软件学报 ISSN 1000-9825, CODEN RUXUEW
[doi: 10.13328/j.cnki.jos.007182] [CSTR: 32375.14.jos.007182]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

Tor 被动流量分析综述*

梅汉涛[1,2,3,4], 程 光[1,2,3,4], 朱怡霖[1,2,3,4], 周余阳[1,2,3,4]

[1](东南大学 网络空间安全学院, 江苏 南京 211189)

[2](江苏省泛在网络安全工程研究中心 (东南大学), 江苏 南京 211189)

[3](网络空间国际治理研究基地 (东南大学), 江苏 南京 211189)

[4](紫金山实验室, 江苏 南京 211189)

通信作者: 程光, E-mail: chengguang@seu.edu.cn

摘 要: 随着网络的蓬勃发展, 用户隐私正面临着前所未有的挑战. 人们开发出多种匿名通信系统来保护隐私, 第2代洋葱路由 Tor (the second-generation onion router) 是目前最为广泛使用的匿名通信系统. 然而, 卓越的匿名性也使之成为不法分子犯罪的温床, 如今 Tor 中充斥着非法交易、网络犯罪等. Tor 被动流量分析通过被动观察网络流量对 Tor 进行去匿名化, 已成为最热门的去匿名化技术. 从 Tor 与流量分析基本概念出发, 介绍 Tor 被动流量分析技术的应用场景与威胁模型. 按照技术类型将现有工作分为流量分类技术与流关联技术, 依据分析流程分别对比其流量采集方法、特征提取方法、使用算法. 最后探讨当前研究面临的主要挑战与未来可能的研究趋势.

关键词: Tor; 流量分析; 流量分类; 网站指纹; 流关联

中图法分类号: TP393

中文引用格式: 梅汉涛, 程光, 朱怡霖, 周余阳. Tor被动流量分析综述. 软件学报. http://www.jos.org.cn/1000-9825/7182.htm

英文引用格式: Mei HT, Cheng G, Zhu YL, Zhou YY. Survey on Tor Passive Traffic Analysis. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7182.htm

Survey on Tor Passive Traffic Analysis

MEI Han-Tao[1,2,3,4], CHENG Guang[1,2,3,4], ZHU Yi-Lin[1,2,3,4], ZHOU Yu-Yang[1,2,3,4]

[1](School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China)

[2](Jiangsu Province Engineering Research Center of Security for Ubiquitous Network (Southeast University), Nanjing 211189, China)

[3](International Governance Research Base of Cyberspace (Southeast University), Nanjing 211189, China)

[4](Purple Mountain Laboratories, Nanjing 211189, China)

Abstract: The growth in the Internet poses privacy challenges, prompting the development of anonymous communication systems like the most widely used Tor (the second-generation onion router). However, the notable anonymity offered by Tor has inadvertently made it a breeding ground for criminal activities, attracting miscreants engaged in illegal trading and cybercrime. One of the most prevalent techniques for de-anonymizing Tor is Tor passive traffic analysis, where in anonymity is compromised by passively observing network traffic. This study aims to delve into the fundamental concepts of Tor and traffic analysis, elucidate application scenarios and threat models, and classify existing works into two categories: traffic identification & classification, and flow correlation. Subsequently, their respective traffic collection methods, feature extraction techniques, and algorithms are compared and analyzed. Finally, the primary challenges faced by current research in this domain are concluded and future research directions are proposed.

Key words: Tor; traffic analysis; traffic classification; website fingerprinting; flow correlation

---

* 基金项目: 国家自然科学基金 (U22B2025,62172093,62202097); 江苏省揭榜挂帅项目 (BE2023004-3); 中国博士后面上基金项目 (2022M70677); 江苏省卓越博士后计划 (2022ZB137)
收稿时间: 2023-09-01; 修改时间: 2023-11-27, 2024-03-05; 采用时间: 2024-03-26; jos 在线出版时间: 2024-11-01

1 引 言

互联网的飞速发展从根本上改变了人们的生活方式, 与此同时, 人们的个人信息通过浏览记录、电子银行、远程医疗等服务与互联网深度绑定, 个人隐私也正面临着前所未有的挑战. 虽然 TLS1.3 等技术的成功应用与 HTTPS 的推广显著地提升了用户访问的安全性, 但数据 IP、DNS、握手等通信过程仍然会泄露许多用户访问信息. 匿名通信技术的出现在极大程度上满足了用户隐私保护的需求[1–3]. 从早期的高延迟匿名技术 MixNet[4]、Mixminion[5]到如今的低延迟匿名通信 Tor、I2P[6]、Freenet[7], 匿名通信技术经过了长足的发展, 终于走向了成熟与应用.

Tor, 即第 2 代洋葱路由[8], 是如今最为广泛应用的匿名通信技术, 它通过流量混淆、数据多跳转发、内容多层加密等措施隐藏通信内容与通信关系, 实现了对用户与服务的隐私保护. 如今, Tor 在全球拥有超过 7000 个节点, 每天用户超过 400 万[9], 已成为许多记者、异见人士、情报人员隐蔽地连接互联网的手段. 数百万人每天通过 Tor 等匿名软件互相连接、传播或获取资源, 形成了“暗网”. 然而, 卓越的匿名性也使之成为了不法分子犯罪的温床, 2013 年 7 月, Freedom Hosting 服务器由于提供违法网络服务被查封; 同年 10 月, 臭名昭著的黑市网站丝绸之路被关闭; 然而, 仅 1 个月后丝绸之路 2.0 上线, 于 2014 年 11 月关闭后又迅速诞生了丝绸之路 3.0[10]. 虽然相关执法机构之后几年接连关闭了多个暗网黑市, 如 AlphaBay 和 Hansa 等, 但暗网黑市并不会消失, 而总是迅速转移到新的网站. 由于 Tor 极高的匿名性, 即便成功地关闭了黑市服务, 执法机构追踪到的违法用户与非法服务提供商也寥寥无几. 在这种情形下, 暗网非法服务愈演愈烈, 2022 年, Hydra 黑市被披露在 7 年内交易已超 50 亿美元[11]. 如今, 暗网中遍布着非法商品交易、金融犯罪、恶意软件等内容, 严重违背了保护用户隐私和匿名性的设计初衷, 已成为全球网络安全的严重威胁.

自 Tor 问世 20 多年来, 学术界开展了大量针对 Tor 的去匿名化攻击研究. 其中被动流量分析是研究最多、对 Tor 匿名威胁最大的研究方向[12]. 流量分析通过监控网络活动, 从网络流量中发现特定模式以收集有价值的信息[13]. 被动流量分析则旨在不影响用户访问 Tor 网络或不妨碍服务器提供服务的情况下, 通过监测用户流量、主机状态、网络态势等挖掘 Tor 实体信息, 对用户或服务器进行去匿名化. 根据技术类型, Tor 被动流量分析又可以分为 Tor 流量分类问题与 Tor 流关联问题. 流量分类判断用户是否在使用 Tor 网络以及识别用户使用 Tor 访问的内容, 往往包括匿名工具、服务类型、应用程序、网站等, 其中识别用户访问网站的技术被称为网站指纹攻击; 流关联则试图关联客户端流与服务器流, 确定双端通信关系. 被动流量分析方法实现成本低, 成功率高, 有大量的理论基础与成功实践, 已成为近些年网络安全研究的热点.

已有一些文献系统地总结了加密流量分析[14]与 Tor 流量分类[15]相关的研究工作. 其中的一些文献专注于对 Tor 去匿名化工作的梳理[16,17], 仅涵盖了部分 Tor 流量分析内容; 另一些文献则专注于 Tor 流量分析的某些细分领域[18], 缺乏了对 Tor 流量分析研究完整、全面的评估. 本文对 Tor 被动流量分析的研究成果进行了分类和梳理, 形成如图 1 的结构脉络. 首先, 根据技术类型, 本文将现有工作分为流量分类与流关联研究. 在流量分类方面, 根据流量分类流程按照单端流量采集方法、流量特征提取方法、算法选择对现有工作进行梳理. 进一步, 根据每个流程中使用的具体方法进行细分. 在流关联方面, 按照双端流量获取方法、流关联特征、流关联方法进行梳理. 在研究挑战部分, 本文讨论了 Tor 被动流量分析领域面临的最大挑战, 包括过于理想的威胁模型、数据集稀缺、基本比率谬误、概念漂移等问题, 最后总结了当前工作的应对方法以及未来工作可能的发展趋势.

本文第 2 节概述 Tor 匿名通信技术的基本原理, Tor 被动流量分析工作的基本威胁模型和当前 Tor 流量分析的相关综述, 厘清本文综述的重点. 在此基础上, 第 3 节总结 Tor 流量分类工作. 第 4 节总结 Tor 流关联工作. 最后, 本文在第 5 节给出当前研究挑战与展望, 并在第 6 节进行总结.

2 Tor 被动流量分析概述

Tor 是基于重路由机制的匿名通信系统, 使用多层加密和多跳转发隐藏通信内容与通信关系. 被动流量分析在不影响用户通信的前提下通过观察网络流量挖掘用户信息. 在第 2.1 节简要地介绍了 Tor 匿名通信技术, 重点介绍了 Tor 数据包格式, 这是 Tor 流量比常规加密流量更难识别的根源. 在第 2.2 节描述了 Tor 被动流量分析涉及

的基本概念与威胁模型, 这是被动流量分析工作中的基本假设条件. 在第 2.3 节梳理了流量分析尤其是 Tor 被动流量分析研究相关的综述, 并厘清了本综述重点关注的方向.

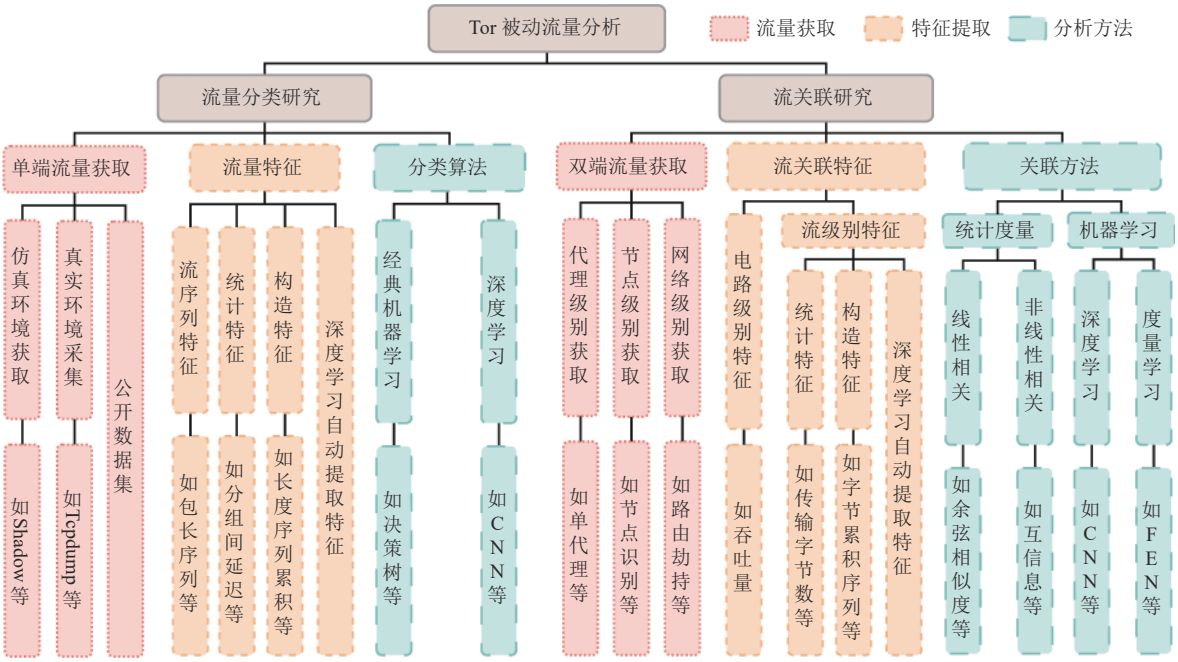

图 1 综述总体框架

2.1 Tor 匿名通信技术

第 2 代洋葱路由 Tor (the second-generation onion router), 是一种建立在 TCP 传输基础上的基于重路由机制的低延迟匿名通信系统, 主要用于及时通信、Web 访问等, 是目前最为广泛使用的匿名通信系统. Tor 于 2004 年的 USENIX 安全讨论会上发表, 同年, 其源码正式公开. 从那时开始, 学术界、技术界、民众开始了 Tor 的使用、研究与讨论. 在长达十几年的时间里, 不断有新的针对 Tor 的攻击技术提出甚至实施, Tor 则不断改进, 变得越来越安全. Tor 匿名通信包含 3 个基本组件: 目录服务器 (directory server, DS)、洋葱代理 (onion proxy, OP) 和洋葱路由器 (onion router, OR), 如图 2 所示.

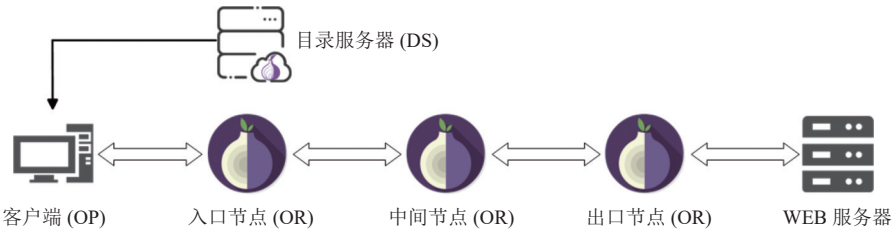

图 2 Tor 匿名通信原理

目录服务器 (directory server, DS): 目录服务器主要用于维护洋葱路由节点信息. 当中间节点想要加入 Tor 网络, 它需要将自己的带宽、地址、密钥、策略等信息提交给目录服务器, 目录服务器对节点进行评估, 并将相关信息保存为共识文件发布给所有用户. 当用户需要使用 Tor 网络, 它可以向目录服务器请求共识文件, 下载相关节点信息, 选择若干节点建立通信连接.

洋葱代理 (onion proxy, OP): 洋葱代理一般由客户端运行, 负责建立洋葱通道, 协商密钥, 并加密和解密客户端

与服务器通信过程中的网络流量. 洋葱代理还可以搭载混淆插件, 实现流量混淆以抵抗流量分析.

洋葱路由器 (onion router, OR): 洋葱路由器是 Tor 网络的核心, 它们由许多个人和组织志愿贡献设备与带宽资源, 提供 Tor 服务. Tor 用户的流量都由洋葱路由器转发和加解密, 让攻击者无法破解通信内容以及对访问溯源. 每次客户端实现 Tor 匿名通信, 都将至少通过 3 个洋葱路由器构建 Tor 通道, 再使用通道进行网络连接. 根据通道中的位置, 洋葱路由器可以被称为入口节点、中间节点和出口节点.

2.1.1 Tor 通道构建

洋葱代理将用户数据代理到 Tor 通道以实现匿名通信. 在开始通信之前, 客户端运行洋葱代理, 访问目录服务器并请求中继节点列表的共识文件, 然后基于加权随机规则选择至少 3 个节点. 接着, 洋葱代理依次与这些节点建立连接并协商密钥以构建 Tor 通道, 也称 Tor 电路 (circuit), 在协商密钥过程中, 使用共识文件中包含的中继节点公钥保证客户端请求不被篡改, 使用 Diffie-Hellman 握手保证密钥安全交换. 根据在通道中的位置, Tor 中继节点 (relay) 可以分为入口节点 (entry node), 中间节点 (middle node) 和出口节点 (exit node).

建立匿名通道后, 洋葱代理将用户数据进行多层加密并向后传输, 各个中继节点将数据解密并发送到下一跳, 最终到达目的服务器. 在整个通道建立和数据传输过程中, 客户端只与入口节点发生通信, 每个中继节点只与上一跳与下一跳通信, 没有任何一方可以同时获知客户端与目的服务器以及数据内容, 从而实现匿名. OP 在应用层将通信内容封装为 512 字节的 cell, 并交于下层协议 (TCP) 转发, 使流量分析变得更加困难.

2.1.2 隐藏服务

洋葱代理软件为服务请求者 (客户端) 提供了快捷、安全的匿名服务, 然而, 服务提供者 (Web 服务器) 却直接暴露在网络中. 当 Web 服务器也想要隐藏自己身份时, 就需要使用隐藏服务. 隐藏服务原理如图 3 所示. Tor 隐藏服务构建过程如下.

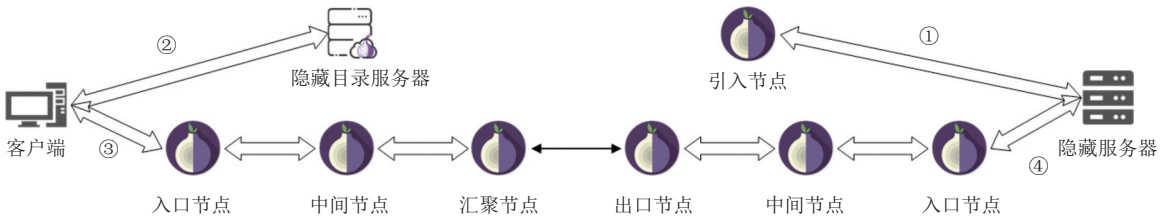

图 3 Tor 隐藏服务原理

● 第 1 步: 隐藏服务器建立一条长期存在的 Tor 通道, 该条通道的出口节点称为引入节点 (introduction point). 隐藏服务器会将引入节点及其公钥信息上传至隐藏服务目录服务器.

● 第 2 步: 客户端通过 Tor 通道获取隐藏服务相关信息, 包括引入节点地址、公钥等.

● 第 3 步: 客户端构建一条通道, 将该通道的末端作为客户端和隐藏服务器通信通道的汇聚点, 称为汇聚节点 (rendezvous point).

客户端通过一条 Tor 通道与引入节点连接, 将汇聚节点信息如地址与公钥等发送给隐藏服务器.

● 第 4 步: 隐藏服务器通过一条通道与汇聚节点建立连接, 开始与客户端的通信. 至此, 客户端与隐藏目录服务器均实现了自己身份的匿名.

2.2 威胁模型

威胁模型是 Tor 流量分析的基本假设. 它反映了攻击者的角色, 具备的能力以及用户的行为假设[1]. 攻击者可以自己运行客户端或服务器, 也可以部署或劫持洋葱路由器以截获流量. Tor 流量分析的通用威胁模型如图 4 所示. 通常, 流量分类研究假设攻击者是能够监听用户流量的设备和节点. 这可以是靠近用户侧的客户端、网关、ISP 等, 也可以是靠近 Tor 通道的入口节点. 从位置上看, 这些模型都处在用户到入口节点之间, 因为在整条 Tor 通道中, 只有这一段路径上的设备直接与用户连接, 可以知晓用户身份. 攻击者不破译通信内容, 仅通过分析数据交

互模式、流量流通计数、数据包统计特征等识别用户是否在使用 Tor、使用 Tor 的应用甚至访问的网站, 其中识别网站的攻击又被称为网站指纹攻击.

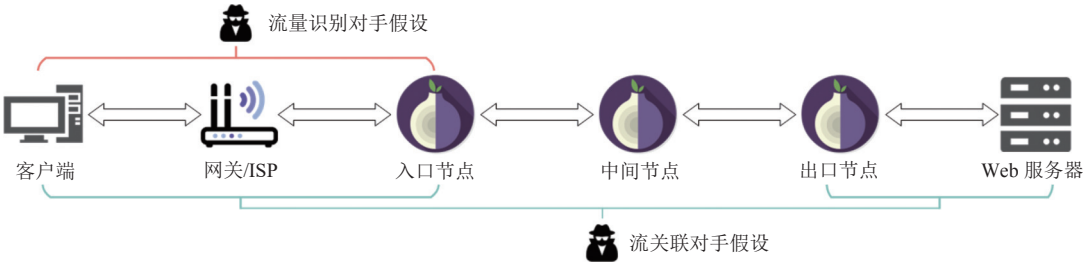

图 4 Tor 被动流量分析威胁模型

流关联研究则通常假设攻击者能够掌控匿名通道的入口和出口流量. 攻击者可以在 Tor 网络中植入足够多的洋葱路由器, 并利用 Tor 节点选中机制或攻击其他节点来提高自己被选中的概率, 也可以作为域级别 (AS) 的监管者[19]截获两端流量. 与流量分类相比, 流关联可以直接确认客户端与服务器关系, 这意味着它的去匿名化程度更高, 同时它的难度也更高. 因此, 流关联的威胁模型往往包含着比流量识别与分类更严格的假设, 且更依赖于其他对 Tor 进行的研究, 例如如何部署蜜罐研究[20], 如何提升节点被选中的概率研究等[12].

流量分析涉及一些基本概念. 如流、双向流、突发等. 流指一段时间内具有相同五元组 (源 IP, 目的 IP, 源端口, 目的端口, 通信协议) 的数据包序列. 双向流则是五元组中源 IP, 源端口与目的 IP, 目的端口互相交换的两个方向的数据包序列, 即同时包含上行与下行的数据包. 其中, 上行指的是从客户端到服务器的数据包, 下行则相反. 一条双向流一般是一条连接的通信往返数据. 流量突发是某个方向上突发的连续快速发送的数据包序列.

2.3 Tor 流量分析相关综述

自 Tor 发布以来, 针对 Tor 的去匿名化工作受到了极大的关注, 许多综述从不同角度梳理了 Tor 去匿名化相关的工作. 表 1 对去匿名化工作的综述, 尤其是被动流量分析相关综述进行了比较分析. 2009 年, Edman 等人[21]梳理了高延迟、低延迟匿名通信技术及其攻击方法. 流量分析是主要攻击方法之一. 2010 年, Salo 等人[16]将当时的匿名攻击技术分为 5 类, 基于概率模型、基于出入口选择、基于 AS 级别和全球级别对手的攻击、流量与时序分析攻击和协议脆弱性分析. 2015 年, Erdin 等人[22]将匿名攻击机制分为基于应用程序的攻击和网络级攻击, 前者依靠程序或网页的插件、脚本以及用户的不规范操作实现攻击, 后者则针对匿名机制本身发起攻击. Cambiaso 等人[23]基于攻击目标将去匿名化工作分类为针对客户端的攻击、针对服务器的攻击和针对网络的攻击, 流量分析和时序分析是针对客户端和服务器的攻击. 赵娜等人[24]将匿名通信系统隐藏服务定位技术分为客户端定位技术和服务端定位技术, 介绍了 3 种交叉攻击方法和 14 种网站指纹技术. 以上工作大多从去匿名化角度对主流工作进行分类, 越来越多工作将 Tor 流量分析作为分类的重点之一. 然而, 以上工作重点往往在介绍去匿名化方法概念与基本方法, 缺乏在细分领域的分析与比较.

随着去匿名化工作的发展, 对去匿名化工作的分类也越来越清晰和细致化, 主动与被动方法由于在先置条件、实现方法和攻击目标等存在较大差异, 逐渐成为去匿名化工作综述的重要分类指标. Yang 等人[25]与罗军舟等人[2]在网络流量分析的去匿名化技术中引入两个维度: 主动和被动攻击、单端和端到端攻击. 根据两个维度的交叉, 流量分析去匿名化工作又可以分为 4 类. Evers 等人[26]在文献 [25] 的分类基础上根据方法和目标进行了更细致地分类: 关联攻击 (端到端被动攻击), 拥塞攻击 (端到端主动攻击), 时序攻击 (端到端主动攻击), 指纹攻击 (单端被动攻击), 拒绝服务攻击 (单端主动攻击), 辅助攻击和隐藏服务攻击. 吕博等人[27]把去匿名化工作分为基于流量分析、协议弱点的攻击, 其中流量分析也分为主动和被动攻击. Basyoni 等人[28]从威胁模型角度, 将 Tor 流量分析攻击分为全球规模的对手、捕捉入口流的对手和控制 Tor 中继的对手, 每种对手下又分为主动和被动的攻击. 2021 年, Karunanayake 等人[17]对针对 Tor 的去匿名化工作进行了较为详细的综述, 他们根据攻击条件将现有工作分为基于

关键组件的攻击、基于入口和出口的攻击、基于侧信道的攻击以及混合攻击. 其中每种攻击又按照主动方法和被动方法进行分类. 以上综述工作从主动和被动角度对 Tor 去匿名化工作进行分类, 重点讨论了两种方法的攻击条件、实现方法与优劣.

表 1 Tor 流量分析综述

文献	年份	内容	分类方法	流量识别	网站指纹	流关联
[21]	2009	匿名通信技术以及针对匿名通信技术的攻击方法	高延迟与低延迟匿名通信系统, 以及流量分析等攻击方法	×	√	√
[16]	2010	Tor去匿名化方法综述	基于概率模型、基于出入口选择、基于AS级别和全球级别对手、基于流量与时序分析、基于协议脆弱性的攻击	×	×	√
[22]	2015	匿名通信网络的去匿名化方法综述	基于应用程序的攻击与网络级攻击	×	√	√
[25]	2015	匿名通信网络的去匿名化方法综述	主动与被动攻击、单端和端到端攻击	√	√	√
[26]	2016	Tor去匿名化方法总结	基于主动与被动攻击、单端和端到端攻击扩展成7种类别	√	√	√
[27]	2017	Tor去匿名化方法综述	基于流量分析和协议弱点的攻击, 流量分析包括主被动攻击	×	√	√
[15]	2018	基于机器学习方法的Tor流量分类综述	机器学习方法、流量分类输入、流量分类输出	√	×	×
[18]	2018	混淆流量识别与追踪综述	识别: 深度包检测与机器学习; 追踪: 主动与被动方法	√	×	×
[2]	2019	匿名通信网络的去匿名化方法综述	主动与被动攻击、单端和端到端攻击	×	√	√
[23]	2019	Tor去匿名化方法综述	基于攻击目标分类: 针对客户端、针对服务器、针对网络	√	×	√
[28]	2020	Tor流量分析工作综述	基于威胁模型分类: 全球规模对手、捕捉入口流对手、控制Tor中继对手, 每种对手分为主动和被动方法	√	√	√
[29]	2021	Tor网页指纹识别综述	针对单标签和针对多标签	×	√	×
[17]	2021	Tor去匿名化方法综述	基于攻击前提分类: 掌握关键组件、掌握入口和出口、掌握侧信道信息、混合攻击, 每种前提分为主动和被动攻击	√	√	√
[30]	2021	匿名通信系统流量识别与分类综述	基于匿名工具: Tor、I2P、ZeroNet	√	√	×
[31]	2022	网站指纹识别与防御综述	网站指纹分为: 基于传统、基于机器学习、基于深度学习方法	×	√	×
[24]	2022	隐藏服务定位技术综述	基于攻击目标分类: 客户端定位技术、服务端定位技术	×	√	×
[32]	2023	基于深度学习的网站指纹识别与防御	识别与防御工作	×	√	×

在上述综述中, 被动流量分析技术由于实现成本低、方案简洁有效等原因, 积累了大量工作, 成为 Tor 去匿名化工作的最热门技术之一. 许多工作梳理了被动流量分析研究. Pacheco 等人[14]对使用机器学习进行加密流量分析工作综述, 他们按照机器学习流程对当时工作进行分类和介绍. Aminuddin 等人[15]开展了使用机器学习方法实现 Tor 流量分类研究的综述, 根据使用的机器学习方法, 现有工作分为有监督、半监督、无监督学习; 流量分类的输入可以包括通道信息、流信息和包信息; 输出包括流聚类、应用类型、应用协议、软件以及更细粒度的分类. Wang 等人[30]也梳理了匿名通信流量识别与分类工作. 综述首先回顾了传统流量分类方法: 深度包检测和加密流量识别方法, 然后介绍了 Tor、I2P 和 ZeroNet 的流量分类工作. 姚忠将等人[18]对混淆流量识别与追踪工作进行了综述. 对混淆流量的识别包括基于深度检测包的技术和基于机器学习的技术, 追踪包括被动关联技术和主动关联技术, 其中被动关联技术分为揭露分析、流量分析和指纹技术. 孙学良等人[29]将 Tor 网页指纹工作分为针对单标签和针对多标签的工作, 其中, 针对单标签的网页指纹工作分为基于传统方法、基于机器学习的方法和基于深度

学习的方法. 邹鸿程等人[31]综述了网站指纹攻击与防御, 其中网站指纹攻击分为基于相似度、基于机器学习、基于深度学习的方法. 2023 年, Liu 等人[32]梳理了将深度学习用于网站指纹攻击和防御的工作. Aminuddin 等人[15]则从分析流程角度综述了针对 Tor 的网站指纹工作. 以上许多综述对去匿名化方法进行了梳理, 但这些工作往往涉及多种攻击方法, 重点在对攻击方法的介绍, 对具体方法的比较和分析有限; 另外一些综述对具体一种技术, 例如隐藏服务揭露技术或网站指纹技术进行梳理.

Tor 被动流量分析具有成本低、去匿名化程度高、易于部署等特点, 历经近 20 年的发展, 已成为最重要的去匿名化技术. 本文从被动流量分析角度对 Tor 去匿名化的现有工作进行梳理和总结, 与之前工作相比, 本文按照分析流程对工作进行梳理, 讨论了许多研究细节, 对比了 研究流程中不同方法的优缺点. 此外, 本文专注于被动分析角度, 除了流量分类和网站指纹工作以外, 还收集了大量网站指纹攻防和流关联研究. 在此基础上, 本文扩展了对被动流量分析工作的讨论, 分析了许多被动流量分析工作中的共性问题, 给出了可能解决的方向.

3 Tor 流量分类研究进展

Tor 流量分类研究主要通过分析流量知晓用户或服务器是否在使用 Tor 以及具体的服务类型. 从最初的基于端口方法[33]、深度包检测方法[34]发展到机器学习以及深度学习方法, Tor 分类研究经历了长足的发展. 本节将流量分类过程分为数据采集、特征提取、流量分类 3 个阶段, 并将主要工作总结在表 2.

表 2　主要的 Tor 流量分类工作

文献	年份	特征	模型	数据集规模	分类目标
AlSabah等人[35]	2012	统计特征	NB, BN, 决策树	899条	3种流量类型
Cai等人[36]	2012	构造相似性特征	SVM, HMM	40×800	网站
何高峰等人[37]	2013	TLS连接和报文长度分布特征	SVM	10G	Tor
Wang等人[38]	2014	统计特征, 构造特征	kNN	100×90+9000	网页
Panchenko等人[39]	2016	构造长度累积和特征	kNN, SVM	100×90+9000	网页
Hayes等人[40]	2016	统计特征, 构造特征	kNN, SVM	Tor55×100+隐藏服务30×80+Tor100000+普通55×30+20000	网页
Wang等人[41]	2016	统计特征	kNN, SVM, NB	Tor55×100+隐藏服务30×80+Tor100000+普通55×30+20000	网页
Cuzzocrea等人[42]	2017	统计特征	6种经典机器学习	22G	8种流量类型
Shahbar等人[43]	2018	统计特征	BN, NB, C45, RF	140万+	21种应用
Rimmer等人[44]	2018	深度学习自动特征	SDAE, CNN, LSTM	900×2500+800000	网页
Sirinam等人[45]	2018	深度学习自动提取	CNN	95×1000+40716	网页
Sirinam等人[46]	2019	构造距离特征	多种深度学习	多个公开数据集	网页
Bhat等人[47]	2019	自动提取和统计特征	CNN	900×2500+500000	网页
Montieri等人[48]	2020	序列特征, 统计特征	BN, NB, C45, RF	140万+	21种应用
Montieri等人[49]	2020	序列特征, 统计特征	BN, NB, C45, RF	140万+	21种应用
Bovenzi等人[50]	2020	序列特征, 统计特征	BN, NB, C45, RF	140万+	21种应用
Hu等人[51]	2020	统计特征	机器学习, 深度学习	16387条	8种流量类型
Wang[52]	2020	多种工作复现	多种工作复现	100×100+10000	网页
Singh等人[53]	2021	时间相关统计特征	多种深度学习	141530条	多种流量类型
Lin等人[54]	2021	深度学习自动特征	CNN, LSTM	16387条	8种流量类型
Zhao等人[55]	2021	深度学习图特征	GCN	36193条	10种匿名服务
Xu等人[56]	2022	长度序列	机器学习, 深度学习	多个公开数据集	多种流量类型
Lan等人[57]	2022	序列特征, 统计特征, 构造特征	深度学习	67834条	多种流量类型
Yin等人[58]	2022	统计特征	XGBoost	50×100+2500	网页
Wang等人[59]	2022	数据包序列	CNN	900×2500+400000	网页

表 2 主要的 Tor 流量分类工作 (续)

文献	年份	特征	模型	数据集规模	分类目标
Cherubin等人[60]	2022	构造距离特征	kNN	100 000	网页
Deng等人[61]	2023	滑动窗口深度学习自动提取	自注意力机制的深度学习	多标签、多版本、多种类数据	多标签网页
Mathews等人[62]	2023	序列特征, 统计特征, 构造特征	机器学习, 深度学习	95×200+19000	网页
Karunanayake等人[63]	2023	深度学习自动提取	GNN	2×15×50+3×15×15×50	DAPPs
Zhou等人[64]	2023	深度学习自动提取	Transformer	95×1000+40716	网页
Bahramali等人[65]	2023	构造特征	深度学习	文献[44]的数据集, 225×550+10000	网页
Jin等人[66]	2023	深度学习自动提取	Transformer	50×200+9900	多标签网页

注: $n\times m+x$是网页数据集格式, n代表监控集网站数量, m代表每个网站收集的数量, x代表非监控集网站

在数据采集部分, 梳理了现有工作生成与收集 Tor 流量的主要方法与目前公开的 Tor 数据集; 在特征提取部分, 本节讨论了现有工作中使用的特征以及使用的特征预处理与特征选择方法; 在流量分类部分, 本节梳理了现有工作中主要使用的分类算法, 包括机器学习算法和深度学习算法等; 最后, 本节整理了目前 Tor 流量分类工作的最新动态.

3.1 单端流量采集

在数据采集阶段, 研究者收集 Tor 流量用于分析. 采集需要考虑整个测试平台的软件设置、硬件设施以及具体配置. 当前的研究主要采用 3 种方式获得数据集: 仿真环境采集、真实网络采集以及使用公开数据集.

3.1.1 仿真环境采集

一些工作通过搭建仿真 Tor 网络环境进行 Tor 网络流量分析. 2010 年, 为了模拟真实 Tor 网络中的带宽、延迟、吞吐量等, Chakravarty 等人[67]在 DETERlab 网络防御技术实验研究实验室上仿真了一个由十几个模拟终端主机组成的小型 Tor 网络. 2011 年, Bauer 等人发布了 ExperimenTor[68], 一个大型 Tor 网络仿真工具包和测试平台. 基于该平台, 研究者可以在若干台主机上模拟 Tor 网络拓扑并生成客户端流量. 2012 年, Jansen 等人发布了 Shadow[69], 这是一个离散事件模拟器, 可以生成多个虚拟节点, 然后在虚拟节点上运行真实 Tor 程序并模拟网络连接. 与 ExperimenTor 平台相比, Shadow 更像是一个仿真黑盒, 它仅在一台主机上作为一个应用程序运行, 占用了更少的资源, 因此研究者可以在 Shadow 上模拟更大规模的网络. 2018 年, 为了更深入地评估 Tor 流量的安全性, Jansen 等人[70]在 Shadow 中开发了流量生成工具 TGEN. TGEN 可以模拟真实客户端, 在仿真环境中生成灵活而复杂的流量模式, 帮助研究者自由地构建流量模型. 他们将在 Shadow 中仿真的流量模型与在真实环境中收集到的流量模型对比, 结果表明在 Shadow 中测量流量的多个网络特性都十分接近真实环境. 仿真环境的优势在于可以模拟复杂的网络环境与大规模的网络场景.

3.1.2 真实网络采集

更多工作自己手动访问并捕捉 Tor 流量. ALSabah 等人[35]收集了 3 种流量: 网页流量、P2P 流量和流媒体流量. 他们设置 3 种客户端, 网页客户端: 使用 Firefox 浏览器中的 iMacros 插件自动访问 Alexa 排名前 100 的网站并收集流量. 为了使流量更真实, 客户端在页面加载完成后会等待一个随机的时间以模拟用户浏览和思考, 再进行下一个动作. BitTorrent 客户端: 收集了一些 torrent 并使用 Vuze BitTorrent 软件在 Tor 网络中下载. 流媒体客户端: 同样使用 Firefox iMacros 插件, 使用关键字搜索视频, 然后打开视频随机观看 1–5 min. 为了消除背景流量的影响, 他们实现了一种新的中继命令单元“MeasureMe”. 该单元会提醒入口节点何时开始记录流量以及流量所属类型. 何高峰等人[37]在校园网络中生成并收集 Tor 流量. 他们拥有 60 台主机通过一台交换机连接网络, 其中, 10 台主机运行 Tor 访问 Web、P2P、FTP 和 IM 等服务. 这些 Tor 流量和其余 50 台主机产生的普通流量通过交换机的端口镜像功能被转发至一台检测主机, 由此获得数据集. Jia 等人[71]使用 wireshark 捕捉网页、视频、收发邮件流量. Cuzzocrea 等人[42]在 Whonix 系统中收集 Tor 流量. 这是一个轻量的 Tor 操作系统, 可以将所有访问经过 Tor 网络传输. 他们使用 ISCXFlowMeter 生成流量, 使用 wireshark 和 tcpdump 捕捉流量. Rimmer 等人[72]基于 Tbselenium

编写分布式爬虫实现 Tor 浏览器的自动访问. 这是基于 selenium 库修改的 Python 库, 可实现 Tor 浏览器自动化功能. 与仿真环境相比, 从真实网络中收集流量可以获得最接近实际的数据集.

3.1.3 公开数据集

一些研究者在公布自己的工作时还提供了源数据集, 这保障了研究工作的可复现性, 也给了其他工作者在统一的数据集上对比不同方法的机会. 表 3 列出了目前最常用的公开 Tor 流量数据集.

表 3 Tor 公开数据集

数据集名称	数据集内容	数据集形式	Tor数据集规模
ISCXTor2016[73]	7种类型Tor与常规流量	原始流量与特征文件	8044 Tor+59790 nonTor
Anon17[74]	匿名工具应用流量	92个特征的特征文件	5283 Tor+252 Tor应用程序+355591混淆流量
DeepCorr-2018[75]	Tor入口与出口流量对	流对特征文件	1000通道×50个网站+500混淆流量
TCUB2020[51]	匿名通信网络流量	Pcap与特征文件	8种类型, 共8 632条Tor流量
CIC-Darknet2020[76]	ISCXTor2016与ISCXVPN2016合并	Pcap与特征文件	8044 Tor +59790 nonTor
Tik-Tok[77]	未防御、防御、隐藏服务	序列特征文件	95000+95000+90000+75000+41426
SJTU-AN21[55]	匿名工具流量	92个特征的特征文件	5种类型, 共1 966条Tor流量
CMU-SYNTRAFFIC-2022[76]	使用生成对抗网络等方法生成样本并与CIC-Darknet合并	64个特征的特征文件	共2 650 467个样本
Wang14[38]	访问网站流量	原始流量与特征文件	100×90+9000
Rimmer17[44]	访问网站流量	序列特征文件	900×2500+400000
Sirinam18[45]	访问网站流量	序列特征文件	95×1000+9000
Wang20[52]	访问网站流量	流特征文件	100×200+80000
Multitab-WF-Datasets[61]	多标签、多种类网页	特征文件	2–5多标签网页, 共超过57万

3.2 流特征提取

特征提取从捕获的流量中计算出不同的指标, 用来反应流量的某些特定属性, 是流量分析最重要的步骤. 研究者通常先选择需要提取的特征列表, 再从每一条流中计算出所有特征, 得到一个向量或数组, 即特征向量. 然后, 研究者需要进行数据清洗, 以得到高效、纯净的原始数据集. 最后, 特征集输入到分类算法或模型, 用于训练或测试分类器.

3.2.1 特征分类

本文从流量分析流程出发, 考虑了威胁模型、特征提取条件等因素, 将 Tor 流量分类工作的特征分为 4 个主要类别: 流序列特征、统计特征、构造特征以及深度学习自动提取特征.

3.2.1.1 流序列特征

流序列特征关注数据包交互顺序, 提取数据包时间顺序、数据包包长顺序等特征. 在 Tor 通道构建与 Tor 数据传输过程中, Tor 客户端与服务器独特的通信交互规则会影响数据包的大小以及在时间上的分布, 因此可以用数据包交互序列表征 Tor 通信事件, 从而实现 Tor 流量的识别与分类.

Montieri 等人[48]研究了匿名工具流量早期分类的问题. 作者使用流的前 20 个数据包的有效载荷长度和时间间隔作为序列特征输入, 评估是否可以动态地识别到匿名工具流量、流量类型及其具体应用. 实验结果显示, 仅需 7 个数据包就可以实现匿名工具应用的识别. Xu 等人[56]也使用了序列特征进行 Tor 和 VPN 流量的识别与分类. 他们从数据包序列中提取数据包长度, 将上行数据包设为负数, 下行数据包设为正数, 由此获得一维流量路径. 为了更深入地探究客户端与服务器的交互, 他们还将该一维流量路径拆解为上行链路序列和下行链路序列, 上行链路序列流量路径中所有下行数据包长度被置零, 下行链路序列流量路径则相反. 这些序列特征在 40–200 个数据包上都可以获得较好的识别性能. Lan 等人[57]同时使用了序列特征与统计特征进行 Tor 和 VPN 流量识别, 其中, 使用的序列特征是流的前若干个数据包长度序列. 作者探究了数据包数、字节数以及输入的数据包长度序列大小

对识别结果的影响. 他们的实验表明, 数据包、字节数、序列大小对识别结果呈现类似的趋势: 增加以上特征数量可以提升识别结果, 但来到一定数量以后, 识别结果达到稳定, 不再有提升. 在他们的实验中, 使识别结果稳定的数据包长度序列大小为 100.

由于序列特征提取简单, 一般只使用少量数据包, 因此可以用于实现流量的在线识别. 流量的在线识别被广泛用于 QoS 服务, 异常识别等[78–80]. Tor 的在线识别可以被用于实现 Tor 服务的 QoS[35], 也可能被用于网络审查. 序列特征是 Tor 流量分析的一类关键特征, 它易于提取, 可解释性强. 但同时所含信息量较小, 在多个工作的对比中, 都证明了序列特征的识别结果低于其他更复杂的特征[48,81].

3.2.1.2 流统计特征

流统计特征通过统计并计算流的特定属性实现对流的表征, 包括方向、时间、长度、包头信息等, 是流量分析中最主要的特征. 大多数流量分析工作都使用了统计特征[82]. 与序列特征只需要前若干个数据包不同, 统计特征中的许多特征需要等待流结束才可以得到, 例如流时长、流平均包长等, 由于可以从完整的流中提取信息, 因此统计特征往往比序列特征表现效果更好.

何高峰等人[37]使用 Tor TLS 流量中抽取的密码套件与数字证书指纹和 Tor 报文长度分布特征实现 Tor 流量的识别. Wang 等人[38]使用了大量统计特征进行 Tor 网站指纹识别研究. 除了总传输大小、总传输时间等常规特征以外, 作者还提取了每条流独特的数据包长度、数据包排序、传出数据包集中度等特征. 作者表明, 影响网页流量的因素很多: 网络条件、随机的广告和内容、网页数据更新、网页资源传输顺序和客户端配置等. 因此, 需要构建多样化的特征, 从每种模式中提取代表性元素, 从不同角度评估复杂的网页流量. 虽然使用了相似的度量和分类算法, 但他们的方法优于文献 [36] 等. 许多工作[48–50]对 Anon17 数据集[74]开展了匿名通信流量的识别与分类工作, 该数据集使用 Tranalyzer[83]提取了 91 个流特征, 删除 ICMP、VLAN 等对流量识别无用的特征后, 最终使用了 81 个统计特征. 他们的实验表明, 大约使用 25 个最好的特征就可以使识别效果达到最好, 在这之后, 增加额外的特征并不会使识别效果获得明显提升.

在特征选择方面, Moore 等人[82]对流量特征提取进行了系统的分析, 他们关注 TCP 协议, 梳理了 248 个最重要的统计特征. 后续 Tor 流量分析的许多工作都使用了该特征集或者子集, 如 Singh 等人[53]将时间相关的特征转化为三维图像, 再用深度学习模型对 Tor 流量进行检测. Hu 等人[51]使用流时长、包计数、长度、间隔、窗口大小、ACK 计数等 26 个特征实现匿名网络流量的分层识别: 最高层首先识别是否是匿名网络, 包括 Tor、I2P、ZeroNet 和 Freenet 这 4 种匿名网络; 第 2 层识别具体流量类型, 包括 8 种流量类型; 最后识别某种类型中具体的应用. Hayes 等人[40]使用包数量、进出包百分比、间隔时间等特征实现网站指纹攻击. Ma 等人[84]使用统计特征实现多流复用的网站指纹攻击等.

统计特征可以分为以下 4 类: 一是方向特征, 包括客户端→服务端、服务端→客户端与双向 3 种情况. 研究者可以根据研究内容选择研究的流方向进行特征统计. 二是时间统计特征, 指流中关于时间相关的统计特征, 包括流时长、分组或数据包间隔时间、流活跃与空闲时间等, 还可以计算对应最大值、最小值、平均值等衍生特征. 三是长度统计特征, 指流中分组或数据包长度的统计特征, 包括字节数、包长度、特定包长频率及其计算出的衍生特征等. 四是计数统计特征, 指对流中一些属性计数统计得到的特征, 包括数据包数量、ACK/SYN/FIN 报文数量、重传数据包数量、RTT (往返时间) 计数、TTL (生存时间) 计数等.

3.2.1.3 构造特征

一些工作提出构造特征, 旨在通过挖掘流量的深层特性实现流量识别. 基于规则的特征是依据特定规则提取出的特征, 一般高度依赖专家知识. 信息熵是香农于 1948 年提出的描述事件发生的不确定性的概念. 高度加密流量的熵明显高于普通流量. 因此熵特征很早被用于加密流量的识别. Jia 等人[71]使用数据包长度熵作为特征进行 Tor 流量的识别. 除此之外, 作者还使用了特殊字节长度数据包频率、平均数据包间隔等特征. Panchenko 等人[39]发现不同网站在结构、内容、执行脚本等元素上的差异会显著影响网站访问的负载行为. 例如, 包含各种主题、视频和嵌入式动态对象 (图片) 的网站比较为固定的网站在数据包序列上呈现出更大的变化. 因此, 作者提出 CUMUL, 长度序列累积特征, 作为网站指纹进行识别. 他们记录流序列中包的大小和方向, 下一个包为接收包时,

序列累积值加上包大小, 下一个包为发送包时, 累计值减去包大小, 由此得到一系列长度序列累积特征 (CUMUL). Oh 等人[85]也使用了长度序列累积特征, 但存在的问题是: 不同的网站访问序列总长不同, 他们使用的分类器输入的向量却是相同维度的, 因此他们倒转了流序列, 这样只保留序列末尾部分的长度累积特征, 而末尾部分的特征包含了序列前部的信息. 他们的实验表明倒转的长度累积特征有着更好的准确性. Shahbar 等人[86]也从数据包大小序列中构造包动量特征实现匿名工具流量的识别. 作者考虑了流中数据包的方向、大小与到达时间, 提取包动量特征. 他将数据包到达的时间间隔与方向视为该数据包的速度, 则该速度与数据包的乘积即为数据包的动量. 除了包动量特征以外, 作者还使用了流最大包、次最大包及其频率等特征. Karunanayake 等人[63]认为每次访问行为会受到多种因素影响, 例如网络波动、电路构造等过程, 传统特征提取方法无法体现这些影响, 因此作者对流量出入序列的突发进行如下操作中的一种: 改变入包突发、插入出包、合并入包突发, 以此来构造流量特征.

除了从流中构造新的特征, 还有研究者通过主动干扰用户访问过程影响原有特征. 当使用 Tor 时, 用户的网站访问请求被全部代理到匿名通道中, Yang 等人[87]认为 Tor 通道构建流量以及网站各种 Web 对象的访问交互流量混合在一起会极大地影响网站识别的准确性. 因此, 他们提出了一种主动流量识别方法, 先分析 Tor 通道构建过程以及数据单元传输特性, 推断出第 1 个 HTTP 请求的位置; 接着, 他们延迟该 HTTP 请求并记录流数据; 最后从流数据中提取特征完成网站识别. 延迟 HTTP 请求不会破坏用户的网站访问, 只让用户感受到些许的延迟, 但分割了 Tor 流中不同交互过程, 提升了识别准确率.

3.2.1.4 深度学习自动提取特征

以上特征大多基于直觉和专家知识并需要手动提取, 一些工作引入深度学习, 自动化提取流量特征. 深度学习中的流量识别输入可以是经过预处理的流量原始数据, 也可以是经过提取的简单特征.

Rimmer 等人[44]开展了使用深度学习自动提取流量特征进行 Tor 网站指纹识别的研究. 他们将 Tor 单元的时间信息、方向和数据包大小作为原始序列输入深度学习模型, 由深度学习模型进行特征提取, 并实现网页的识别与分类, 根据所使用模型的不同, 他们分别使用了流的前 150、前 3000 和前 5000 个 Tor 单元. Lin 等人[54]使用包原始数据, 基于不同深度学习模型提取时空特征实现 Tor 流量分类. 他们将前 15 个包的原始数据按照字节读取, 并通过填充 0x00 将每个数据包扩充到 1500 字节 (最大传输单元), 这个过程称为矢量化. 矢量化的数据经过规范化等操作即为深度学习模型的输入. 作者使用了 CNN 模型提取抽象的空间特征, 使用 RNN 提取抽象的时间特征. Lan 等人[57]引入自注意力深度学习方法研究 Tor 流量分类. 作者选择了侧信道特征与内容特征. 在他们的工作中, 侧信道特征是统计特征与序列特征的合集, 而内容特征是选择了前 N 个数据包, 每个数据包保留 M 字节作为输入. 对于流长度, 如果流长度超过 N 个包则阶段, 如果少于 N 个包则用 0 填充, 数据包字节同理. 在他们的实验中, 30 个数据包和每个数据包取 256 个字节可以取得较好的识别性能. Zhou 等人[64]也引入了自注意力机制, 使用 Transformer 从流量序列中自动提取特征实现网站指纹攻击. 作者认为, 由于自注意力机制可以关注流量序列中不同部分之间的相关性, 可以提取更深层和稳定的特征, 因此具备更高的识别性能.

深度学习自动提取特征的优点在于不依赖专家知识, 一些工作[44]认为深度学习可以提取更深层次的流量信息, 使得提取的特征更普遍和稳定, 所以在流量识别领域具有更大的潜力.

3.2.2 预处理

由于网络波动、流量采集工具错误、Tor 网址失效、Tor 连接失败等原因, 研究人员采集的流量中不可避免地包含一些无效流和噪声. 因此需要对原始流量或原始数据集进行预处理, 以去除无效流量对数据集的影响. 另一方面, 使用过多的特征可能导致计算资源的浪费和识别效率的下降 (维度诅咒), 少数特征甚至会对识别效果产生负面影响, 因此需要进行特征选择.

● 数据预处理: (1) 去除无效流: Jansen 等人[70]抓取了 5000 个洋葱网站的 Tor 流量. 在搜集洋葱服务网址地址时, 作者使用 Torsocks 和 cURL 工具测试洋葱站点的可访问性并删除了无效的地址. 在捕获 Tor 流量时, 作者删除访问失败的流, 以此获得有效 Tor 网站流量. 部分工作[48]在开展早期包识别研究时删除了前 20 个数据包负载为 0 的实例以保证输入数据的有效性. (2) 平衡数据集: 使用抽样下采样, 从样本数量多的类中选择部分样本, 使得不同类的样本数量接近, 实现不平衡的数据集到平衡数据集的转化[48]; Xu 等人[56]使用 ISCX-Tor2016 数据集开展 Tor

流量分类研究, 然而该数据集中 Tor 流量只有几十个样本, 存在明显的样本不均衡问题. 作者采用 3 种抽样方法从 Tor 流中提取多条流: 随机抽样、固定步长抽样和混合抽样 (结合时间间隔和数据包长度规定抽样点). Lin 等人[54]同时使用流量分割和采样获得平衡数据集; 还有方法使用 BalanceCascade 算法构造平衡训练集[58]. 该方法使用上一轮训练的分类器筛选出容易识别的多数类并删除这些样本, 使得下一轮训练保留容易分类错误的样本, 以此保证平衡训练集的质量.

● 特征选择: Shen 等人[88]分析了 Tor 等加密流量的特征选择方法. 一般来说, 特征选择的主要目的是筛选有效特征, 提升模型泛化能力, 防止过拟合. 对特征的评估需要综合考虑特征贡献度与特征开销. 特征贡献度衡量方法主要有卡方检验、TF-IDF 和基于模型的排名等, 特征开销主要需要考虑提取特征的时间复杂度与空间复杂度等. Oh 等人[85]使用 Kruskal-Wallis 检验 (H 检验) 评估特征重要性. 除此之外, 他们还对特征开销进行了实验. 首先, 他们在 4 组特征中选取不同数量的特征进行模型训练和测试, 然后评估不同特征数量下的识别准确率与训练时间开销. 实验表明, 增加特征数量在使训练时长不断上升的同时也会使准确率不断上升. Yin 等人[58]基于模型排名对特征进行排名. 他们使用了文献 [40] 中的特征, 并扩充了文献 [41] 中使用的特征, 例如将某个数据包周围 5 个包时间间隔特征扩充到数据包周围的 50, 45,…, 10, 5 个包时间间隔特征等, 共获得 302 个特征. 然后, 他们使用递归特征消除技术[89]来筛选最重要的特征.

数据预处理与特征选择是一个可选步骤, 通过优化数据集和选择最相关的特征来提升分类的准确性、泛化性并降低计算开销[88,90]. 在大部分研究中, 这一步骤最主要的作用是提升分类准确性, 少数工作在这一步骤解决类不平衡问题和计算开销问题.

3.3 分类算法

经过数据预处理与特征选择, 研究者可以选择分类算法进行 Tor 流量分类的训练和测试. 本文基于当前工作使用的经典机器学习算法和深度学习算法对文献进行分类. 进一步, 根据具体使用的分类算法梳理相关工作.

3.3.1 经典机器学习

● 贝叶斯分类器: 是基于贝叶斯理论的经典机器学习分类算法, 它根据概率和误判损失对事件实施决策. 贝叶斯分类器中最常用的方法有朴素贝叶斯 (naïve Bayes, NB)、多项式朴素贝叶斯 (multinomial Naïve Bayes, MNB) 和贝叶斯网络 (Bayesian networks, BNs) 等. AlSabah 等人[35]提取 Tor 电路寿命、传输数据、单元间隔时间与最近发送的单元数等特征, 使用 NB 分类器和 BNs 分类器对网页流量、P2P 流量和流媒体流量进行在线分类和离线分类. 结果显示, 在在线分类中, BNs 可以实现 97.8% 的准确率, 而 NB 的准确率不超过 31%. 在离线分类中, BNs 的准确率可以达到 85%. 一些工作[48–50]研究了匿名工具的流量识别与分类问题. 作者使用了包括时序特征和统计特征在内的 74 个特征, 应用包括 NB、BNs 在内的分类器, 研究了特征数量、时间特征与非时间特征、时序特征对分类结果的影响. NB 分类器在识别 21 种匿名应用程序流量方面获得了 62.97% 的准确率, BNs 则是 71.39%. 然而, 作者发现在使用少量特征时, 贝叶斯分类器可以获得较高性能, 在增加使用的特征时, NB 和 BNs 会有准确率的下降.

综合来看, 贝叶斯分类器分类效率稳定, 对缺失和噪声适应性强, 在数据分析领域具有极为重要的地位. 然而, NB 分类假设所有属性具有独立性, 这使得研究者使用大量特征或相关性强的特征时准确率大大下降.

● 决策树: 通过构造一个树结构, 每个树节点根据特征属性做出判断最终到达叶节点, 判断样本类别, 是最常见的机器学习算法之一. AlSabah 等人[35]也使用了函数树 (functional tree, FT)[91]和逻辑模型树 (logistic model tree, LMT)[92]两种决策树算法进行流量分类, 实现了超过 90% 的准确率. 实验表明, 两种决策树的识别准确率接近, 且决策树算法明显优于 BNs 等算法. Cuzzocrea 等人[42]同样使用了决策树算法识别 Tor 流量并进行流量类型分类. 首先使用 Mann-Whitney 检验和 Kolmogorov-Smirnov 检验验证 Tor 流量与普通流量特征分布存在显著差异, 然后使用机器学习算法分类. 作者使用的决策树算法包括 J48, J48Consolidated 和 REPTree. Jia 等人[71]提出了一种改进的决策树算法, 称为 Tor-IDT, 利用信息增益选择划分属性点, 利用信息增益率选择划分属性, 他们的结果表明, 该算法对 Tor 的识别率超过 99%. Shahbar 等人[43]进行了匿名工具流量识别与分类研究. 作者使用 C4.5 决策树算法

实现了 97.2% 的准确率.

决策树可以被直观地理解和解释, 构造简单, 分类性能出色. 然而, 当样本数量较少, 样本噪声较多或者创建较为复杂的树时, 决策树会产生过拟合问题, 需要设置剪枝、规定叶节点样本数等策略.

● 随机森林: 随机森林是以决策树为基学习器, 基于 Bagging 集成学习方法构建的分类器算法. Hayes 等人[40]训练了一个随机森林模型, 将每棵树的叶节点对属性的判断导出为一个特征向量, 使用汉明距离评估指纹距离, 实现网站指纹识别. Xu 等人[58]提出一种针对多标签页的网站指纹识别方法. 该方法从流中提取包括统计特征、序列特征、相似性特征在内的 452 个特征, 使用随机森林分类器进行分类. 在针对 50 个 Tor 网站的识别中, 使用 48 个特征获得了 68.72% 的真阳性率. Montieri 等人也将随机森林算法应用在匿名工具流量识别与分类研究中, 他们发现在不同级别的流量分类任务[48]、对匿名流量的多级分类任务[49]和数据与模型并行的流量分类任务[50]中, 随机森林都优于贝叶斯分类器和决策树, 几乎在所有任务中都获得了最高的分类性能.

综合来看, 随机森林在 Tor 流量分类和网站指纹识别中都具有极高的真阳性率和准确率, 在特征评估、对抗噪声、不平衡数据集方面也有不俗的表现, 优于大多数其他算法. 随机森林算法既有决策树高效、准确的优点, 也有较好的泛化能力和鲁棒性.

● Boosting: 是另外一种常用的集成学习方法, 通过不断调整样本分布, 增加后续分类器在训练时对错误样本的关注, 逐渐在多轮训练中从弱学习器中生成强学习器. Hu 等人[51]研究了 4 种类型的暗网、8 种流量类型共 25 种暗网行为的流量分类问题. 作者提取了 26 个时间相关的特征, 使用包括 GBDT、XGBoost、LightGBM 等多个 Boosting 集成分类模型在内的多个算法对流量进行识别和分类. 实验表明, Boosting 算法取得了比决策树、随机森林、多层感知机等分类模型更好的性能. Zhao 等人[55]在匿名流量识别中使用 LightGBM 对特征进行排序和选择. LightGBM 算法使用基于梯度的单侧采样算法减少样本处理时间, 而在迭代训练过程中, 具有大梯度的样本往往意味着更多的信息增益, 因此可以通过保留大梯度的样本选择出最重要的特征. Yin 等人[58]改进了文献 [93] 中的方法, 对多标签页的网站指纹识别问题开展了深入研究. 首先训练一个 XGBoost 二元分类器, 用于识别一条流的每个传出数据包, 判断是否是流分割点 (新流开始数据包); 然后再次使用一个 XGBoost 多分类器, 用于识别网站. 实验表明, XGBoost 在识别流分割点任务中取得了 84.6% 的准确率, 在网站识别任务中取得了 94.7% 的准确率, 优于 k-FP[40]、CUMUL[39]、DF[44]等方法.

Boosting 算法能对特定的数据分布进行学习, 通过重赋权、重采样等方法, 可以从多个性能较低、泛化能力弱的分类器中构造出强分类器.

● k 近邻 (k-nearest neighbor, kNN): kNN 通过度量样本在特征上与训练样本的距离进行分类. Wang 等人[38]提取了统计特征、独特数据包长度、传出数据包集中度、流量突发等接近 4 000 个特征 (其中 3 000 个是独特数据包长度特征), 使用 kNN 算法对 100 个监控网页实现了 85% 的 TP 率和 0.6% 的 FP 率. 为了提升识别率, 尤其是对应用了网站指纹防御方法流量的识别率, 作者提出了特征权重调整方法. Wang 等人[41]还研究了多页面访问的网站指纹识别问题. 作者首先使用 kNN 和 LF-kNN 方法判断一条流是否需要进行流分割 (是否是多页面访问), 然后用一个 kNN 机器学习模型判断分割点, 最后实现网站指纹识别. Jansen 等人[70]针对 Tor 的中间中继节点流量开展 Tor 流量分析研究. 在使用的算法上, 作者复现了 Wang-kNN, k-FP 中的 kNN 方法. 对于 10 个网站实现了 95% 的准确率, 50 个网站 85% 的准确率以及 100 个网站的 68% 准确率. Cherubin 等人[60]应用 Triplet 指纹方法[46]从网站流量序列中生成特征向量, 使用 kNN 算法识别网站. Triplet 指纹方法基于 *N*-shot 学习, 计算同类网站若干个样本的平均嵌入向量 (mean embedded vector) 作为特征向量. 作者的方法对 5 个网站可以实现 95% 以上的准确率, 但在 25 个网站时准确率降到 80%.

kNN 算法思想简单, 训练快捷, 是一种常用的监督学习方法. 由于网站访问过程中相对固定的交互模式, kNN 与 SVM 算法常被用于开展网站指纹识别研究, 且在多种方法中表现良好. 然而作为一种慵懒学习方法, kNN 计算量大, 预测时间往往较长, 且存在维度爆炸的障碍, 因此更依赖于研究者提出高效的特征.

● 支持向量机 (support vector machine, SVM): SVM 通过在样本集中寻找一个超平面将不同类别划分开实现分类. Cai 等人[36]构建了一个网站指纹识别模型, 可以识别客户端是否在访问特定网站, 以及识别多个网页是否来

自同一个网站. 作者将流量序列转换为字符串, 并计算字符串间的 Damerau-Levenshtein 距离[94], 使用 SVM 实现网页分类. 即便使用当时的几种 Tor 流量防御方法[95,96]对流量进行整形和混淆, 该方法依然能够识别出 80% 的网页. 何高峰等人[37]基于报文长度分布特征, 使用 SVM 实现 Tor 流量识别. 作者首先分析并统计 Tor 流量中报文长度分布, 观察到若干特定报文长度, 在识别过程中, 使用特定报文长度过滤非 Tor 流量, 再将流报文长度出现频率由高到低排序, 由支持向量机判断函数判断是否为 Tor 流量, 该方法实现了 91% 的识别率. 在 Panchenko 等人[39]的网站指纹研究中, 基于 CUMUL 特征的 SVM 分类器取得了 91.38%, 比使用了 3 736 个传统特征的 kNN 算法更高 (90.86%). Oh 等人[85]也使用了包括反转后的 CUMUL 特征、统计特征在内的 247 个特征开展网站指纹识别研究. 结果显示 SVM 算法的准确率也明显优于 kNN.

SVM 算法具有优秀的泛化能力, 且可以抵御维度爆炸, 在某些流量识别任务中具有良好的表现. 与其他算法相比, SVM 似乎在统计特征等传统特征上表现平淡, 却在构造特征上有更好的表现. 这可能是因为构造特征是基于专家知识对数据的一次信息萃取, 而 SVM 恰好更适合基于少数支持向量决定最终结果.

3.3.2 深度学习

● 堆叠降噪自编码器: 自编码器 (autoencoder) 是一套用于数据降维或特征提取的算法, 通常使用一个神经网络将输入重建为一组更具信息性的特征向量. 降噪自编码器 (denosing autoencoder, DAE) 在自编码器基础上以一定概率分布将输入向量的某些维度擦除, 在去除噪声的同时增加了模型的泛化性能. 堆叠降噪自编码器 (stacked denoised autoencoder, SDAE) 则是将多个 DAE 堆叠在一起, 更好地捕捉数据语义信息. Rimmer 等人[44]提取网页序列的前 5 000 个数据包的时间、方向和长度信息, 使用 SDAE 对 200 个目标网站实现了 80.25% 的真阳性率, 超越了使用 kNN 和 SVM 的 CUMUL 方法. 同时, SDAE 具有强大的泛化性, 可以抵御概念漂移问题, 在前 10 天的时间里下降的准确率很小, 在两个月的时间里, 模型的准确性下降了 22%. 作者认为这表明深度学习方法比传统机器学习算法更具潜力.

● 卷积神经网络 (convolutional neural network, CNN): 是一种前馈神经网络, 能够从具有卷积结构的数据中提取特征. Rimmer 等人[44]也使用了 CNN 进行 Tor 网站的识别. 作者使用了 3 000 个数据包, 对仅包含 100 个网站的封闭数据集获得了 96.66% 的准确率, 即便加入背景流量对比, CNN 仍然对 200 个网站获得了 80.11% 的真阳性率. Sirinam 等人[45]研究了针对网站指纹防御的网站指纹识别问题. 作者只考虑流序列的数据包方向, 将前 5000–7000 个数据包序列简化为 [−1, +1] 的值序列作为输入. 为了提取深层特征, 该方法在池化之前应用了 2 层连续的卷积, 为了防止过拟合, 在每个卷积层后进行了批处理标准化 (batch normalization, BN), 在池化后进行了 Dropout 处理. 实验表明, 该方法不但能对 95 个网站实现 98.3% 的准确率, 高于 SDAE[97], AWF[44], kNN[38], CUMUL[39], k-FP[40]等经典方法, 还能成功地识别几种著名的网站指纹防御方法流量. Lin 等人[54]使用 CNN 从流中提取抽象的空间特征, 开展 Tor 和 VPN 流量识别与分类研究. 作者读取流序列的每个数据包的原始字节, 用 0 填充到 1 500 字节长度 (最大传输单元, MTU), 经过规范化后将数据输入深度学习模型进行分类. 由于 CNN 具有局部相关性共享机制, 作者认为 CNN 可以从流中捕获特定模式, 提取抽象的长度特征.

Chen 等人[98]提出了基于小样本的网站指纹攻击方法. 针对以往方法需要大量样本训练模型, 难以频繁变化监控网站问题, 该方法构建了一个迁移学习指纹攻击模型: 首先使用大量普通流量训练由 8 个卷积层组成的 CNN 模型, 从非目标网站流量中学习到嵌入特征. 接着, 使用少量目标监控网站流量即可微调出可用的分类器. 该方法在使用少量样本时 3 种分类器均优于 CUMUL、k-FP 等方法. Lan 等人[57]提出 DarknetSec, 一种用于暗网流量分类的新型自注意力深度学习方法. 该方法结合了手动提取的特征与自动提取特征方法, 同时应用了侧信道特征和内容特征, 侧信道特征包括序列特征和统计特征, 内容特征则是原始数据包字节序列. DarknetSec 使用 CNN 从内容特征中提取信息, 从原始数据包字节序列中学习流空间特征. 该 CNN 由两层一维卷积网络、批量归一化层和激活函数组成. Zhao 等人[55]使用图残差神经网络开展 Tor 流量分类研究. 该方法将流序列视为一个图, 序列中的每条流都是一个节点, 根据流关系形成节点间连接. 该方法的输入是包含 8 个流的序列, 接着从序列中提取每条流的特征, 将特征向量按照应用程序和时间构建图结构, 使用 4 个残差图神经网络 (ResGCN) 模块从图结构中提取深度特征, 3 层多层感知机实现流量分类. Wang 等人[59]基于神经网络快照集成 (snapshot ensembles) 实现 Tor 网站指

纹识别. 神经网络采取合适的学习率, 使目标函数收敛到局部最小值来寻找最优解. 在迭代过程中目标函数可能会找到不同的局部最小值, 这些局部最小值可能学习到不同程度的知识. 快照集成方法在一次训练中通过调整学习率收集到多个局部最优模型, 结合多个子模型结果得到最优解.

CNN 使用卷积自动提取流量深度特征, 其输入可以是经过简单提取的特征, 也可以是数据包原始字节, 避免了复杂的预处理和特征提取过程. 局部连接和共享权值等特点使其具有极高的准确率和较强的泛化能力, 在流量识别与分类任务中具有出色的表现.

● 循环神经网络 (recurrent neural network, RNN): 是在神经网络中引入了定向循环, 非常适合处理具有前后关联特性的序列. LSTM (long short-term memory) 是 RNN 的一种变体, 可以解决梯度消失问题. Rimmer 等人[44]应用 LSTM 实现网站指纹识别. 与 SDAE 使用前 5000 个数据包的特征、CNN 使用前 3000 个数据包特征相比, 作者仅使用前 150 个数据包特征作为 LSTM 的输入. 这是因为 LSTM 的深度由输入序列长度决定, 更长的输入序列会导致训练时间过长, 以及梯度消失问题. Hu 等人[51]提取了时间、长度、计数等统计特征作为输入, 应用了多种机器学习算法和 LSTM 等深度学习算法开展暗网流量识别研究. 结果显示, 传统机器学习算法的分类性能普遍高于 LSTM 等深度学习算法. 作者指出, 精心提取的特征可能并不适合深度学习, 能够自动提取深度特征才是深度学习算法独特的优势. 虽然 LSTM 擅长处理序列数据, 但对数据包中长度分布等特征的学习能力却不如其他模型. 为了弥补该不足, DarknetSec[57]和 TSCRNN[54]引入了 CNN 和 LSTM 级联模型, 先将流原始数据输入 CNN, 提取抽象的空间特征, 再将 CNN 的输出特征向量输入 LSTM, 提取抽象的时间特征, 这样, 输出的特征向量就同时学习了时空特征. RNN 能够记录历史输入信息, 适合处理序列信息. 流具有明显的数据包序列, 且大部分网页访问行为具有相对固定的交互关系, 因此 RNN 模型在流识别任务中有良好的表现.

深度学习方法学习能力强, 具有极高的上限. 一些方法可以直接从原始流量中自动提取特征, 避免了复杂的预处理过程. 也有一些方法结合了传统手工特征实现流量分析. 深度学习还有极强的适应能力, 在流量的概念漂移问题上比传统机器学习有更好的表现. 然而, 许多深度学习方法需要大量真实数据进行训练, 这对 Tor 流量分析研究是一个挑战.

3.4 主要研究动态

Tor 流量分类通过观察匿名用户流量识别用户访问行为, 对用户进行去匿名化. 自 Tor 诞生以来, 业界就致力于发展各种 Tor 流量分类技术. 最初, 学术界的研究聚焦于 Tor 流量的识别与分类和混淆流量的识别, 后来, 一些学者开始开展网站指纹攻击研究. 网站指纹的研究方法与流量分类接近, 但网站指纹旨在识别用户具体访问的网站. 为了抵御网站指纹攻击, 还有许多工作提出了网站指纹防御方法. 在流量识别技术发展过程中, 学术界不断提出更严格的假设和更广泛的研究场景, 诸如封闭与开放世界场景和多流复用场景下的流量识别研究等.

3.4.1 Tor 流量分类

流量分类检测流量是否来自 Tor、给出具体的流量类型以及识别具体的应用或服务, 可以用于监测匿名用户, 提升服务质量 (quality of service, QoS) 等. 最初的工作使用深度包检测方法实现 Tor 流量检测[99], 已经渐渐不适用于 Tor. 何高峰等人[37]较早地引入机器学习实现 Tor 流量识别. 作者基于 Tor 报文特点分析出 TLS 连接特征和报文长度分布特征, 使用 SVM 识别 Tor 流量.

在实现 Tor 流量识别后, 学术界开始致力于 Tor 流量类型分类. AlSabah 等人[35]较早地开始流量类型分类的研究. 作者提出, 少量的下载服务会占据大量 Tor 带宽, 而 Tor 主要的网页服务却只会消耗少量流量, 这影响了 Tor 的 QoS. 于是论文对 Tor 流量中的网页浏览, P2P 和流媒体 3 种类型流量进行实时分类, 然后基于流量类型提供不同 QoS 策略, 以此改善 Tor 客户端体验. Cuzzocrea 等人[42]则探究了机器学习对 Tor 流量主要类型的分类问题. 还有多项工作不断改进针对流量类型的分类方法, 例如针对多种匿名通信流量的流量分类方法[51], 基于路径签名特征的流分类方法[56], 引入深度学习的流量分类方法[54]等.

应用或服务识别是 Tor 流量分类工作的最细粒度. Shahbar 等人[86]收集了背景流量对 Anon17 中的匿名流量进行了流量分类的初尝试[43]. Montieri 等人[49]探究了不同特征组合、不同特征数量、早期包识别在匿名流量分类

中的表现. 结果表明, 随着分类层面的深入, 识别效果不断下降. 为了提高识别效果和效率, 还有工作使用分层架构实现匿名流量分类[49], 引入分布式计算种的模型并行和数据并行等方法提升流量分类效率[50]等.

3.4.2 混淆流量识别

Tor 提供了多种流量混淆插件, 利用加密、随机填充、隐蔽网桥等方法随机化流量特征, 使攻击者难以识别或屏蔽 Tor. 目前, 使用最多的流量混淆插件包括 Obfs4, Meek, Snowflake 等. Obfs4[100]利用随机填充和加密躲避 Tor 审查. He 等人[101]使用随机性检测算法检测握手负载部分的随机性, 再基于统计特征实现 Obfs4 流量的识别. Liang 等人[102]使用多种策略获取节点特征、协议握手特征、流量特征等实现 Obfs4 流量识别. Meek[103]基于域名前置技术, 将真实的访问隐藏在 TLS 握手协议中, 躲避 Tor 审查. Yao 等人[104]使用高速混合描述流量概率密度分布, 并使用隐马尔可夫模型实现 Meek 流量识别. Snowflake[105]利用 WebRTC 和 DTLS 协议进行点对点的数据传输加密, 躲避审查监管. Chen 等人[106]基于快速规则匹配和 DTLS 握手阶段指纹分析实现 Snowflake 流量的快速精准识别.

Tor 混淆技术主要用于躲避 Tor 审查和抵御流量分析. 随着流量分析技术的发展, 目前主流的 Tor 混淆技术流量都已被攻破[18], 已有研究均能在真实的 Tor 混淆流量上取得高识别率. 然而, Tor 混淆技术并不仅基于加密和随机化流量, 这些技术还经常利用无法一刀切的大型云服务器或经常跳变的普通用户绕过 Tor 审查, 这对未来的 Tor 混淆流量识别技术提出了更高的要求.

3.4.3 网站指纹攻击

网站指纹攻击旨在推测用户正在访问的网站或网页, 可以被攻击者或监管者用于锁定特定用户, 识别特定行为. Cai 等人[36]较早在 Tor 上成功实现了网站指纹攻击. 该方法利用了流相似性特征, 将流量序列转换为字符串, 并计算流字符串之间的 Damerau-Levenshtein 距离, 最后用 SVM 识别流所属网页, 并使用隐马尔可夫模型 (HMM) 判断不同的网页是否来自同一个网站.

许多工作致力于寻找高可用的网站流量特征以及网站分类方法. Panchenko 等人[39]提出长度累积和特征进行网站指纹攻击, 该特征利用了不同网站的交互特点在流量序列上的影响, 具有优秀的分类性能. 作者指出, 面对近乎无限的网站数量以及每个网站中的许多子网页, 网站指纹攻击者只能收集少部分网站进行训练, 因此网站指纹技术本身的准确性可能受到限制. k-FP [40]收集随机森林每个叶节点的决策并计算决策向量间的距离, 作为随机森林指纹. 根据指纹的距离, 测试实例将被分类为距离最近的几个训练样本所属的网站. 通过调整判断策略和测试 k-FP 对特定网站的分类错误率, 该方法可以提升对关注的网站的识别率. Rimmer 等人[44]引入深度学习自动提取流量特征实现网站指纹攻击. 作者评估了 SDAE、CNN 和 LSTM 这 3 种深度学习方法, 针对不同深度学习模型特点, 输入不同大小和特性的特征. 结果表明, 深度学习的识别效果与 Wang-kNN[38], k-FP[40], CUMUL[39]等表现最好的网站指纹攻击方法相近, 其中 CNN 速度更快, LSTM 泛化性更好.

一些工作尝试利用网站指纹识别更复杂的内容. Oh 等人[85]提出关键字指纹识别 (keyword fingerprinting, KF), 识别用户正在用搜索引擎查询的内容. KF 首先判断流是否是一次 Tor 搜索, 然后识别使用的搜索引擎类型, 最后识别特定的搜索词组. 该方法研究了谷歌、必应和 DuckDuckGo 这 3 种流行搜索引擎, 结果表明, 搜索查询的流量比大多数网页都小, 这使得它们很容易被识别. 最终, KF 对 300 个查询词组获得了 91% 的精度与 80% 的召回率. 大多数网站指纹攻击只考虑对网站主页访问行为的识别, Hasselquist 等人[107]则关注网站内容, 开展新闻网站中新闻文章页面的识别. 作者分析了纽约时报、雅虎、福克斯新闻等多个新闻网站的证书大小与 TLS 传输的块大小, 并提取相关指纹识别流量所属文章. Bahramali 等人[65]关注近几年爆发增长的去中心化流量, 如以太坊流量等, 他们使用图神经网络 (GNN) 探索了对去中心化流量实现网站指纹攻击的效果.

为了更进一步展示已有网站指纹工作, 本文在封闭世界与开放世界场景下对 8 种已开源网站指纹工作进行了准确率的对比. 本文选取的数据集为 Tik-Tok[77]中公开的无防御数据集, 该数据集共包括 95 个监控集网站, 每个网站 1 000 个实例, 以及 40 716 个非监控集网站, 每个网站一个实例. 封闭世界与开放世界场景是网站指纹研究中的重要讨论场景, 封闭世界假设监控者知道被监控者将会访问哪些网站, 分类器只需要将流量归属到 95 个监控集网站中的某一种即可完成识别. 开放世界则认为用户可能访问目标之外的网站, 因此分类器面临一个 95+1 类分类任

务 (详见第 3.4.5 节). 本文选取的方法包括: (1) Wang-kNN[38]: 基于流量序列最近邻实现分类; (2) CUMUL[39]: 提出长度序列累积特征实现网站识别; (3) k-FP[40]: 将随机森林叶节点决策作为特征; (4) DF[45]: 基于 CNN 的专门针对防御的网站指纹; (5) AWF[44]: 使用 SDAE、LSTM 和 CNN 这 3 种深度学习方法实现高性能网站指纹; (6) RF[108]: 基于时间片内进出包数量差实现网站指纹.

图 5 展示了上述 8 种方法的实验结果. 可以看出, 所有网站指纹方法都在 Tik-Tok 数据集的封闭世界和开放世界场景中取得了较高的准确率. 对于封闭世界场景, RF 方法取得了 98.7% 的准确率; 对于开放世界, DF 方法取得了 96.7% 的准确率. 可以看出 DF 和 RF 这样的深度学习方法已经在识别效果上超越了传统机器学习方法. 由于加入了非监控类, 网站指纹面临一个更复杂的分类任务, 因此 DF 和 RF 这样高准确率攻击的准确率在开放世界场景中出现了明显下降. 这是因为分类器将许多非监控类误判为监控的网站, 增加了许多误判实例. 其他的方法都同时能够在封闭世界和开放世界中获得较高的准确率, 这说明这些方法不仅可以准确识别监控网站, 也可以把非监控网站排除出去. AWF-CNN 使用深度学习自动提取了更通用和稳定的特征, 在开放世界场景中可以把更多的非监控网站排除出去, 因此准确率有较大提升. 总体而言, 以上方法均能很好地执行网站指纹攻击任务, 最近的一些工作开始结合以上工作, 例如同时使用不同文献中的特征、多层神经网络等, 实现了更高效果的网站指纹攻击.

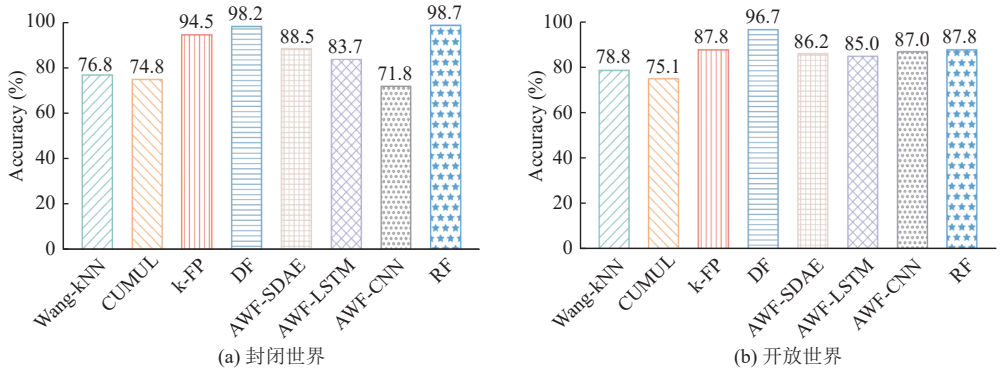

图 5 封闭世界与开放世界场景下多种网站指纹攻击效果对比

3.4.4 网站指纹防御

网站指纹防御通过流量整形、填充数据包或随机化链路等手段消除流特征, 达到抵御流量分析的目的. 不断有新的网站指纹防御技术提出, 抵御网站指纹攻击; 同时也有许多工作集成了更新的技术, 突破网站指纹防御[31]. 许多技术通过在流量中填充数据包以改变流特性. BuFLO[109]严格控制流量, 使用固定大小数据包和固定流速以抹除数据流特征. Tamaraw[110]改进了 BuFLO, 优化了填充策略与上下行数据的速率, 减少了部分开销. 以上方法虽然能有效抵御流量分析, 但引入了较高的延迟和带宽开销. 一些工作尝试实现较小开销的网站指纹防御方法. WTF-PAD[111]通过在流间隔比较大的位置插入虚假的数据包, 实现轻量级的网站指纹防御. WTF-PAD 对正常访问行为的影响较小, 且仅额外增加了少量的带宽. 与 WTF-PAD 类似, BiMorphing[112]也是一种轻量级网站指纹防御. 不同的是, BiMorphing 对上行和下行的突发分别进行采样和填充, 并使用数字优化技术最小化带宽开销. Gong 等人[113]提出了一种零延迟轻量级防御, 分为 FRONT 和 GLUE 两种策略: 针对流量前部包含较多网站行为信息, 容易被用于实现网站指纹特点, 作者提出 FRONT 策略, 在流量前部加入高度随机的伪造数据包使得流量前部变得难以识别; GLUE 策略用于在数据流较大间隔处加入数据包, 使得流量连续, 没有停顿, 攻击者无法判断流结束位置. 最近有工作证明, 基于填充数据包的方法会增加数据包队列长度, 实际上总是会增加延迟, 因此并不能称为“零延迟”[114].

除了填充数据包, 还有工作通过拆分流实现网站指纹的防御. TrafficSliver[115]将一条流拆分到多个 Tor 通道, 限制攻击者能够观察到的流量, 使得流量分析方法失效. Abolfathi 等人[116]也利用了多路径路由和欺骗抵御流量分析. 首先建立多条 Tor 路径, 将流量拆分并分发到不同路径; 欺骗策略则在真实数据包中穿插伪造的数据包. 最近一些工作成功地使用流量变形实现了网站指纹防御. Ling 等人[117]提出专门针对深度学习网站指纹攻击的防御方法.

首先选择一个最小相似网站, 然后注入少量假 cell, 使得流量变形成相似的网站. Surakav[118]基于生成对抗网络(GAN) 学习正常的网页流量模式, 再调制流量模仿这些流量模式, 抵御网站指纹攻击.

表 4 展示了以上网站指纹防御方法在策略、额外负载、抵御流量分析能力方面的对比. 一般来说, 为了增强抵御流量分析能力, 必然需要引入较高的带宽或时间负载; 相反, 想要实现高性能通信就难以获得很强的抵抗流量分析能力. 而随着网站指纹防御技术的逐渐成熟, 许多网站指纹研究者也开始关注这些技术. 一些工作评估了在网站指纹防御上的识别效果[62,108]. 这些工作收集使用了防御方法的网站流量, 并用自己的方法识别网站. 多项工作表明, 流量变形方法和流量拆分方法产生的流量较容易识别, 而基于填充数据包的防御方法中, WTF-PAD 等轻量级防御方法也较容易识别, 而 Tamaraw 等高开销的防御方法几乎无法识别.

表 4　部分网站指纹防御方法对比

防御	年份	防御策略	延迟	额外带宽负载	额外时间负载	抵御流量分析能力
BuFLO[109]	2012	固定数据包发送速率	高	高	高	强
Tamaraw[110]	2014	优化了BuFLO填充策略与上下行数据速率	高	高	高	强
WTF-PAD[111]	2015	在间隔较大的位置插入伪造数据包	低	中	无	中
Walkie-Talkie[119]	2017	以半双工模式实现防御	高	低	低	弱
BiMorphing[112]	2019	上下行数据分别采样和填充	低	中	无	中
FRONT[113]	2020	在流量前部加入高度随机的伪造数据包	低	中	无	中
TrafficSliver[115]	2020	将一条流拆分到多个通道	低	无	无	强
Surakav[118]	2022	调制流量模仿正常流量模式	低	中	低	中

3.4.5　研究假设

流量分类研究通常都在一定假设前提条件下开展, 假设条件涉及威胁模型、用户访问行为、分类目标等. 这些假设使得研究人员得以量化具体场景参数, 但同时也限制了这些研究在多样化场景中应用的能力. 随着科学研究的循序渐进, 不断有工作挑战更加严格的研究假设.

● 封闭世界与开放世界假设: 最初的 Tor 流量分类工作在封闭世界场景下进行. 封闭世界假设监控者知道被监控者会访问哪些网站, 将流量识别转化为一个多分类问题. AlSabah 等人[35]使用机器学习方法对 Web、P2P、流媒体 3 种流量进行分类. Montieri 等人[48]开展了针对匿名工具流量的识别与分类工作. 作者认为, 封闭世界中的流量分类可以作为真实应用的上限. Cai 等人[36]研究了封闭世界中的 Tor 网站指纹攻击问题.

2014 年, 流量分析工作中的封闭世界假设受到了 Tor 开发人员[120]与 Juarez 等人[121]的批评, 他们认为这些方法在现实开放世界中会面临诸多问题. 自那时起, 许多工作开始了开放世界中的流量分析. 开放世界接受用户访问目标之外的内容, 分类器除了需要识别出目标监控流量以外, 还要将非监控流量排除出去. Cuzzocrea 等人[42]设置了一组普通流量和 Tor 流量模拟开放世界中的 Tor 流量分类研究. 普通流量与 Tor 流量一一对应, 这些方法首先从混合流量中识别出 Tor 流量, 然后对 Tor 流量进行分类. 在网站指纹研究中, 研究者通常将开放世界中的网站指纹转化为监控网站集与非监控网站集的识别问题[40], 首先收集 n 个网站的流作为监控集, 每个网站收集 m 次, 再选择 x 个网站作为非监控集, 每个访问一次, 形成一个 $n\times m+x$ 的数据集, 分类器接受一条流, 判断属于监控集中的哪个网站或属于非监控集. Wang 等人[59]认为当前网站指纹中的开放世界假设仍然不能准确反映真实世界情况. 当前工作收集大量非目标网站样本作为非监控集, 这远远不能代表现实中用户访问的几乎无限的网站. 针对该问题, 作者提出了广世界 (wide-world) 的概念, 将监控集和非监控集外的网站集也考虑在内.

封闭世界、开放世界和广世界是对流量识别工作在现实场景的假设. 过于简单的假设讨论出的方法存在有效性不足等问题, 而过于复杂的假设则不利于研究的开展. 随着各种方法在开放世界中取得的成功, Tor 流量识别的假设会越来越接近现实.

● 单流与多流访问假设: Tor 将所有请求复用到匿名通道, 因此从匿名通道观察到的 Tor 流常是多条流的重叠. 该问题又存在两种情况: 一是多流的复用问题, 当用户打开一个应用或网页, 可能产生多条流, 与多个服务器建

立连接. 二是多标签问题, 用户很可能同时打开多个标签页, 访问多个网站. 重叠的流改变了流的原有特征, 使得常规针对单流的识别方法失效.

一些工作利用流子序列相似性从复杂流中识别网站. Zhuo 等人[122]认为网页中包括许多链接, 对不同链接的访问产生了不同的子序列, 用户点击链接的行为可能导致子序列随机排列和重叠. 因此, 作者引入了 profile 隐马尔可夫模型 (PHMM), 通过描述序列中每个位置的概率分布以及一个位置转换到另一个位置的概率, 并允许灵活地插入或删除元素, 实现对与原始序列具有相同或类似特征的其他序列的搜索. 该方法可以很好地识别出通过超链接连接多个页面的网站流量. Ma 等人[84]提出了上下文感知的识别方法, 首先使用聚类提取出网站的代表性流量并量化代表性流量与网站访问的关联程度, 然后通过识别流匹配到的子序列来帮助分类器判断识别的网站.

其他的工作通过寻找流的起始点分割流实现流识别. Wang-kNN[41]分析了访问多标签网页的 4 种情况: 第 1 个网页加载完成后一段时间打开第 2 个网页, 中间有较大时间间隔; 加载完第 1 个网页立即打开第 2 个网页, 中间几乎没有时间间隔; 同时打开两个网页, 流量重叠; 单网页访问. 作者首先使用一个 kNN 分类器根据 cell 前后的时间特征判断流量分割点, 然后再进行网站指纹分析. Yin 等人[58]也采用先判断分割点再分类的方法. 作者扩展了 Wang-kNN 的特征集, 对两个及以上的多标签网页流量进行识别. 他们引入了 BalanceCascade 方法训练 XGBoost 分类器, 获得一个可以真实判断分割点的分类器. Yang 等人[87]考虑了 Tor 通道构建部分的流量, 提出了一种主动延迟网站指纹攻击方法. 该方法首先推断 Tor 通道构建过程, 并以此定位第 1 个 HTTP 请求的位置, 接着, 设计了基于统计分析的算法和基于目标函数优化的算法推断后续 HTTP 请求的第 1 个单元位置, 然后通过延迟 HTTP 请求分离访问不同 Web 对象的流, 最终实现对多流的识别.

Deng 等人[61]认为与识别元素相关的数据包往往集中在一个小流量段中, 因此他们使用滑动窗口从重叠的流中提取片段, 使用改进的自注意力机制实现多标签页的识别. Jin 等人[66]受图像领域目标检测算法的启发, 将多标签识别转化为一个长度序列中的有序集预测问题, 使用深度融合网络 (DFNet) 实现对网站流量的知识提取, 使用 Transformer 实现多标签的查询, 最终实现多标签网页的识别. 多流的识别问题一直是流量分析领域的难题, 而 Tor 的匿名通道机制使得 Tor 上的多流识别难度更大.

● 网页与网站访问假设: 大部分网站指纹工作将网站主页作为网站识别的唯一目标. 这隐含地假设用户每次都通过主页访问网站, 忽略了对网站子页访问的讨论. Mani 等人[123]对 Tor 用户使用情况的调研表明, Tor 用户访问网站的方式在很大程度上是未知的, 因此目前的网站指纹研究与真实情况仍然存在较大差距. Cherubin 等人[60]在真实的 Tor 出口中继收集流量开展网站指纹研究以解决该问题. 作者认为, 在出口中继收集至少具有以下优点: (1) 完全真实的 Tor 流量; (2) 出口中继可以根据 DNS 请求判断域名, 而用户访问不仅包含网站主页, 还包含网站子页; (3) 出口中继流量包含了用户访问的多样性以及更少的噪声. 他们在出口中继捕捉流量, 根据 DNS 结果给流量打上标签并训练模型, 将模型部署在入口节点以实现网站指纹部署. 该方法是首次基于真实 Tor 用户流量开研究的网站指纹研究. 他们的结果表明了流量识别方法在真实环境中部署的可行性. 然而, 该方法对流量的标签来自出口中继的 DNS 请求, 只能实现对某个域名的识别, 无法执行更细粒度的分析, 也无法应用于隐藏服务等更具体的应用场景.

流量分类通过分析用户流量判断用户访问的内容与行为, 易于实现, 准确率高, 是最重要的 Tor 去匿名化方法之一. 在当前的研究中, 已有多种仿真环境、真实网络流量采集方法, 然而, 由于 Tor 网络环境的复杂性、Tor 访问对象的隐蔽性、用户行为的多样性, 目前还没有一个统一的权威公开数据集, 使得不同方法的可对比性受限; 在使用的特征与算法方面, 已有工作多年来已经提出了大量流量特征和算法, 且许多特征已经被证明有效, 许多分类算法也依据所提取的特征定制, 因此, 如何根据需求进行特征筛选并选择合适的算法也是一项值得探讨的工作. 在未来的工作中, 则需要探讨面向复杂背景流量、大量高速网络环境、网站指纹防御、网站与网页、概念漂移等复杂场景下的 Tor 流量分类工作.

4 Tor 流关联研究进展

Tor 流关联研究主要解决如下问题: 如何获取并分析 Tor 入口端和出口端流量, 关联通信两端, 实现双端去匿

名化. 与流量分类研究相比, 流关联的威胁模型假设更强, 例如需要掌握双端流量等, 但对 Tor 的去匿名化程度更高, 可以确定 Tor 通道两端的通信关系. 本节将主要工作梳理在表 5, 首先在第 4.1 节梳理了现有工作获取双端流量的方法, 并按捕获手段分为代理级别、节点级别和网络级别的获取方法; 在第 4.2 节梳理了流关联研究使用的特征; 在第 4.3 节将流关联方法分为基于统计度量的关联方法和基于机器学习的关联方法.

表 5　主要的流关联工作

文献	年份	特征	攻击/防御方法	评价指标
Steven等人[124]	2005	统计延迟特征	利用流量模式和探测延迟之间的相关性	FP、FN
Edman等人[125]	2009	拓扑特征	提出了AS感知路径选择算法	路径暴露率 (path compromise)
Gorman等人[126]	2009	统计计数特征	改进包计数、互相关系数和互信息的方法进行关联攻击	K-匿名性 (K-anonymity)
Mittal等人[127]	2011	电路吞吐量	皮尔逊积矩相关关联吞吐量识别入口节点	节点分布熵
Akhoondi等人[128]	2012	拓扑特征	LASTor: 具有可调节AS感知WPS路径选择算法的Tor客户端	Accuracy、FPR和FNR
Johnson等人[19]	2013	拓扑特征	一个Tor对于流关联攻击的安全分析框架	路径暴露率
Sun等人[12]	2015	统计特征、拓扑特征	斯皮尔曼等级相关算法进行非对称相关攻击和最长前缀攻击来获取更高比例的Tor链接	Accuracy
Nithyanand等人[129]	2016	拓扑特征	Astoria: 具有考虑不对称相关攻击的AS感知路径选择算法的Tor客户端	可观察带宽
Barton等人[130]	2016	拓扑特征	DeNASA: 目的地无关的AS感知路径选择算法	Tor流的脆弱率 (Tor stream vulnerability)
Sun等人[131]	2017	拓扑特征	Counter-RAPTOR: 考虑节点弹性的Tor守卫节点保护算法	对劫持攻击的恢复能力 (hijack resilience)
Nasr等人[132]	2017	统计延迟特征	利用余弦相似度关联压缩处理后的IPD特征, 并提出一系列降低关联开销的方法	真阳性(TP)和假阳性(FP)
Johnson等人[133]	2017	拓扑特征	信任感知路径选择算法TAPS	路径暴露率
Nasr等人[75]	2018	深度学习自动提取特征	DeepCorr: 利用深度学习CNN对Tor流量进行流关联攻击	TPR、 FPR和Accuracy
Wails等人[134]	2018	拓扑特征	利用时间动态信息提升获取Tor双端流量机会的Tempest攻击	AS的后验分布熵
Guan等人[135]	2020	构造累积序列特征	ResTor: 流关联预处理降噪模型, 利用余弦相似度关联字节累计序列等提升关联性能	TP、FP
Palmieri[136]	2021	统计特征	利用皮尔逊积差相关不同分辨率级别关联流的小波系数	TP、FP
Rimmer等人[72]	2022	深度学习自动提取特征	更接近真实网络的多代理设置获取双端流里	平均精度
Oh等人[137]	2022	深度学习自动提取特征	DeepCoFFEA: 利用FEN度量学习和放大技术提升关联性能的流关联攻击	TPR、FPR和BDR
Tan等人[20]	2022	电路特征	通过改变Tor路由选择算法掌握出、入口节点, 进而关联电路指纹的Trapper攻击	路径暴露率、Accuracy和FPR
Zhang等人[138]	2022	深度学习自动提取特征	利用虚拟包注入打破CNN已学习的流模式进行流关联防御	保护成功率 (protection success rate)
Lopes等人[139]	2024	统计特征	SUMo: 根据双端流滑动子集总和计算双端流相关性	TPR和精度

4.1 双端流量获取方法

在流关联攻击中, 攻击者拦截 Tor 连接两端的流量, 通过关联双端流量实现对发送方和接受方的去匿名化. 在流关联攻击研究中, 研究人员需要同时收集入口流量与对应的出口流量, 这在实验中与实际攻击过程中都是实现的难点. 在当前研究中, 双端流量获取方法可以分为代理级别、节点级别与网络级别.

4.1.1 代理级别的流量获取

当前研究通常使用代理获取 Tor 出口流量[75]: 首先搭建一个代理服务器, 在本地将请求转发到代理客户端, 代理客户端再将请求转发给 Tor 客户端, Tor 客户端又将请求转发到 Tor 通道. 这样, Tor 出口节点解析请求后, 会将请求数据发往代理服务器, 代理服务器访问网页. 即, 请求转发路径为“浏览器-代理客户端-Tor 通道-代理服务器-Web 服务器”. 研究人员在本地获取入口端流量, 在代理服务器处获取出口端流量. 基于代理的流量获取解决了出口流量难以获取的问题, 但该方法为连接创建了额外的跳点, 干扰了流量的某些计时特性, 影响流量之间的可区分性. Rimmer 等人[72]在研究中使用多个代理捕获出口流量, 最小化了基于代理的端到端额外延迟, 使其更接近于现实的计时测量. 在他们的多代理设置中, 现有最好方法的平均性能下降了 7.95%, 这说明了以往基于单代理方法的某些缺陷.

4.1.2 节点级别的流量获取

基于代理的流量获取解决了实验数据集问题. 但在实际攻击中, 如何捕获目标客户端的双端流量是流关联研究的关键难点之一. 许多研究提出基于节点的流量获取方法. 攻击者运行自己的 Tor 节点, 通过一系列的手段使得这些恶意节点成为 Tor 电路的入口节点和出口节点, 继而获取 Tor 连接流量. Murdoch 等人[124]提出了一种节点识别方法, 用于推断节点是否被用来转发目标流, 从而可以进一步更改节点的选择. 攻击者的恶意服务器向目标用户发送特定流量模式数据, 判断监测节点上的流量是否与该流量模式有相关性. 由于攻击者与普通的 Tor 客户端行为相近, 这种攻击不容易被检测. Mittal 等人[127]提出了一种更加隐蔽的节点识别方法, 不同 Tor 节点的容量存在显著差异, 其中容量低的中继成为 Tor 流的瓶颈, 而 Tor 入口节点往往容易成为瓶颈节点, 因此共享同一入口节点的两个电路将具有高度相关的吞吐量, 利用相关系数测量吞吐量的相似性判断两个电路是否共享同个子路径, 多次电路重构后可以识别入口节点.

在如何使 Tor 选择自己的恶意节点作为入口和出口节点方面, Johnson 等人[19]证明了 Tor 客户端的入口与出口节点选择策略与节点带宽高度相关, 攻击者在部署恶意节点时应该为不同类型节点分配不同带宽. 作者的实验表明, 当拥有的恶意节点总带宽资源固定时, 5∶1 的入口和出口带宽分配可以最大化双端流捕获概率. 为了降低攻击者通过控制恶意中继节点执行流关联的机会, Tor 引入了守卫节点的机制. 然而, Tan 等人[20]提出 Trapper 攻击, 引入具有防检测机制的蜜罐节点 (honey relays), 通过阻塞对其他守卫节点的选择增大目标客户端选择恶意守卫节点的概率; 该方法在掌握入口节点后, 通过破坏电路, 迫使目标客户端同时选择其控制的出口节点. 此外, Trapper 攻击还可以加快 Tor 客户端更新守卫节点列表的速度, 将更新间隔缩短到 3 min, 从而增加目标客户端选择蜜罐节点的机会.

4.1.3 网络级别的流量获取

更强大的攻击者可以通过控制、监听自治系统 (AS) 或互联网交换点 (IXP) 来增加执行流关联的机会, 也可以通过一些路由操作将更多 Tor 连接重新路由通过他们的恶意 AS 和 IXP, 增加记录双端传输 Tor 流量的机会. Nithyanand 等人[129]展示了大约 40% 的 Tor 电路容易受到单个恶意 AS 的流关联攻击, 此后大量工作从 Tor 的控制平面入手获得双端流量. Sun 等人[12]提出的 RAPTOR 攻击, 使 AS 级别的攻击者可以利用网络路由的不对称特性、路由波动来增加观察到 Tor 流量的机会, 并提出最长前缀攻击实现 BGP 路由劫持, 使攻击者能够识别用户通信. Tan 等人[140]对 RAPTOR 方法进行量化, 并提出了类似原理的最短路径攻击, 两种攻击都不易被用户发现. 此外, Johnson 等人[19]的分析表明, 网络级别流量获取的成功率与客户端行为也有一定的关系, 客户端目的地多样性越低, 双端流越容易被掌握, 因为攻击者必须覆盖的目标区域集合更窄, 用户更有可能同时选择对手掌握区域的入口和出口. Wails 等人[134]同样证明了客户端的移动性、用户行为和网络路由更新等时间动态信息对攻击者掌握双端流量能力的提升. 网络级别流量获取大多利用路由选择算法, 将 Tor 连接重定向至恶意 AS 或 IXP, 实现流关联. 也有很多工作通过优化 Tor 路由选择算法以防御流关联攻击[125,128–131,133,141].

除了以上流量获取方法, Palmieri[136]也使用公开数据集 ISCXTor2016 作为双端流量. 作者巧妙地将 Tor 流量作为入口流量, 将对应的 nonTor 流量模拟为解密后的出口节点流量. 然而, 此数据集是在受控环境下收集的, 没有

经过真实完整的 Tor 网络, nonTor 流量并不能体现出口节点特性, 因此该方法没有广泛使用.

Tor 流关联中目前公开的流量集较少, 集中在 2018 年后, 因为节点级别和网络级别的数据收集会造成节点位置多样性的损失, 同时涉及强烈的道德问题, 威胁 Tor 用户的安全性, Nasr 等人[75]利用代理级别的流量收集方法, 收集了目前为止最大、最全面的相关流量数据集 DeepCorr[142]; 之后的 2022 年, Rimmer 等人[72]通过使用多代理替代单代理收集改进了洲际延迟, 收集了数据集 Trace Oddity[143], 使数据更接近真实 Tor 网络; 同时, Oh 等人[137]模仿 DeepCorr 的收集方法收集了数据集 DCF[144], 测试 DeepCoFFEA 对于长训练间隔以及使用大型非训练集进行关联的能力.

4.2 流关联特征

流关联研究通常将双端流量转化为多个特征值或特征向量, 对比二者的相似度以判断它们是否关联. 在 Tor 流关联研究中, 使用的特征通常包括电路特征和流量特征.

● 电路特征: 在 Tor 流关联中, 电路特征是指用于描述 Tor 电路属性或信息的指标, 包括电路的构建时间、电路中传输 cell 的类型、数量和方向等. 这些特征能够区分不同的 Tor 流, 从而确定哪些流属于同一个 Tor 连接. 其中, 吞吐量是电路特征中的一个重要指标, 用来衡量该电路的传输能力和效率. Mittal 等人[127]证明共享同一入口节点的两个电路将具有高度相关的吞吐量, 因此在宏观层面的时间尺度上通过电路的吞吐量特征进行节点识别. Chakravarty 等人[145]通过对两个流时间序列在相同时间节点上的吞吐量值进行关联, 实现了大规模流量中使用 Netflow 进行的轻量级 Tor 流关联攻击. Tan 等人[20]对潜在的源-目的地电路对使用吞吐量特征进行快速关联分析, 识别最相似的双端流量.

● 流量特征: 在流量特征方面, Tor 流关联主要关注统计特征、构造特征以及深度学习自动提取特征. O'Gorman 等人[126]使用字节计数和包计数两种统计特征进行关联攻击可行性的研究, Sun 等人[12]通过提取 TCP 报头中的 TCP 序列号和 TCP 确认号字段来统计单位时间内传输的数据字节数, 还有一种常见的统计特征分组间延迟 (inter-packet delay, IPD) 被 Nasr 等人[132]用于更有效地捕获时序变化. 为了更进一步提升关联性能, Guan 等人[135]使用数据包大小和时间间隔的字节累积序列对流进行了更具鲁棒性的特征描述, Palmieri[136]则关注数据包速率、比特速率、平均包到达时间、平均包大小等多个特征, 使用小波多分辨率分析技术将其捕获为小波系数, 以此构造出更好处理瞬时和局部特性的特征. Lopes 等人[139]关注双端流每对流之间单位滑动窗口内出入包的总和特征, 最后根据所有滑动窗口间相关性实现流关联. 此外, 近年来还有一些工作引入深度学习, 自动化提取流量特征[75], 使得提取的特征更具有普遍性和稳定性.

4.3 Tor 流关联方法

在判断双端流量是否关联方面, 本文按照匹配算法将现有工作分为基于统计度量的关联方法与基于机器学习的关联方法.

4.3.1 基于统计度量的关联方法

基于统计度量的关联方法是 Tor 流关联的主要方法, 该方法使用统计度量来表示数据流的相似性, 每条流被抽象为表征流形状的矢量, 例如分组方向、分组大小、分组间隔、字节突发的序列等, 然后将矢量输入到度量计算公式中以判定相似性. 如果相似性超过某个阈值, 则认为流对是相关的. 不同度量方法的差异在于描述流的特征以及计算特征相似度的算法. 在 Tor 流关联中, 常用"线性相关"和"非线性相关"描述两个或多个流特征之间关系的性质.

4.3.1.1 线性相关

当两个流特征之间的关系可以用一个线性方程或线性函数来表示时, 它们之间即存在线性相关. 简单来说, 如果一个流特征的变化与另一个流特征成正比或成反比, 且变化的关系是直线的, 其相关性就可以用线性相关系数来衡量, 线性相关包括交叉相关、余弦相似度和皮尔逊积矩相关.

● 交叉相关 (cross-correlation): 是一种用于衡量两个时间序列之间相关性的方法, 其衡量方式考虑了第 1 个序列相对于第 2 个序列的位移. 因此, 它在估计两个时间序列之间的延迟或检测较短序列在较长序列中的出现时非

常有效, 可以应用于流关联中. 当进行归一化处理后, 它会得到一个时间依赖的相关性估计器, 其取值范围在 [−1, 1] 之间, 1 表示完全正相关, 0 表示不存在相关, −1 表示完全负相关. 当完全负相关时, 对于变量中的每个增量, 在另一个变量中存在对应的固定比例递减.

O'Gorman 等人[126]首次将交叉相关的方法应用在 Tor 流关联领域, 并使用 K-匿名性曲线衡量关联的性能. 此工作描述了两种网络对流量的影响并提出了相应的补偿技术, 将关联的精度提高了 10%–40%. 在模拟的 Tor 网络上, 可以立即关联约 50% 的流量; 而在真实 Tor 网络中, 10% 的流量可以被成功关联, 这是最初成功的 Tor 流关联攻击研究.

交叉相关的方法存在一些局限性. 首先, 它可能受到观测窗口的选择的显著影响, 且对于时间序列中存在的局部强度波动非常敏感. 此外, 当处理线性动态并非显著相关但数据波动变化很小的时间序列时, 交叉相关性可能完全失效.

● 余弦相似度 (cosine): 是一种度量向量相似性的方法, 更适用于向量数据的相似性分析, 因此常被用于流关联相关研究. 它计算多维空间中投影的随机变量相关的样本向量之间的夹角余弦. 该度量得到的相似度值在−1 (完全相反) 和 1 (完全相关) 之间, 其中 0 表示向量之间的正交性或完全缺乏相关性, 中间值对应相似或不相似的程度.

Nasr 等人[132]依靠降维方法来提高流关联的效率, 他们使用高斯随机投影算法压缩流量特征, 并提出了一整套减少开销的流关联步骤. 作者选择了非性能最优但是计算速度较快的余弦相似度相关算法来关联压缩后的数据包间延迟 IPD 向量, 最终取得了良好的关联性能. 此外, 作者还使用了低开销抵抗数据包级修改的算法来提升关联算法的同步性, 并基于位置敏感哈希 (LSH) 数据结构设计了一种用于存储压缩 IPD 特征的快速数据结构, 大幅度降低了关联方发在计算、通信、存储方面的开销. Guan 等人[135]提出了第一个流量相关领域的去噪模型. 它使用数据包大小和时间间隔的字节累积序列对 Tor 流的形状进行特征的描述, 并使用二者累计序列的乘积表征流观察流持续时间. 作者在测试中发现, ResTor 结合余弦相似度关联的效果最好, 在达到了与当时最先进的 Tor 流关联方法 DeepCorr 相当的高精度结果的同时, 他们的方法的关联速度提升了 2 倍. 这也是 Oh 等人[137]选择使用余弦相似度关联方法降低计算成本的原因.

由于余弦相似度衡量的是两个向量之间的方向相似性, 而不考虑其大小或幅度, 在处理线性相关性较弱的特征数据时可能表现较好, 但对于具有明显线性趋势的流关联特征数据, 可能无法很好地捕捉其相关性.

● 皮尔逊积矩相关 (Pearson product moment correlation, PPMC): 是在流关联中使用最多的线性相关统计度量, 它通过衡量两个流量特征之间的线性关系强度判断关联度. 其取值区间与其他线性相关方法相同. 皮尔逊相关系数可以灵活地处理不同形式的特征, 包括时间序列、向量等, 适用于连续和离散数据, 且简单易用, 因此被大数流关联研究工作采纳. Mittal 等人[127]证明了共享同一低转发容量节点的两个电路将具有高度相关的吞吐量, 因此利用皮尔逊积矩相关测量吞吐量相似性能够判断两个电路是否共享一个子路径, 经过多次电路重构, 即可识别电路中的节点. 此外, 由于 Tor 拥塞控制机制中的批处理行为, 当以较小的时间尺度对吞吐量进行采样时, 可以观察到两个流对电路的互斥使用. 作者通过在宏观层面的时间尺度 (如 5 s) 上判断吞吐量的相关程度, 在微观层面的时间尺度 (如 0.5 s) 上判断吞吐量的互斥程度, 实现了在 5 min 内关联来自同一发起者的两个连接, 交叉错误率小于 1.5%. Palmieri[136]提出了使用小波多分辨率分析 (wavelet multi-resolution analysis) 对数据进行预处理的方法, 小波的分解和重构机制能够更好地处理瞬时特征和局部特征. 作者通过在不同分辨率级别上使用皮尔逊积矩相关对小波系数进行关联, 完全避免了特定误差 (FP 和 FN), 同时保证累积误差接近于 0. 由于依赖于轻量级技术, 这种方法能够以分布式的方式部署在大量的观察点上, 形成一个系统的观察框架, 支持执法机构在侦查犯罪活动中开展多种复杂的关联工作. Tan 等人[20]提出了一种两阶段提高准确性的关联方法. 第 1 阶段利用 Tor 电路构建的方式, 标记可能来自被攻击的 Tor 用户的源-目的地电路对, 第 2 阶段通过计算 Tor 电路的吞吐量, 利用皮尔逊相关系数来识别最匹配的服务器和客户端.

皮尔逊积矩相关在小样本数据上可能会产生不稳定的结果, 相比之下, 适用非线性相关的斯皮尔曼等级相关通过改进皮尔逊相关对原始数据的处理, 在小样本数据中也能提供较为可靠的关联度量.

4.3.1.2 非线性相关

当两个流特征之间的关系不能用一个线性方程或线性函数来表示时, 则称它们之间存在非线性相关. 非线性相关意味着特征之间的关系不是直线的, 而可能是曲线、指数、对数等非线性形式. 应用于 Tor 流关联领域的非线性统计度量主要包括互信息和斯皮尔曼等级相关.

● 互信息 (mutual information): 是一种用于度量随机变量之间相关性的指标, 反映了两个变量之间的信息交流程度. 在流关联分析中, 互信息可以用来衡量流之间的依赖程度和相关性. 当两条流的特征之间存在较高的互信息时, 意味着它们的特征值更可能在同一时间点出现, 从而表明两条流具有较强的相关性. 归一化后, 互信息可表征 0 (没有任何相关性) 到 1 (完全匹配) 之间的值. Zhu 等人[146]首次在匿名流关联中使用了互信息作为时域上的相似度度量, 证明了当时的匿名保护机制对于流关联攻击的脆弱性. O'Gorman 等人[126]使用互信息关联窗口内的数据包数量序列, 在相同 K-匿名性要求下, 无论在模拟 Tor 网络还是真实 Tor 网络中, 该方法性能都优于传统包计数和互相关性方法.

互信息的计算需要足够的数据量和较长的特征向量才能获得可靠的结果. 这是因为互信息需要对目标流的流量特征进行经验分布的重构和比较, 较短的流或较少的特征都可能导致不准确的结果.

● 斯皮尔曼等级相关 (Spearman's rank correlation): 通过度量两个变量排名 (或称为秩) 之间的统计依赖性评估两个变量的相关性. 与皮尔逊积差相关一样, 其取值区间为 [−1, 1]. Sun 等人[12]利用 Internet 路由的不对称特性、路由波动等特性增加观察到流量的机会. 该工作使用了 50 对流, 每对流包含 300 秒的流量, 对每条流提取 TCP 报头中的 TCP 序列号、TCP 确认号字段统计单位时间内传输的数据字节数, 通过计算 Tor 连接之间的斯皮尔曼等级相关系数, 选择与入口流量相关性最高的出口流量作为最佳匹配, 来实施流关联攻击.

斯皮尔曼等级相关基于两个变量的排名而非原始数值进行计算, 对于数据中的小幅度变动不敏感, 因此在轻度噪声或数据包不完整的情况下, 仍能保持一定的准确率. 然而, 由于对原始数据的排序和排名操作, 在处理大规模数据时使用斯皮尔曼等级相关可能会导致计算复杂度较高, 影响关联效率.

表 6 展示了基于统计度量的关联方法性能对比. 线性相关基于线性函数评估两条流序列的相关性, 与非线性相关算法相比, 消耗了较小的计算开销的同时取得了不俗的准确率, 因此近年来被越来越多的工作使用.

表 6 基于统计度量的关联方法的比较

关联方法		取值范围	计算开销	最优性能下的准确率
线性相关	交叉相关	[−1, 1]	O(n)	≈75%[126]
	余弦相似度	[−1, 1]		≈90% (CTA)[132] ≈95% (ResTOR)[135] ≈98% (DCF)[137]
	皮尔逊积矩相关	[−1, 1]		≈98%[136]
非线性相关	互信息	[0, 1]	O($n\log n + M\times N$)	≈80%[126]
	斯皮尔曼等级相关	[−1, 1]	O($n\log n$)	≈95%[12]

注: M 和 N 代表不同特征离散化后的取值数量

4.3.2 基于机器学习的关联方法

由网络状态引起的噪声 (例如, 拥塞、动态带宽分配、设备故障) 等动态复杂的变化会在一定程度上影响矢量表示, 导致基于统计度量的关联方法的相关精度降低. 近些年, 基于机器学习的流关联方法逐渐兴起, 为解决噪声问题提供了新的思路. 基于机器学习的方法可以学习到流深层特征, 具有更强的非线性拟合能力, 相比于简单的统计度量, 对网络状态噪声具有更好的抵抗力. 在 Tor 流关联攻击中, 使用到的基于机器学习的关联方法主要分为深度学习和度量学习两种方法.

4.3.2.1 深度学习方法

深度学习算法能够从原始输入中提取复杂、有效的特征, 有效抵抗网络噪声实现流关联. Nasr 等人[75]首次将深度学习应用于 Tor 流关联攻击研究. 其提出的 DeepCorr 基于双端流的 IPD 和数据包大小特征, 学习了一个

CNN 模型作为 Tor 的复杂网络流量模型的流相关函数. 该方法无需学习任何源、目的信息或电路信息, 仅训练了一个可用于链接任意双端流的相关函数. 该方法可以用于关联任意的电路和目标, 不受训练过程中使用的数据集的限制, 具有高可迁移性. DeepCorr 在不同的测试集上均达到了 96% 以上的准确率, 在相同条件下, 与当时最先进的 Raptor 攻击[12]相比, 攻击精度大大提高. 该方法还可以应用到其他除 Tor 以外的流关联中, 如脚踏石攻击[147]. 然而, 考虑到 DeepCorr 攻击对于 Obfs4 等混淆流量机制的关联能力较差, Guan 等人[135]将去噪思想同时应用于 Tor 网络流量的预处理阶段, 使得处理后的 Tor 出口流量流与其对应的入口流量更接近. 作者使用比原始特征更稳定字节累积序列来对流的形状进行特征描述, 并利用堆叠自编码器架构有效地去除了普通状态噪声和混淆噪声, 对于 Obfs4 混淆的流量, 真阳性率保持在 80% 以上, 大大超过 DeepCorr.

许多研究提出了针对 DeepCorr 攻击的防御措施. Tian 等人[148]提出了新的通用扰动生成算法, 通过对流量注入微小扰动来抵御流关联攻击. 该方法仅需 10 ms 微小变化的扰动就可以使 DeepCorr 的 TP 从 82% 降低到 60% (当 FP 为 10^{-3} 时). Zhang 等人[138]通过预先计算对抗实例, 将虚拟数据包注入流中, 成功地打破了 CNN 模型学习到的流模式, 实现了 97% 以上的高成功防御率. 抵御基于深度学习的 Tor 流关联攻击已成为新的研究热点.

4.3.2.2 度量学习方法

度量学习是一种机器学习方法, 其目标是学习数据的相似性或距离度量, 将相似的数据点映射到相近的位置, 将不相似的数据点映射到较远的位置, 实现流关联的判断. 度量学习的目的是改变数据的表示, 使得在新的表示空间中, 数据点之间的距离或相似性更具有意义, 具有计算简单快捷的优点, 可以用于改进深度学习计算开销巨大的问题. 在 Tor 流关联攻击中, 度量学习可以应用在特征嵌入网络 (feature embedding network, FEN) 中. FEN 旨在学习一个有效的特征表示, 使得样本之间的距离能够在特定任务上得到良好的度量. 这包含了深度学习用于学习特征表示的能力和度量学习用于优化距离度量的能力. Oh 等人[137]利用度量学习和放大技术克服了 DeepCorr 计算开销较大的问题, 并进一步降低了误报率. 其提出的 DeepCoFFEA 攻击训练 FEN, 将入口和出口流映射到低维空间, 使得相关的流在嵌入空间中更加相似. 此外, 该方法使用放大技术将流分成短时间窗口, 并在这些窗口之间进行投票, 放大了 TP 和 FP 的差异, 显著降低了误报率. DeepCoFFEA 在实验中表现出比 DeepCorr 明显的优势, 在 FPR 为 2×10^{-4} 时 TPR 高达 93%, 关联速度提高了两个数量级. 此外, 该攻击也考虑了一些潜在的防御措施, 现有的混淆或填充等轻量级防御无法有效保护 Tor 网络免受此威胁.

为了比较机器学习关联方法的性能, 本文选用目前公开的最大的 DeepCorr 数据集对两种代表性的开源模型进行对比实验, 结果如图 6 所示. 本文分别测试 1k 和 6k 规模的 DeepCorr(DC) 和 DeepCoFFEA(DCF) 进行对比, 两个模型的 ROC 曲线如图 6(a) 所示. 可以看出, 这两种基于机器学习的关联方法基本不受测试集规模的影响, 且在 FPR 为 10^{-3} 时均达到了 80% 以上的 TPR. 其中 DCF 对于任何给定的 FPR 都能达到更优的 TPR, 因此也相应拥有更好的 BDR 表现. 不同 TPR 下的 BDR 曲线如图 6(b) 所示. BDR 更能反映在基本率很低的情况下的实际关联能力, DC 在实现较高 TPR 的同时有较低的 BDR, 并且随着训练集的增加有较明显的退化, 而 DCF 的 BDR 曲线反映出其对于降低误报率有着更明显的优势. 总的来说, DC 和 DCF 都具有很高的可扩展性, 可以应用于大规模流关联. 在我们的实验中, DC 需要的训练时间更短, 而 DCF 更适用于追求低误报率的场景.

Tor 被动流关联通过关联客户端流与服务器流, 确定双端的通信关系. 与 Tor 流量分类相比, 双端流量的获取是实施流关联攻击的难点之一, 代理级别的流量获取方法被广泛用于实验中, 但额外延迟使得数据的真实性受到影响; 节点级别和网络级别的流量获取方法可以应用于实际的攻击环境, 但是需要攻击者具备让 Tor 连接经过自己的节点和 AS 的能力. 被动流关联与流量分类都会使用流量特征, 而流关联加入了电路特征以增强对不同连接的流的区分能力. 许多已有工作基于统计度量或机器学习算法实现流对的匹配, 由于双端流量的成对性, 选择关联方法的关键在于适当的计算复杂度和适应不同数据集规模的能力. 在应用方面, 被动流关联攻击去匿名化程度更高, 能够在不影响正常网络服务的情况下对通信的双端进行识别, 有效实现 Tor 客户端与服务端通信关系去匿名化. 然而, 流关联对攻击者能力的要求更高, 威胁模型受限, 发起条件较为苛刻. 在未来, 可以考虑将 Tor 流量分类和流关联两种被动流量分析方法结合使用, 在攻击中相互辅助补充, 增强去匿名化的效果.

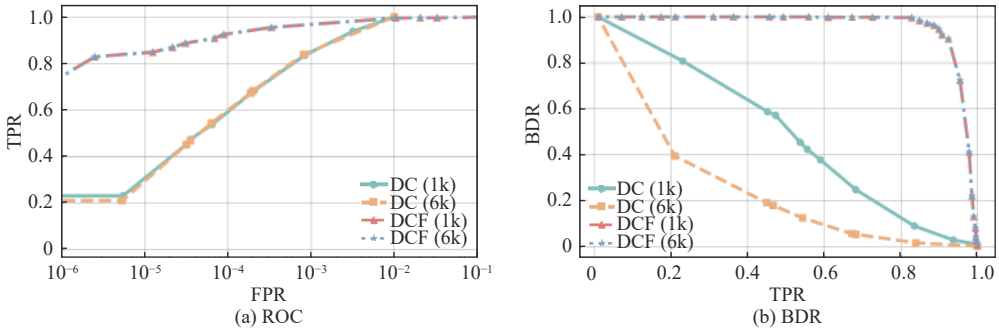

图 6　DeepCorr (DC) 和 DeepCoFFEA (DCF) 模型在不同规模数据集下的性能对比

5　研究挑战与展望

Tor 被动流量分析研究经过 20 多年的发展, 取得了许多重要成果, 已经成为 Tor 去匿名化最重要的研究领域之一. 本文将该领域关键工作按照时间梳理为图 7. 在流分类与网站指纹方面, 越来越多工作引入了深度学习、多层分类架构, 在多流识别、防御等多个场景实现了高性能识别; 在流关联领域, 也有许多工作引入深度学习、降噪模型实现了新型流关联. 然而, 这些方法的实际应用仍然面临着较多的挑战, 主要包括以下几个方面.

图 7　部分 Tor 被动流量分析工作发展历史

(1) 理想威胁模型. 威胁模型描述了流量分析中的各项假设, 包括攻击者能力、用户行为与能力等, 是开展流量分析工作的基础. 当前的工作中设定了较为理想的威胁模型, 利于研究工作的开展, 但也容易让研究结果在实际部署和应用时面临较多的问题.

在攻击者能力方面, Tor 流量分类研究一般假设攻击者可以掌控客户端到入口节点的所有流量, 即攻击者为

用户的网关, ISP 或用户经过的入口节点; 而 Tor 关联研究则假设用户访问的流量入口与出口同时被攻击者检测到, 攻击者需要掌握很多 Tor 入口与出口节点, 或者是 AS 域级别的掌控者. 相比之下, 流关联研究对攻击者能力的要求更高, 应用更困难. 随着网络环境的复杂化与 Tor 网络规模的巨大化, 这些研究的可用性也将下降. 另一方面, Tor 流量在实际网络中占比极低, 因此目前的流量分析方法在实际部署时会浪费绝大多数资源在普通流量上, 这也导致了流量分析方法实用性的下降.

在用户假设方面, 当前的研究通常只研究有限行为下的用户流量. 例如封闭世界与开放世界是对用户访问内容集合的假设. Xu 等人[56]提出广世界的概念, 指出大多数工作中的开放世界假设完全不能模拟近乎无限的非监控空间. 单流或多流是另一种常用的用户假设. 多数流量分析工作假设用户只进行单流访问, 这显然与事实不符. Xu 等人[93]通过寻找分割点进行多标签网页分类的研究, 但这些方法通常只在用户连续访问两个网页的情况奏效, 面对更复杂的用况, 准确率将大大下降. 除此之外, 当前研究通常隐含地规定了用户 Tor 应用, 软件版本与网络环境[41], 这些默认限制也影响着流量分析方法的适用性. 在 Tor 流关联研究中, 受到收集 Tor 流量方法的限制, 大部分研究仍然无法在隐藏服务双端流上开展.

未来的 Tor 被动流量分析工作应该在逐渐弱化许多假设的同时, 考虑更接近现实的威胁模型. 在攻击者能力角度, 需要考虑的问题包括攻击者资源限制、攻击者目的、流关联中对手级别等; 在用户假设角度, 需要考虑的问题包括开放真实世界假设、单标签与多标签访问假设、普通网站与隐藏服务访问假设、网页与网站假设、用户网络环境等. 在方法上, 可以根据分析场景或问题推进应用场景的深化, 挖掘如分布式检测[149]、主动响应[150]、分层分析架构[151]等新型方法的潜力.

(2) 数据集稀缺. 虽然很多工作都公开了使用的数据集[152], 但目前的数据集仍然存在样本数量小、种类有限、数据集不通用等问题. 由于较慢的访问速度以及负载均衡机制, Tor 流量收集需要比普通加密流量收集高得多的时间和硬件成本. 同时, Tor 将所有流代理到了长时间会话的匿名通道中, 基于五元组的流聚合只能得到比真实访问行为少得多的 Tor 流. 另外, Tor 提供了多种流量混淆插件, Tor 网络中还存在多种流量类型, 以上原因使得各流量分析工作只能收集少量且零散的 Tor 流量. 目前, 除了少部分工作提供了原始 Pcap 外, 多数数据集仅提供提取后的流量特征, 而不同研究中使用的特征又往往各不相同, 这使得不同研究之间使用的数据集互不相同, 这进一步限制了许多研究的有效性. 在流关联研究中, 除了上述问题以外, 流关联数据集的收集还存在收集难、真实性低等问题. 首先, 大部分双端流量都通过基于代理的方式收集, 与真实的 Tor 流量相比, 增加代理对流量元数据的影响仍然是一个未解决的问题. 收集真实的出口流量还可能涉及道德上的风险. 另一方面, 目前的实验数据集仍然仅限于普通网站的访问, 尚未有收集隐藏服务双端流量的解决方案.

未来的工作需要解决 Tor 数据集稀缺的问题. 除了呼吁更多的研究者公开数据集以外, 还可以构建自动化 Tor 流量采集方法, 形成兼顾研究需要和隐私保护的流量采集规范. 既可以提高流量采集的效率, 也方便不同方法相互比较, 推进 Tor 流量分析研究的发展. 另一方面, 可以引入图像、普通加密流量分析等其他领域进行数据处理的方法, 如生成对抗网络[76]、小样本学习[98]等弥补 Tor 流量收集难、现有样本不平衡、体量小等问题.

(3) 基本比率谬误. 基本比率谬误 (base rate fallacy) 是指统计上对基本比率不敏感导致的推论谬误. 当检测目标显著少于非目标时, 即便使用非常高性能的分类器, 也会有大量非目标被错误地分类为目标, 导致分类结果无效. 例如, 一个 99.9% 精度的分类器每识别 1 000 条非目标流量就会将一条流量误分类为目标流量. 在实际部署中, 如果非目标流量是目标流量的 1 000 倍以上, 分类器误分类的结果就会多于正确分类的结果, 实际精度将低于 50%. 基本比率谬误问题广泛存在于流量分析研究中[153].

通过提升分类器精度或降低误报率可以显著改善基本比率谬误问题. 在流量分类方面, Wang 等人[38]和 Hayes 等人[40]通过改变训练样本分布和分类结果判断方法, 实现用假阳性率换取真阳性率, 从而提高最终分类结果的精度. Pulls 等人[154]和 Greschbach 等人[155]结合流量之外的信息 (如缺乏保护的 DNS 信息[156]) 降低流量误报率. 分类器给出分类结果后, 攻击者可以查询短时间内是否存在对应的 DNS 行为、广告实时竞价行为等, 确定分类是否准确. 然而, 这些工作并没有解决基本比率谬误问题, 只是具有提升分类结果精度的潜力. Wang[52]提出精度

优化器, 将分类为目标流量但不可靠的结果拒绝为非目标流量. 该策略会将许多目标流量归类为非目标, 导致召回率下降, 但保留了高可信的分类结果, 显著改善了基本比率谬误. 虽然如此, 该方法的最高精度仍然低于 90%, 同时召回率最低只有 20%. 基本比率谬误还存在于流量分类的其他方面, 例如, 在解决多标签网页分类问题时, 使用一个分类器遍历每个数据包, 判断该数据包是否是重叠流的开始. 在该任务中, 真实的分割点只有一个, 但是非分割点却有几百个, 即便分类器将所有点都识别为非分割点, 分割识别率仍然可以高达 99.76%. 这是基本比率谬误在识别分割点中的体现[58].

该问题还存在于 Tor 流关联中, 由于攻击的成对性质, 对 N 对 Tor 连接进行去匿名化需要进行 N^2 次比较, 其中只有 N 次结果是正确的检测目标. 这会导致流关联中的大量误报. Rimmer 等人[72]使用优化平均准确率 (average precision, AP) 替代传统的假阳性率 (FPR) 和真阳性率 (TPR) 来反映攻击者的能力, 因为 AP 综合权衡了准确率和召回率两方面因素, 对真正关联的流量对和非关联的流量对都设置了更公平的权重, 更全面地评估攻击的效果. Oh 等人[137]将流分割成多个窗口, 并在每个窗口内计算相似性, 经过投票来聚合结果. 通过调整窗口数和阈值, 可以将误报率降低到预期水平.

以上方法从提升分类器精度、降低误报、更改评估指标等方面缓解了基本比率谬误, 但目前几乎还没有工作真正解决了该问题. 由于 Tor 在普通流量中的极少占比, 基本比率谬误也存在 Tor 被动流量分析的多个方面. 在未来的研究中, 针对该问题的研究将有非常广阔的空间. 基本比率谬误长期存在于医学 (如癌症检测)、网络 (入侵检测) 等领域, 可以考虑借鉴其他领域解决基本比率谬误的方法提升 Tor 流量分析的性能.

(4) 概念漂移. 模型预测目标的统计特性可能随着时间、环境变化, 导致分类性能下降, 这种现象被称为概念漂移, 该现象广泛存在于流量分析、异常检测等领域[157]. 在 Tor 流量分析中, 引起概念漂移的原因包括 Tor 版本的更新、网站的更新、Tor 网络的改变、网络环境的变化等. 虽然可以通过重新训练模型来保持较高的分类性能, 但 Tor 训练数据难以收集, 模型训练成本高等问题极大地限制了该策略的可行性.

从流量中提取深度信息可以抵抗概念漂移. 例如基于 Snapshot 集成深度学习方法[59], 使用训练集的不同子集训练多个局部最优子模型, 深度提取流量知识, 可以有效抵抗概念漂移; Cherubin 等人[60]的工作表明, 对真实流量的学习可以提取深度信息, 减轻概念漂移的影响. 通过降低重新训练成本也可以抵抗概念漂移问题. 例如, Wang 等人[41]的方法仅需 6 800 个样本就可以保持对 100 个网站的高精度识别. Attarian 等人[158] 提出的基于自适应挖掘算法的自适应在线网站指纹攻击可以通过随时间更新模型来抵抗概念漂移. 一些工作引入小样本学习[98], 先从大量非目标流量中学习流量知识, 然后使用少量目标流量样本即可训练目标分类器. 在流关联中, 可以通过提升攻击模型的可迁移性来减轻概念漂移的影响. 例如, DeepCorr[75]使用 CNN 训练了一个为 Tor 网络定制的相关函数, 可以用于关联任意的电路和目标, 不受训练过程中使用的数据集的限制, 且不需要频繁重新训练. DeepCoFFEA[137]在不同的电路、网站和长达 14 个月的时间尺度上进行了迁移学习, 训练模型后, DeepCoFFEA 在不同的时间和环境中都具有较好的泛化能力, 且每年重新训练一次即可维持性能.

然而, 提取深度信息的方法只能让分析模型性能不至于下降过快, 而降低训练成本的方法虽然使得快速更新模型成为可能, 但这些方法的分类性能普遍低于其他的高性能方法, 同时这些方法仍然受到其他挑战问题的制约. 提升可迁移性的办法则往往需要较大规模的训练集, 且不能彻底解决概念漂移的问题. 因此, Tor 流量分析中的概念漂移问题仍然是一个极具挑战的问题. 在未来的 Tor 被动流量分析研究中, 需要综合考虑可能导致概念漂移的因素, 包括使用流量样本的代表性、均衡性, 用户网络环境如浏览器版本、Tor 版本、用户类型等, 同时研究能够抵御概念漂移的流量分析方法, 如深层特征提取方法, 增量学习模型等.

6 总 结

Tor 是当前最为广泛使用的匿名通信技术. Tor 被动流量分析技术具有成本低、准确率高等特点, 历经近二十年的发展, 已成为去匿名化技术最重要的研究方向. 该技术不仅可以用于实现犯罪行为的监控与检测, 威慑非法网络行为, 也不断挑战着当前 Tor 技术的安全性, 帮助改进隐私保护技术. 本文梳理了 Tor 被动流量分析的主要工作,

从 Tor 和流量分析的基本概念入手, 接着介绍了相关的威胁模型、Tor 去匿名化相关综述. 本文将 Tor 被动流量分析技术按照技术类型分为流量分类研究与流关联研究两种, 从流量分析流程出发, 进一步探讨了两种技术的流量采集方法、使用的特征以及算法选择. 最后, 本文对当前研究面临的主要问题进行了梳理与分析, 并给出了未来研究方向, 希望可以推动 Tor 被动流量分析的研究.

References:

[1] Ma CW, Zhang Y, Fang BX, Zhang HL. Survey on anonymous networks. Ruan Jian Xue Bao/Journal of Software, 2023, 34(1): 404–420 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6513.htm [doi: 10.13328/j.cnki.jos.006513]

[2] Luo JZ, Yang M, Ling Z, Wu WJ, Gu XD. Anonymous communication and darknet: A survey. Journal of Computer Research and Development, 2019, 56(1): 103–130 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2019.20180769]

[3] Zhao H, Wang LM, Shen TH, Huang L, Ni XL. Survey on anonymity metrics in communication network. Ruan Jian Xue Bao/Journal of Software, 2021, 32(1): 218–245 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6103.htm [doi: 10.13328/j.cnki.jos.006103]

[4] Chaum DL. Untraceable electronic mail, return addresses, and digital pseudonyms. Communications of the ACM, 1981, 24(2): 84–90. [doi: 10.1145/358549.358563]

[5] Danezis G, Dingledine R, Mathewson N. Mixminion: Design of a type III anonymous remailer protocol. In: Proc. of the 2003 Symp. on Security and Privacy. Berkeley: IEEE, 2003. 2–15. [doi: 10.1109/SECPRI.2003.1199323]

[6] I2P. The invisible Internet project. 2018. https://geti2p.net/zh/

[7] Clarke I, Sandberg O, Wiley B, Hong TW. Freenet: A distributed anonymous information storage and retrieval system. In: Proc. of the 2000 Int'l Workshop on Design Issues in Anonymity and Unobservability. Berkeley: Springer, 2000. 46–66. [doi: 10.1007/3-540-44702-4_4]

[8] Dingledine R, Mathewson N, Syverson PF. Tor: The second-generation onion router. In: Proc. of the 13th USENIX Security Symp. San Diego: USENIX, 2004. 303–320.

[9] Tor Project. Tor metrics. 2024. https://metrics.torproject.org/.

[10] Buxton O. Silk road: History+accessing the black market. LifeLock. 2024. https://lifelock.norton.com/learn/internet-security/silk-road

[11] van Wegberg R, Tajalizadehkhoob S, Soska K, Akyazi U, Gañán CH, Klievink B, Christin N, van Eeten M. Plug and Prey? Measuring the commoditization of cybercrime via online anonymous markets. In: Proc. of the 27th USENIX Security Symp. Baltimore: USENIX Association, 2018. 1009–1026.

[12] Sun YX, Edmundson A, Vanbever L, Li O, Rexford J, Chiang M, Mittal P. Raptor: Routing attacks on privacy in Tor. In: Proc. of the 24th USENIX Security Symp. Washington: USENIX Association, 2015. 271–286.

[13] Shen M, Ye K, Liu XT, Zhu LH, Kang JW, Yu S, Li Q, Xu K. Machine learning-powered encrypted network traffic analysis: A comprehensive survey. IEEE Communications Surveys & Tutorials, 2023, 25(1): 791–824. [doi: 10.1109/COMST.2022.3208196]

[14] Pacheco F, Exposito E, Gineste M, Baudoin C, Aguilar J. Towards the deployment of machine learning solutions in network traffic classification: A systematic survey. IEEE Communications Surveys & Tutorials, 2019, 21(2): 1988–2014. [doi: 10.1109/COMST.2018.2883147]

[15] Aminuddin MAIM, Zaaba ZF, Singh MKM, Singh DSM. A survey on Tor encrypted traffic monitoring. Int'l Journal of Advanced Computer Science and Applications (IJACSA), 2018, 9(8): 113–120. [doi: 10.14569/IJACSA.2018.090815]

[16] Salo J. Recent attacks on Tor. Aalto University, 2012. http://www.cse.hut.fi/en/publications/B/11/papers/salo.pdf

[17] Karunanayake I, Ahmed N, Malaney R, Islam R, Jha SK. De-anonymisation attacks on Tor: A survey. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2324–2350. [doi: 10.1109/COMST.2021.3093615]

[18] Yao ZJ, Ge JG, Zhang XD, Zheng HB, Zou Z, Sun KK, Xu ZH. Research review on traffic obfuscation and its corresponding identification and tracking technologies. Ruan Jian Xue Bao/Journal of Software, 2018, 29(10): 3205–3222 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5620.htm [doi: 10.13328/j.cnki.jos.005620]

[19] Johnson A, Wacek C, Jansen R, Sherr M, Syverson P. Users get routed: Traffic correlation on Tor by realistic adversaries. In: Proc. of the 2013 ACM SIGSAC Conf. on Computer & Communications Security. Berlin: ACM, 2013. 337–348. [doi: 10.1145/2508859.2516651]

[20] Tan QF, Wang XB, Shi W, Tang J, Tian ZH. An anonymity vulnerability in Tor. IEEE/ACM Trans. on Networking, 2022, 30(6): 2574–2587. [doi: 10.1109/TNET.2022.3174003]

[21] Edman M, Yener B. On anonymity in an electronic society: A survey of anonymous communication systems. ACM Computing Surveys,

2009, 42(1): 5. [doi: 10.1145/1592451.1592456]

[22] Erdin E, Zachor C, Gunes MH. How to find hidden users: A survey of attacks on anonymity networks. IEEE Communications Surveys & Tutorials, 2015, 17(4): 2296–2316. [doi: 10.1109/COMST.2015.2453434]

[23] Cambiaso E, Vaccari I, Patti L, Aiello M. Darknet security: A categorization of attacks to the Tor network. In: Proc. of the 3rd Italian Conf. on Cyber security. Pisa: CEUR-WS.org, 2019.

[24] Zhao N, Su JS, Zhao BK, Han B, Zou HC. A survey on hidden service location technologies in anonymous communication system. Chinese Journal of Computers, 2022, 45(2): 393–411 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2022.00373]

[25] Yang M, Luo JZ, Ling Z, Fu XW, Yu W. De-anonymizing and countermeasures in anonymous communication networks. IEEE Communications Magazine, 2015, 53(4): 60–66. [doi: 10.1109/MCOM.2015.7081076]

[26] Evers B, Hols J, Kula E, Schouten J, den Toom M, van der Laan RM, Pouwelse JA. Thirteen years of Tor attacks. 2016. https://github.com/Attacks-on-Tor/Attacks-on-Tor

[27] Lv B, Liao Y, Xie HY. Survey on attack technologies to Tor anonymous network. Journal of CAEIT, 2017, 12(1): 14–19 (in Chinese with English abstract). [doi: 10.3969/j.issn.1673-5692.2017.01.003]

[28] Basyoni L, Fetais N, Erbad A, Mohamed A, Guizani M. Traffic analysis attacks on Tor: A survey. In: Proc. of the 2020 IEEE Int'l Conf. on Informatics, IoT, and Enabling Technologies. Doha: IEEE, 2020. 183–188. [doi: 10.1109/ICIoT48696.2020.9089497]

[29] Sun XL, Huang AX, Luo XP, Xie Y. Webpage fingerprinting identification on Tor: A survey. Journal of Computer Research and Development, 2021, 58(8): 1773–1788 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2021.20200498]

[30] Wang RN, Zhao YF. A survey on anonymous communication systems traffic identification and classification. In: Proc. of the 3rd Int'l Conf. on Advanced Information Science and System. Sanya: ACM, 2021. 36. [doi: 10.1145/3503047.3503087]

[31] Zou HC, Su JS, Wei ZL, Zhao BK, Xia YS, Zhao N. A review of the research of website fingerprinting identification and defense. Chinese Journal of Computers, 2022, 45(10): 2243–2278 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2022.02243]

[32] Liu PD, He LT, Li ZJ. A survey on deep learning for website fingerprinting attacks and defenses. IEEE Access, 2023, 11: 26033–26047. [doi: 10.1109/ACCESS.2023.3253559]

[33] Pope S. Port-scanning resistance in Tor anonymity network. 2009. https://www.cs.utexas.edu/ftp/techreports/TR-1943.pdf

[34] Lapshichyov V, Makarevich O. Technology of deep packet inspection for recognition and blocking traffic of the Tor network. In: Proc. of the 7th Int'l Conf. on Security of Information and Networks. Sochi, 2019.

[35] AlSabah M, Bauer K, Goldberg I. Enhancing Tor's performance using real-time traffic classification. In: Proc. of the 2012 ACM Conf. on Computer and Communications Security. Raleigh: ACM, 2012. 73–84. [doi: 10.1145/2382196.2382208]

[36] Cai X, Zhang XC, Joshi B, Johnson R. Touching from a distance: Website fingerprinting attacks and defenses. In: Proc. of the 2012 ACM Conf. on Computer and Communications Security. Raleigh: ACM, 2012. 605–616. [doi: 10.1145/2382196.2382260]

[37] He GF, Yang M, Luo JZ, Zhang L. Online identification of Tor anonymous communication traffic. Ruan Jian Xue Bao/Journal of Software, 2013, 24(3): 540–556 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4253.htm [doi: 10.3724/SP.J.1001.2013.04253]

[38] Wang T, Cai X, Nithyanand R, Johnson R, Goldberg I. Effective attacks and provable defenses for website fingerprinting. In: Proc. of the 23rd USENIX Conf. on Security Symp. San Diego: USENIX Association, 2014. 143–157. [doi: 10.5555/2671225.2671235]

[39] Panchenko A, Lanze F, Pennekamp J, Engel T, Zinnen A, Henze M, Wehrle K. Website fingerprinting at Internet scale. In: Proc. of the 23rd Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2016. 1–15.

[40] Hayes J, Danezis G. k-fingerprinting: A robust scalable website fingerprinting technique. In: Proc. of the 25th USENIX Security Symp. Washington: USENIX Association, 2016. 1187–1203.

[41] Wang T, Goldberg I. On realistically attacking Tor with website fingerprinting. Proc. on Privacy Enhancing Technologies, 2016, 2016(4): 21–36. [doi: 10.1515/popets-2016-0027]

[42] Cuzzocrea A, Martinelli F, Mercaldo F, Vercelli G. Tor traffic analysis and detection via machine learning techniques. In: Proc. of the 2017 IEEE Int'l Conf. on Big Data. Boston: IEEE, 2017. 4474–4480. [doi: 10.1109/BigData.2017.8258487]

[43] Shahbar K, Zincir-Heywood AN. How far can we push flow analysis to identify encrypted anonymity network traffic? In: Proc. of the 2018 IEEE/IFIP Network Operations and Management Symp. Taipei: IEEE, 2018. 1–6. [doi: 10.1109/NOMS.2018.8406156]

[44] Rimmer V, Preuveneers D, Juarez M, van Goethem T, Joosen W. Automated website fingerprinting through deep learning. In: Proc. of the 25th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2018. 1–15.

[45] Sirinam P, Imani M, Juarez M, Wright M. Deep fingerprinting: Undermining website fingerprinting defenses with deep learning. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 1928–1943. [doi: 10.1145/3243734.3243768]

[46] Sirinam P, Mathews N, Rahman MS, Wright M. Triplet fingerprinting: More practical and portable website fingerprinting with n-shot learning. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security. London: ACM, 2019. 1131–1148. [doi: 10.1145/3319535.3354217]

[47] Bhat S, Lu D, Kwon A, Devadas S. Var-CNN: A data-efficient website fingerprinting attack based on deep learning. Proc. on Privacy Enhancing Technologies, 2019, 2019(4): 292–310. [doi: 10.2478/popets-2019-0070]

[48] Montieri A, Ciuonzo D, Aceto G, Pescapé A. Anonymity services Tor, I2p, JonDonym: Classifying in the dark (Web). IEEE Trans. on Dependable and Secure Computing, 2020, 17(3): 662–675. [doi: 10.1109/TDSC.2018.2804394]

[49] Montieri A, Ciuonzo D, Bovenzi G, Persico V, Pescapé A. A dive into the dark Web: Hierarchical traffic classification of anonymity tools. IEEE Trans. on Network Science and Engineering, 2020, 7(3): 1043–1054. [doi: 10.1109/TNSE.2019.2901994]

[50] Bovenzi G, Aceto G, Ciuonzo D, Persico V, Pescapé A. A big data-enabled hierarchical framework for traffic classification. IEEE Trans. on Network Science and Engineering, 2020, 7(4): 2608–2619. [doi: 10.1109/TNSE.2020.3009832]

[51] Hu YZ, Zou FT, Li LS, Yi P. Traffic classification of user behaviors in Tor, I2p, ZeroNet, freenet. In: Proc. of the 19th IEEE Int'l Conf. on Trust, Security and Privacy in Computing and Communications. Guangzhou: IEEE, 2020. 418–424. [doi: 10.1109/TrustCom50675.2020.00064]

[52] Wang T. High precision open-world website fingerprinting. In: Proc. of the 2020 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2020. 152–167. [doi: 10.1109/SP40000.2020.00015]

[53] Singh D, Shukla A, Sajwan M. Deep transfer learning framework for the identification of malicious activities to combat cyberattack. Future Generation Computer Systems, 2021, 125: 687–697. [doi: 10.1016/j.future.2021.07.015]

[54] Lin KD, Xu XL, Gao HH. TSCRNN: A novel classification scheme of encrypted traffic based on flow spatiotemporal features for efficient management of IIoT. Computer Networks, 2021, 190: 107974. [doi: 10.1016/j.comnet.2021.107974]

[55] Zhao RJ, Deng XW, Wang YH, Chen LB, Liu M, Xue Z, Wang YJ. Flow sequence-based anonymity network traffic identification with residual graph convolutional networks. In: Proc. of the 30th IEEE/ACM Int'l Symp. on Quality of Service. Oslo: IEEE, 2022. 1–10. [doi: 10.1109/IWQoS54832.2022.9812882]

[56] Xu SJ, Geng GG, Jin XB, Liu DJ, Weng J. Seeing traffic paths: Encrypted traffic classification with path signature features. IEEE Trans. on Information Forensics and Security, 2022, 17: 2166–2181. [doi: 10.1109/TIFS.2022.3179955]

[57] Lan JH, Liu XD, Li B, Li YA, Geng TT. DarknetSec: A novel self-attentive deep learning method for darknet traffic classification and application identification. Computers & Security, 2022, 116: 102663. [doi: 10.1016/j.cose.2022.102663]

[58] Yin QL, Liu ZT, Li Q, Wang T, Wang Q, Shen C, Xu YX. An automated multi-tab website fingerprinting attack. IEEE Trans. on Dependable and Secure Computing, 2022, 19(6): 3656–3670. [doi: 10.1109/TDSC.2021.3104869]

[59] Wang YB, Xu HT, Guo ZH, Qin Z, Ren K. snWF: Website fingerprinting attack by ensembling the snapshot of deep learning. IEEE Trans. on Information Forensics and Security, 2022, 17: 1214–1226. [doi: 10.1109/TIFS.2022.3158086]

[60] Cherubin G, Jansen R, Troncoso C. Online website fingerprinting: Evaluating website fingerprinting attacks on Tor in the real world. In: Proc. of the 31st USENIX Security Symp. Boston: USENIX Association, 2022. 753–770.

[61] Deng XH, Yin QL, Liu ZT, Zhao XY, Li Q, Xu MW, Xu K, Wu JP. Robust multi-tab website fingerprinting attacks in the wild. In: Proc. of the 2023 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2023. 1005–1022. [doi: 10.1109/SP46215.2023.10179464]

[62] Mathews N, Holland JK, Oh SE, Rahman MS, Hopper N, Wright M. SoK: A critical evaluation of efficient website fingerprinting defenses. In: Proc. of the 2023 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2023. 969–986. [doi: 10.1109/SP46215.2023.10179289]

[63] Karunanayake I, Jiang JJ, Ahmed N, Jha SK. Exploring uncharted waters of website fingerprinting. IEEE Trans. on Information Forensics and Security, 2024, 19: 1840–1854. [doi: 10.1109/TIFS.2023.3342607]

[64] Zhou Q, Wang LM, Zhu HJ, Lu T, Sheng VS. WF-Transformer: Learning temporal features for accurate anonymous traffic identification by using Transformer networks. IEEE Trans. on Information Forensics and Security, 2024, 19: 30–43. [doi: 10.1109/TIFS.2023.3318966]

[65] Bahramali A, Bozorgi A, Houmansadr A. Realistic website fingerprinting by augmenting network traces. In: Proc. of the 2023 ACM SIGSAC Conf. on Computer and Communications Security. Copenhagen: ACM, 2023. 1035–1049. [doi: 10.1145/3576915.3616639]

[66] Jin ZX, Lu TB, Luo S, Shang JZ. Transformer-based model for multi-tab website fingerprinting attack. In: Proc. of the 2023 ACM SIGSAC Conf. on Computer and Communications Security. Copenhagen: ACM, 2023. 1050–1064. [doi: 10.1145/3576915.3623107]

[67] Chakravarty S, Stavrou A, Keromytis AD. Traffic analysis against low-latency anonymity networks using available bandwidth estimation. In: Proc. of the 15th European Symp. on Research in Computer Security. Athens: Springer, 2010. 249–267. [doi: 10.1007/

978-3-642-15497-3_16]

[68] Bauer KS, Sherr M, Grunwald D. ExperimenTor: A testbed for safe and realistic Tor experimentation. In: Proc. of the 4th Workshop on Cyber Security Experimentation and Test. San Francisco: USENIX Association, 2011. 1–8.

[69] Jansen R, Hopper N. Shadow: Running Tor in a box for accurate and efficient experimentation. In: Proc. of the 19th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2012. 1–22.

[70] Jansen R, Juarez M, Gálvez R, Elahi T, Díaz C. Inside job: Applying traffic analysis to measure Tor from within. In: Proc. of the 25th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2018.

[71] Jia LY, Liu Y, Wang BL, Liu HR, Xin GD. A hierarchical classification approach for Tor anonymous traffic. In: Proc. of the 9th IEEE Int'l Conf. on Communication Software and Networks. Guangzhou: IEEE, 2017. 239–243. [doi: 10.1109/ICCSN.2017.8230113]

[72] Rimmer V, Schnitzler T, van Goethem T, Romero AR, Joosen W, Kohls K. Trace oddity: Methodologies for data-driven traffic analysis on Tor. Proc. on Privacy Enhancing Technologies, 2022, 2022(3): 314–335. [doi: 10.56553/popets-2022-0074]

[73] Habibi Lashkari A, Draper Gil G, Mamun MSI, Ghorbani AA. Characterization of Tor traffic using time based features. In: Proc. of the 3rd Int'l Conf. on Information Systems Security and Privacy. Porto: SciTePress, 2017. 253–262.

[74] Shahbar K, Zincir-Heywood AN. Anon17: Network traffic dataset of anonymity services. Technical Report, Faculty of Computer Science Dalhousie University, 2017.

[75] Nasr M, Bahramali A, Houmansadr A. DeepCorr: Strong flow correlation attacks on Tor using deep learning. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 1962–1976. [doi: 10.1145/3243734.3243824]

[76] Lashkari AH, Kaur G, Rahali A. DiDarknet: A contemporary approach to detect and characterize the darknet traffic using deep image learning. In: Proc. of the 10th Int'l Conf. on Communication and Network Security. Tokyo: ACM, 2020. 1–13. [doi: 10.1145/3442520.3442521]

[77] Rahman MS, Sirinam P, Mathews N, Gangadhara KG, Wright M. Tik-Tok: The utility of packet timing in website fingerprinting attacks. Proc. on Privacy Enhancing Technologies, 2020, 2020(3): 5–24. [doi: 10.2478/popets-2020-0043]

[78] Bernaille L, Teixeira R, Akodkenou I, Soule A, Salamatian K. Traffic classification on the fly. ACM SIGCOMM Computer Communication Review, 2006, 36(2): 23–26. [doi: 10.1145/1129582.1129589]

[79] Bernaille L, Teixeira R, Salamatian K. Early application identification. In: Proc. of the 2006 ACM CoNEXT Conf. Lisbon: ACM, 2006. Article No. 6. [doi: 10.1145/1368436.1368445]

[80] Marnerides AK, Schaeffer-Filho A, Mauthe A. Traffic anomaly diagnosis in Internet backbone networks: A survey. Computer Networks, 2014, 73: 224–243. [doi: 10.1016/j.comnet.2014.08.007]

[81] Wang LM, Mei HT, Sheng VS. Multilevel identification and classification analysis of Tor on mobile and PC platforms. IEEE Trans. on Industrial Informatics, 2021, 17(2): 1079–1088. [doi: 10.1109/TII.2020.2988870]

[82] Moore A, Zuev D, Crogan M. Discriminators for use in flow-based classification. Technical Report RR-05-13. Queen Mary University of London, 2013.

[83] Tranalyzer2. 2023. https://tranalyzer.com

[84] Ma XB, Shi MW, An BY, Li JF, Luo DX, Zhang JJ, Guan XH. Context-aware website fingerprinting over encrypted proxies. In: Proc. of the 2021 IEEE Conf. on Computer Communications. Vancouver: IEEE, 2021. 1–10. [doi: 10.1109/INFOCOM42981.2021.9488676]

[85] Oh SE, Li S, Hopper N. Fingerprinting keywords in search queries over Tor. Proc. on Privacy Enhancing Technologies, 2017, 2017(4): 251–270. [doi: 10.1515/popets-2017-0048]

[86] Shahbar K, Zincir-Heywood AN. Packet momentum for identification of anonymity networks. Journal of Cyber Security and Mobility, 2017, 16(1): 27–56. [doi: 10.13052/2245-1439.612]

[87] Yang M, Gu XD, Ling Z, Yin CX, Luo JZ. An active de-anonymizing attack against Tor Web traffic. Tsinghua Science and Technology, 2017, 22(6): 702–713. [doi: 10.23919/TST.2017.8195352]

[88] Shen M, Liu YT, Zhu LH, Xu K, Du XJ, Guizani N. Optimizing feature selection for efficient encrypted traffic classification: A systematic approach. IEEE Network, 2020, 34(4): 20–27. [doi: 10.1109/MNET.011.1900366]

[89] Granitto PM, Furlanello C, Biasioli F, Gasperi F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemometrics and Intelligent Laboratory Systems, 2006, 83(2): 83–90. [doi: 10.1016/j.chemolab.2006.01.007]

[90] Yan JH, Kaur J. Feature selection for website fingerprinting. Proc. on Privacy Enhancing Technologies, 2018, 2018(4): 200–219. [doi: 10.1515/popets-2018-0039]

[91] Gama J. Functional trees for classification. In: Proc. of the 2001 IEEE Int'l Conf. on Data Mining. San Jose: IEEE, 2001. 147–154. [doi: 10.1109/ICDM.2001.989512]

[92] Landwehr N, Hall M, Frank E. Logistic model trees. Machine Learning, 2005, 59(1–2): 161–205. [doi: 10.1007/s10994-005-0466-3]

[93] Xu YX, Wang T, Li Q, Gong QY, Chen Y, Jiang Y. A multi-tab website fingerprinting attack. In: Proc. of the 34th Annual Computer Security Applications Conf. San Juan: ACM, 2018. 327–341. [doi: 10.1145/3274694.3274697]

[94] Navarro G. A guided tour to approximate string matching. ACM Computing Surveys, 2001, 33(1): 31–88. [doi: 10.1145/375360.375365]

[95] Luo XP, Zhou P, Chan EWW, Lee W, Chang RKC, Perdisci R. HTTPOS: Sealing information leaks with browser-side obfuscation of encrypted flows. In: Proc. of the 2011 Network and Distributed System Security Symp. San Diego: The Internet Society, 2011.

[96] Tor. Experimental defense for website traffic fingerprinting. 2023. https://blog.torproject.org/blog/experimental-defense-website-traffic-fingerprinting

[97] Abe K, Goto S. Fingerprinting attack on Tor anonymity using deep learning. Proc. of the 2016 Asia-Pacific Advanced Network, 2016, 42: 15–20.

[98] Chen MT, Zhu XJ, Xu HZ, Wang YT. Few-shot website fingerprinting attack. Computer Networks, 2021, 198: 108298. [doi: 10.1016/j.comnet.2021.108298]

[99] Bai XF, Zhang Y, Niu XM. Traffic identification of Tor and Web-mix. In: Proc. of the 8th Int'l Conf. on Intelligent Systems Design and Applications. Kaohsuing: IEEE, 2008. 548–551. [doi: 10.1109/ISDA.2008.209]

[100] Tor. Obfs4. 2023. https://gitWeb.torproject.org/pluggable-transports/obfs4.git/tree/doc/obfs4-spec.txt

[101] He YZ, Hu LP, Gao R. Detection of Tor traffic hiding under Obfs4 protocol based on two-level filtering. In: Proc. of the 2nd Int'l Conf. on Data Intelligence and Security. South Padre Island: IEEE, 2019. 195–200. [doi: 10.1109/ICDIS.2019.00036]

[102] Liang D, He YZ. Obfs4 traffic identification based on multiple-feature fusion. In: Proc. of the 2020 IEEE Int'l Conf. on Power, Intelligent Computing and Systems. Shenyang: IEEE, 2020. 323–327. [doi: 10.1109/ICPICS50287.2020.9202018]

[103] Fifield D, Lan C, Hynes R, Wegmann P, Paxson V. Blocking-resistant communication through domain fronting. Proc. on Privacy Enhancing Technologies, 2015, 2015(2): 46–64. [doi: 10.1515/popets-2015-0009]

[104] Yao ZJ, Ge JG, Wu YL, Zhang XD, Li Q, Zhang L, Zou Z. Meek-based Tor traffic identification with hidden markov model. In: Proc. of the 20th IEEE Int'l Conf. on High Performance Computing and Communications; IEEE the 16th Int'l Conf. on Smart City; IEEE the 4th Int'l Conf. on Data Science and Systems. Exeter: IEEE, 2018. 335–340. [doi: 10.1109/HPCC/SmartCity/DSS.2018.00075]

[105] Fifield D. Threat modeling and circumvention of Internet censorship [Ph.D. Thesis]. Berkeley: University of California, 2017.

[106] Chen JQ, Cheng G, Mei HT. F-ACCUMUL: A protocol fingerprint and accumulative payload length sample-based Tor-snowflake traffic-identifying framework. Applied Sciences, 2023, 13(1): 622. [doi: 10.3390/app13010622]

[107] Hasselquist D, Lindblom M, Carlsson N. Lightweight fingerprint attack and encrypted traffic analysis on news articles. In: Proc. of the 2022 IFIP Networking Conf. Catania: IEEE, 2022. 1–9. [doi: 10.23919/IFIPNetworking55013.2022.9829796]

[108] Shen M, Ji KX, Gao ZB, Li Q, Zhu LH, Xu K. Subverting website fingerprinting defenses with robust traffic representation. In: Proc. of the 32nd USENIX Security Symp. Anaheim: USENIX Association, 2023. 607–624.

[109] Dyer KP, Coull SE, Ristenpart T, Shrimpton T. Peek-a-boo, I still see you: Why efficient traffic analysis countermeasures fail. In: Proc. of the 2012 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2012. 332–346. [doi: 10.1109/SP.2012.28]

[110] Cai X, Nithyanand R, Wang T, Johnson R, Goldberg I. A systematic approach to developing and evaluating website fingerprinting defenses. In: Proc. of the 2014 ACM SIGSAC Conf. on Computer and Communications Security. Scottsdale: ACM, 2014. 227–238. [doi: 10.1145/2660267.2660362]

[111] Juarez M, Imani M, Perry M, Diaz C, Wright M. Toward an efficient website fingerprinting defense. In: Proc. of the 21st European Symp. on Research in Computer Security. Heraklion: Springer, 2016. 27–46. [doi: 10.1007/978-3-319-45744-4_2]

[112] Al-Naami K, El-Ghamry A, Islam S, Khan L, Thuraisingham BM, Hamlen KW, Alrahmawy M, Rashad MZ. BiMorphing: A Bi-directional bursting defense against website fingerprinting attacks. IEEE Trans. on Dependable and Secure Computing, 2021, 18(2): 505–517. [doi: 10.1109/TDSC.2019.2907240]

[113] Gong JJ, Wang T. Zero-delay lightweight defenses against website fingerprinting. In: Proc. of the 29th USENIX Security Symp. USENIX Association, 2020. 717–734.

[114] Witwer E, Holland JK, Hopper N. Padding-only defenses add delay in Tor. In: Proc. of the 21st Workshop on Privacy in the Electronic Society. Los Angeles: ACM, 2022. 29–33. [doi: 10.1145/3559613.3563207]

[115] de la Cadena W, Mitseva A, Hiller J, Pennekamp J, Reuter S, Filter J, Engel T, Wehrle K, Panchenko A. TrafficSliver: Fighting website fingerprinting attacks with traffic splitting. In: Proc. of the 2020 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2020. 1971–1985. [doi: 10.1145/3372297.3423351]

[116] Abolfathi M, Shomorony I, Vahid A, Jafarian JH. A game-theoretically optimal defense paradigm against traffic analysis attacks using multipath routing and deception. In: Proc. of the 27th ACM on Symp. on Access Control Models and Technologies. New York: ACM,

2022. 67–78. [doi: 10.1145/3532105.3535015]

[117] Ling Z, Xiao G, Wu WJ, Gu XD, Yang M, Fu XW. Towards an efficient defense against deep learning based website fingerprinting. In: Proc. of the 2022 IEEE Conf. on Computer Communications. London: IEEE, 2022. 310–319. [doi: 10.1109/INFOCOM48880.2022.9796685]

[118] Gong JJ, Zhang WQ, Zhang C, Wang T. Surakav: Generating realistic traces for a strong website fingerprinting defense. In: Proc. of the 2022 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2022. 1558–1573. [doi: 10.1109/SP46214.2022.9833722]

[119] Wang T, Ian G. Walkie-talkie: An efficient defense against passive website fingerprinting attacks. In: Proc. of the 26th USENIX Security Symp. Vancouver: USENIX Association, 2017. 1375–1390.

[120] Perry M. Experimental defense for website traffic fingerprinting. 2013. https://blog.torproject.org/blog/critique-website-traffic-fingerprinting-attacks

[121] Juarez M, Afroz S, Acar G, Diaz C, Greenstadt R. A critical evaluation of website fingerprinting attacks. In: Proc. of the 2014 ACM SIGSAC Conf. on Computer and Communications Security. Scottsdale: ACM, 2014. 263–274. [doi: 10.1145/2660267.2660368]

[122] Zhuo ZL, Zhang Y, Zhang ZL, Zhang XS, Zhang JZ. website fingerprinting attack on anonymity networks based on profile hidden Markov model. IEEE Trans. on Information Forensics and Security, 2018, 13(5): 1081–1095. [doi: 10.1109/TIFS.2017.2762825]

[123] Mani A, Wilson-Brown T, Jansen R, Johnson A, Sherr M. Understanding Tor usage with privacy-preserving measurement. In: Proc. of the 2018 Internet Measurement Conf. Boston: ACM, 2018. 175–187. [doi: 10.1145/3278532.3278549]

[124] Murdoch SJ, Danezis G. Low-cost traffic analysis of Tor. In: Proc. of the 2005 IEEE Symp. on Security and Privacy. Oakland: IEEE, 2005. 183–195. [doi: 10.1109/SP.2005.12]

[125] Edman M, Syverson P. As-awareness in Tor path selection. In: Proc. of the 16th ACM Conf. on Computer and Communications Security. Chicago: ACM, 2009. 380–389. [doi: 10.1145/1653662.1653708]

[126] O'Gorman G, Blott S. Improving stream correlation attacks on anonymous networks. In: Proc. of the 2009 ACM Symp. on Applied Computing. Honolulu: ACM, 2009. 2024–2028. [doi: 10.1145/1529282.1529732]

[127] Mittal P, Khurshid A, Juen J, Caesar M, Borisov N. Stealthy traffic analysis of low-latency anonymous communication using throughput fingerprinting. In: Proc. of the 18th ACM Conf. on Computer and Communications Security. Chicago: ACM, 2011. 215–226. [doi: 10.1145/2046707.2046732]

[128] Akhoondi M, Yu C, Madhyastha HV. LAStor: A low-latency as-aware Tor client. IEEE/ACM Trans. on Networking, 2014, 22(6): 1742–1755. [doi: 10.1109/TNET.2013.2291242]

[129] Nithyanand R, Starov O, Gill P, Zair A, Schapira M. Measuring and mitigating as-level adversaries against Tor. In: Proc. of the 23rd Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2016.

[130] Barton A, Wright M. Denasa: Destination-naive as-awareness in anonymous communications. Proc. on Privacy Enhancing Technologies, 2016, 2016(4): 356–372. [doi: 10.1515/popets-2016-0044]

[131] Sun YX, Edmundson A, Feamster N, Chiang M, Mittal P. Counter-raptor: Safeguarding Tor against active routing attacks. In: Proc. of the 2017 IEEE Symp. on Security and Privacy. San Jose: IEEE, 2017. 977–992. [doi: 10.1109/SP.2017.34]

[132] Nasr M, Houmansadr A, Mazumdar A. Compressive traffic analysis: A new paradigm for scalable traffic analysis. In: Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security. Dallas: ACM, 2017. 2053–2069. [doi: 10.1145/3133956.3134074]

[133] Johnson A, Jansen R, Jaggard AD, Feigenbaum J, Syverson PF. Avoiding the man on the wire: Improving Tor's security with trust-aware path selection. In: Proc. of the 24th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2017.

[134] Wails R, Sun YX, Johnson A, Chiang M, Mittal P. Tempest: Temporal dynamics in anonymity systems. Proc. on Privacy Enhancing Technologies, 2018, 2018(3): 22–42. [doi: 10.1515/popets-2018-0019]

[135] Guan Z, Xiong G, Li Z, Gou GP. ResTor: A pre-processing model for removing the noise pattern in flow correlation. In: Proc. of the 2020 IEEE Symp. on Computers and Communications. Rennes: IEEE , 2020. 1–6. [doi: 10.1109/ISCC50000.2020.9219544]

[136] Palmieri F. A distributed flow correlation attack to anonymizing overlay networks based on wavelet multi-resolution analysis. IEEE Trans. on Dependable and Secure Computing, 2021, 18(5): 2271–2284. [doi: 10.1109/TDSC.2019.2947666]

[137] Oh SE, Yang TJ, Mathews N, Holland JK, Rahman MS, Hopper N, Wright M. DeepCoFFEA: Improved flow correlation attacks on Tor via metric learning and amplification. In: Proc. of the 2022 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2022. 1915–1932. [doi: 10.1109/SP46214.2022.9833801]

[138] Zhang ZW, Ye DP. Defending against deep-learning-based flow correlation attacks with adversarial examples. Security and Communication Networks, 2022, 2022: 2962318. [doi: 10.1155/2022/2962318]

[139] Lopes D, Dong JD, Medeiros P, Castro D, Barradas D, Portela B, Vinagre J, Ferreira B, Christin N, Santos N. Flow correlation attacks

on tor onion service sessions with sliding subset sum. In: Proc. of the 2024 Network and Distributed System Security (NDSS) Symp. San Diego: NDSS, 2024.

[140] Tan H, Sherr M, Zhou WC. Data-plane defenses against routing attacks on Tor. Proc. on Privacy Enhancing Technologies, 2016, 2016(4): 276–293. [doi: 10.1515/popets-2016-0040]

[141] Juen J, Johnson A, Das A, Borisov N, Caesar M. Defending Tor from network adversaries: A case study of network path prediction. Proc. on Privacy Enhancing Technologies, 2015, 2015(2): 171–187. [doi: 10.1515/popets-2015-0021]

[142] DeepCorr dataset. 2018. https://people.cs.umass.edu/amir/FlowCorrelation.html

[143] Trace Oddity dataset. 2022. https://distrinet.cs.kuleuven.be/software/tor-tc-dl/

[144] DeepCoFFEA dataset. 2022. https://github.com/traffic-analysis/deepcoffea

[145] Chakravarty S, Barbera MV, Portokalidis G, Polychronakis M, Keromytis AD. On the effectiveness of traffic analysis against anonymity networks using flow records. In: Proc. of the 15th Int'l Conf. on Passive and Active Network Measurement. Los Angeles: Springer, 2014. 247–257. [doi: 10.1007/978-3-319-04918-2_24]

[146] Zhu Y, Fu XW, Graham B, Bettati R, Zhao W. On flow correlation attacks and countermeasures in mix networks. In: Proc. of the 4th Int'l Workshop on Privacy Enhancing Technologies. Toronto: Springer, 2004. 207–225. [doi: 10.1007/11423409_13]

[147] Zhang Y, Paxson V. Detecting stepping stones. In: Proc. of the 9th USENIX Security Symp. Denver: USENIX Association, 2000.

[148] Tian J, Gou GP, Guan YY, Xia W, Xiong G, Liu C. Universal perturbation for flow correlation attack on Tor. In: Proc. of the 2021 IEEE Int'l Performance, Computing, and Communications Conf. Austin: IEEE, 2021. 1–9. [doi: 10.1109/IPCCC51483.2021.9679433]

[149] Dong F, Wang L, Nie X, Shao F, Wang HY, Li D, Luo XP, Xiao XS. DISTDET: A cost-effective distributed cyber threat detection system. In: Proc. of the 32nd USENIX Security Symp. Anaheim: USENIX Association, 2023. 1–18.

[150] Wei F, Li HD, Zhao ZM, Hu HX. xNIDS: Explaining deep learning-based network intrusion detection systems for active intrusion responses. In: Proc. of the 32nd USENIX Conf. on Security Symp. Anaheim: USENIX Association, 2023. 243. [doi: 10.5555/3620237.3620480]

[151] Qu J, Ma XB, Li JF, Luo XP, Xue L, Zhang JJ, Li ZH, Feng L, Guan XH. An input-agnostic hierarchical deep learning framework for traffic fingerprinting. In: Proc. of the 32nd USENIX Conf. on Security Symp. Anaheim: USENIX Association, 2023. 34. [doi: 10.5555/3620237.3620271]

[152] Cullen D, Halladay J, Briner N, Basnet R. CMU-SYNTRAFFIC-2022. IEEE Dataport. [doi: 10.21227/wc3q-jz97]

[153] Bozorgi A, Bahramali A, Rezaei F, Ghafari A, Houmansadr A, Soltani R, Goeckel D, Towsley D. I still know what you did last summer: Inferring sensitive user activities on messaging applications through traffic analysis. IEEE Trans. on Dependable and Secure Computing, 2023, 20(5): 4135–4153. [doi: 10.1109/TDSC.2022.3218191]

[154] Pulls T, Dahlberg R. website fingerprinting with website oracles. Proc. on Privacy Enhancing Technologies, 2020, 2020(1): 235–255. [doi: 10.2478/popets-2020-0013]

[155] Greschbach B, Pulls T, Roberts LM, Winter P, Feamster N. The effect of DNS on Tor's anonymity. In: Proc. of the 24th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2017.

[156] Hilton A, Deccio CT, Davis J. Fourteen years in the life: A root server's perspective on DNS resolver security. In: Proc. of the 32nd USENIX Security Symp. Anaheim: USENIX Association, 2023. 3171–3186.

[157] Han DQ, Wang ZL, Chen WQ, Wang K, Yu R, Wang S, Zhang H, Wang ZH, Jin MH, Yang JH, Shi XG, Yin X. Anomaly detection in the open world: Normality shift detection, explanation, and adaptation. In: Proc. of the 30th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2023.

[158] Attarian R, Abdi L, Hashemi S. AdaWFPA: Adaptive online website fingerprinting attack for Tor anonymous network: A stream-wise paradigm. Computer Communications, 2019, 148: 74–85. [doi: 10.1016/j.comcom.2019.09.008]

附中文参考文献:

[1] 马传旺, 张宇, 方滨兴, 张宏莉. 匿名网络综述. 软件学报, 2023, 34(1): 404–420. http://www.jos.org.cn/1000-9825/6513.htm [doi: 10.13328/j.cnki.jos.006513]

[2] 罗军舟, 杨明, 凌振, 吴文甲, 顾晓丹. 匿名通信与暗网研究综述. 计算机研究与发展, 2019, 56(1): 103–130. [doi: 10.7544/issn1000-1239.2019.20180769]

[3] 赵蕙, 王良民, 申屠浩, 黄磊, 倪晓铃. 网络匿名度量研究综述. 软件学报, 2021, 32(1): 218–245. http://www.jos.org.cn/1000-9825/6103.htm [doi: 10.13328/j.cnki.jos.006103]

[18] 姚忠将, 葛敬国, 张潇丹, 郑宏波, 邹壮, 孙焜焜, 许子豪. 流量混淆技术及相应识别、追踪技术研究综述. 软件学报, 2018, 29(10):

3205–3222. http://www.jos.org.cn/1000-9825/5620.htm [doi: 10.13328/j.cnki.jos.005620]

[24] 赵娜, 苏金树, 赵宝康, 韩彪, 邹鸿程. 匿名通信系统隐藏服务定位技术研究综述. 计算机学报, 2022, 45(2): 393–411. [doi: 10.11897/SP.J.1016.2022.00373]

[27] 吕博, 廖勇, 谢海永. Tor 匿名网络攻击技术综述. 中国电子科学研究院学报, 2017, 12(1): 14–19. [doi: 10.3969/j.issn.1673-5692.2017.01.003]

[29] 孙学良, 黄安欣, 罗夏朴, 谢怡. 针对 Tor 的网页指纹识别研究综述. 计算机研究与发展, 2021, 58(8): 1773–1788. [doi: 10.7544/issn1000-1239.2021.20200498]

[31] 邹鸿程, 苏金树, 魏子令, 赵宝康, 夏雨生, 赵娜. 网站指纹识别与防御研究综述. 计算机学报, 2022, 45(10): 2243–2278. [doi: 10.11897/SP.J.1016.2022.02243]

[37] 何高峰, 杨明, 罗军舟, 张璐. Tor 匿名通信流量在线识别方法. 软件学报, 2013, 24(3): 540–556. http://www.jos.org.cn/1000-9825/4253.htm [doi: 10.3724/SP.J.1001.2013.04253]

梅汉涛(1994－), 男, 博士生, 主要研究领域为流量分析, 网络测量, 网络态势感知.

朱怡霖(2000－), 女, 硕士生, 主要研究领域为流量分析, 网络态势感知.

程光(1973－), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为网络空间安全监测和防护, 网络大数据分析.

周余阳(1994－), 男, 博士, 助理研究员, CCF 专业会员, 主要研究领域为网络测量, 网络安全主动防御.