163.

[72] 骆子铭, 许书彬, 刘晓东. 基于机器学习的 TLS 恶意加密流量检测方案[J]. 网络与信息安全学报, 2020, 6(1):77-83.
LUO Zi-ming, XU Shu-bin, LIU Xiao-dong. TLS malicious encrypted traffic detection scheme based on machine learning[J]. Journal of Network and Information Security, 2020, 6(1): 77-83.

[73] 蒋彤彤, 尹魏昕, 蔡冰, 张琨. 基于多头注意力的恶意加密流量识别[J/OL]. 计算机工程, 2021:1-14.
JIANG Tong-tong, YIN Wei-xin, CAI Bing, ZHANG Kun. Identification of malicious encrypted traffic based on multi-head attention [J/OL]. Computer Engineering, 2021:1-14.

[74] Lashkari, A Habibi, Kadir, et al. Toward Developing a Systematic Approach to Generate Benchmark Android Malware Datasets and Classification. ICCST[C]. New York, NY: IEEE Communications Society, 2018. 1-7.

[75] 韦佶宏, 郑荣锋, 刘嘉勇. 基于混合神经网络的恶意 TLS 流量识别研究[J]. 计算机工程与应用, 2021, 57(07):107-114.
WEI Ji-hong, ZHENG Rong-feng, LIU Jia-yong. Research on malicious TLS traffic identification based on hybrid neural network[J]. Computer Engineering and Applications, 2021, 57(7): 107-114.

[76] Malware-Traffic-Analysis[EB/OL]. [2019].https://www.malware-traffic-analysis.net.

[77] Y Yang, C Kang, G Gou, et al. TLS/SSL Encrypted Traffic Classification with Autoencoder and Convolutional Neural Network. 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems[C]. New York, NY: IEEE Communications Society. 2018. 362-369.

[78] Y Zeng, H Gu, W Wei, et al. Deep-Full-Range : A Deep Learning Based Network Encrypted Traffic Classification and Intrusion Detection Framework[J]. IEEE Access, 2019, 7(): 45182-45190.

[79] G Draper-Gil, A H Lashkari, M S I Mamun, et al. Characterization of Encrypted And VPN Traffic Using Time-related[J]. ICISSP, 2016, 407-414.

[80] A Shiravi, H Shiravi, M Tavallaee, et al. Toward Developing a systematic approach to generate benchmark datasets forintrusion detection[J]. Comput. Secur, 2012, 31(3):357-374.

[81] I Torroledo, L D Camacho, A CorreaBahnsen. Hunting Malicious TLS Certificates with Deep Neural Networks. Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security[C]. New York, NY, USA:ACM, 2018. 64–73.

[82] M Lotfollahi, M J Siavoshani, M Saberian. Deep packet: a novel approach for encrypted traffic classification using deep learning[J]. Soft Computing, 2020, 24() : 1999-2012.

[83] P Velan, P Celeda, M Drasar, et al. A survey of methods for encrypted traffic classification and analysis[J]. International Journal of Network Management, 2015, 25(5): 355-374.

[84] G Poh, D Divakaran, H Lim, et al. A Survey of Privacy-Preserving Techniques for Encrypted Traffic Inspection over Network Middleboxes[J]. arXiv:2101.04338. 2021.

[85] A GHARIB, I SHARAFALDIN, L A HABIBI, et al. An evaluation framework for intrusion detection dataset. 2016 International Conference on Information Science and Security (ICISS)[C]. New York, NY: IEEE Communications Society, 2016. 1-6.

[86] F Pendlebury, F Pierazzi, R Jordaney. TESSERACT: Eliminating Experimental Bias in Malware Classification across Space and Time[J]. USENIX Security Symposium, 2019,

[87] Sinha, Gaurav, Kanagarathinam, et al. CQUIC: Cross-Layer QUIC for Next Generation Mobile Networks. 2020 IEEE Wireless Communications and Networking Conference (WCNC)[C]. Seoul, Korea (South): IEEE, 2020. 1-8.

[88] Draft-ietf-quic-http-34, Hypertext Transfer Protocol Version 3 (HTTP/3)[EB/OL]. [2021]. https://datatracker.ietf.org/doc/draft-ietf-quic-http.