网站指纹识别与防御研究综述

邹鸿程[1)]苏金树[1)]魏子令[1)]赵宝康[1)]夏雨生[1)]赵娜[1),2)]

[1)](国防科技大学计算机学院,长沙 410073)

[2)](长沙师范学院信息科学与工程学院,长沙 410100)

摘　要　随着互联网的普及，人们越来越多地通过浏览网站获取消息、社交娱乐或者从事商业活动。用户的浏览兴趣往往暴露了个人的隐私。为了保护自己的浏览隐私，人们开始通过集成了隐私增强技术的网络来访问网站。然而，网站指纹识别与防御的研究成果表明通过隐私增强技术来保护用户访问网站的隐私已不再安全。因此，该研究引发了学术界和产业界的广泛关注。为此，本文以系统化网站指纹研究为目标和牵引，围绕网站指纹研究的最新成果，首先概述了网站指纹研究的基本概念、识别假设、威胁模型、防御模型和研究意义，随后分两节分别详细论述了网站指纹识别和防御的评价指标、分类方法和主要研究成果。在此基础上，本文对网站指纹识别和防御的相关研究分为主体研究和辅助研究两个方面进行论述，置重点于主体研究部分。具体地，在识别方向，对其主体研究即识别方法按照相似度判别法、传统机器学习方法和深度学习方法等三种类别展开阐述，再进一步细分为具体的数学模型进行讨论分析；在防御方向，对其主体研究即防御方法则按照网络层防御、应用层防御和复合层防御进行划分，然后根据各种防御方法具体使用的防御模式进行二次分类，并进行比较综述。最后，本文提出了网站指纹研究面临的三大挑战，并对未来的研究方向进行了展望。

关键词网络空间安全；网站指纹；流量分析；隐私保护；应用安全

中图法分类号　TP391

A Review of the Research of Website Fingerprinting Identification and Defense

ZOUHong-Cheng[1)]SUJin-Shu[1)]WEIZi-Ling[1)]ZHAOBao-Kang[1)] XIAYu-Sheng[1)] ZHAONa[1),2)]

[1)](College of Computer, National University of Defense Technology, Changsha 410073)

[2)](Department of Information Science and Technology, Changsha Normal University, Changsha 410100)

Abstract　With the popularity of the Internet, more and more people browse websites to obtain news, social entertainment or engage in business activities. The user's browsing interest often exposes personal privacy. To protect their browsing privacy, people begin to visit websites through a network integrated with privacy-enhancing technologies. Under this background, the research of website fingerprinting is proposed and studied widely. It includes two perspectives, namely website fingerprinting identification and defense. Website fingerprinting identification is a kind of traffic analysis technique. It enables a local adversary to identify a target user’s browsing websites without decrypting packets or modifying the traffic. In a general process of identification, the adversary first collects the traffic of the target user, then extracts a set of alternative traffic features, such as packet lengths, packet directions, packet statistics, and so on. The selected features together with the labels (i.e., true websites) are input into a given mathematical model to learn the optimal parameters automatically. After training, the model can

本课题得到国家重点研发计划项目（2018YFB0204301）、国家自然科学基金面上项目（61972412）、湖南省科技创新计划（2020RC2047）、湖南省教育厅科学研究一般项目（19C0140）和长沙师范学院校级课题（2019xjzkpy16）资助.**邹鸿程**，博士研究生，主要研究领域为网络空间安全、机器学习.**苏金树**，博士，研究员，博士生导师，主要研究领域为网络空间安全、因特网体系结构.**魏子令（通信作者）**，博士，助理研究员，主要研究领域为网络空间安全、无线通信和边缘计算. E-mail: weiziling@nudt.edu.cn.**赵宝康**，博士，副教授，中国计算机学会（CCF）高级会员，主要研究领域为网络空间安全、计算机网络.**夏雨生**，博士研究生，主要研究领域为网络空间安全、匿名通信系统.**赵娜**，博士研究生，主要研究领域为网络空间安全、匿名通信系统.

predict the labels of the unknown traffic with high accuracy by similarly extracting its features like the training phase. To date, researchers have proposed a large number of identification methods by combining different models with various kinds of features. To lessen the threat of website fingerprinting identification, researchers try to obfuscate the key traffic features in various ways, including packet padding, link padding, traffic splitting, and so on. The kind of research is called website fingerprinting defense in the literature. Obviously, the defense methods will introduce extra bandwidth or time overhead to a different degree, even requiring modifying the code of the explorer or web server. These great limit their application in reality. Nevertheless, these defense methods enrich the variety of privacy-enhancing technologies. Moreover, some of them are the candidates to be integrated into Tor. However, none of them can prevent all kinds of website fingerprinting identifications. In all, existing research fruits of website fingerprinting identification and defense show that it is no longer safe to protect the privacy of users visiting the website through privacy-enhancing technologies. Therefore, this research has aroused widespread concern in both academia and industry. To this end, we take systematic website fingerprinting research as the goal and guidance, focusing on its latest achievements. First, it outlines the basic concepts, identification hypotheses, threat models, defense models, and research significance of website fingerprinting, then elaborates the evaluation indicators, classification methods, and main research achievements of website fingerprinting identification and defense in two sections. On this basis, we discuss the related research on website fingerprinting identification and defense into two aspects: main research and auxiliary research, focusing on the former. Specifically, in terms of website fingerprinting identification, the main research, namely, the identification method, is expounded in three categories: similarity discrimination method, traditional machine learning method, and deep learning method, and then further subdivided into specific mathematical models for discussion and analysis. Regarding the auxiliary research, we summarize six related research of website fingerprinting identification except for identification method, such as, how to improve the identification accuracy in scenarios where identification assumptions are weakened,how to solve the "base rate fallacy" problem, and so on. In terms of website fingerprinting defense, its main research, namely defense methods, is divided into network-layer defense, application-layer defense, and composite-layer defense, then classified according to the specific defense modes used by various defense methods, and a comparative review is done. Regarding the auxiliary research, we introduce four related research of website fingerprinting defense, e.g.,security and bandwidth analysis, feature information leakage analysis, and so forth. Finally, we put forward the three significant challenges faced by website fingerprinting research and explore some future research trends.

Key words cyberspace security; website fingerprinting; traffic analysis; privacy protection; application security

1 引言

随着互联网的不断发展和广泛应用，网络成为人们日常工作生活必不可少的组成部分。截至 2020 年 6 月，全球约有 48 亿网民，互联网普及率是 62%，中国网民达到 13.19 亿[①]。早期，人们通过未加安全防护的网络直接浏览网页、收发邮件或聊天通信，不可避免地出现严重的安全隐患。概括地，主要的网络安全问题包括窃听风险(Eavesdropping)、篡改风险(Tampering)和冒充风险(Masquerading)等。窃听风险是指第三方可以截获并提取通信内容；篡改风险是指第三方可以修改并重放通信内容；冒充风险是指第三方可以伪造通信双方身份进行通信。为了更好的保护个人隐私，各种隐私增强技术(Privacy Enhancing Technologies, PETs)应运而生，如图 1 所示。

最初，互联网服务提供商(Internet Service Provider, ISP)通过支持 TLS/SSL 加密连接，为流量载荷数据提供了保护，大大降低了上述三种安全风险。根据 Google 透明度报告，截止 2020 年 9 月 20 日，Google 的所有产品和服务提供加密的比例达到了 95%[②]。目前，这个比例还在上升。Google 的最终目标是做到 100%加密。除了对网络流量加密之外，研究人员还提出了匿名网络、加密代理和隧道技术等一系列的隐私增强技术。匿名网络是在公共网络之上构建的覆盖网络，具有不可关联、不可辨识和不可观测等性质。目前广泛应用的匿名网络是 Tor(The Onion Router)[1]。它将数据以定长封装在 Cell 元胞内，并综合使用了流量加密、多重中继和动态路由等技术。加密代理则是在用户与网络服务之间部署中继节点，转发加密流量数据，从而使得用户与网络服务之间变得不可关联。常见的加密代

图 1 常见的隐私增强技术

①https://zh.wikipedia.org/wiki/互联网

②https://transparencyreport.google.com/https/overview

理有 SSH 代理、Shadowsocks 代理等。隧道技术实际是一种隧道协议。它通过将另一种网络协议封装在协议载荷中，实现在不兼容的网络间通信或者向外提供一种安全通道服务。VPN 则是应用隧道技术的典型代表。上述三种 PETs 之间虽然有些技术重叠，但是各自的侧重点不同。

各种隐私增强技术通过流量加密、单跳或多跳代理、载荷固定长度填充等手段极大地提高了用户使用网络的安全性。然而，网站指纹研究表明，即使在使用 PETs 的网络环境下，用户的浏览兴趣隐私仍然具有泄露的危险。网站指纹研究包括识别和防御两个方面。网站指纹识别是一种流量分析攻击。它利用网页浏览过程中暴露的网站指纹特征对用户的浏览隐私进行推测与识别。网站指纹特征简称网站指纹。它是基于数据包时间戳、顺序、方向和大小等基础信息而产生的一系列特征。具体产生的方法可以是直接的，也可以是间接的。直接产生就是将基础信息直接作为特征。间接产生就是一种或多种基础信息经过一定的数学变换从而得到新型特征。网站指纹也称为元数据信息。网站指纹识别者通常假定来自本地，而且实施的是被动的识别。识别者来自本地意味着识别者可以将截获的流量与用户关联，比如本地的代理或本地局域网管理员等。实施被动识别则要求识别者不能对流量进行解密、篡改、丢弃或者重放。

与网站指纹识别相对应，研究人员设计或提出了一系列的网站指纹防御技术以抵抗识别的危胁。网站指纹防御通过一定的算法对流量数据进行修改、模糊或混淆，掩盖具有区分信息的关键流量特征，同时尽量做到减少带宽和时延的耗费。值得指出的是，当网站指纹防御技术具体应用部署之后，它也构成了新型 PETs。比如，数据持久化技术、并行流水线技术以及在 Tor 浏览器上完成实验性部署的 WTF-PAD 防御等实际上也构成了轻量级的 PETs。

关于网站指纹的学术研究已经发展了二十多年，相继取得了一系列值得关注的研究成果。鉴于网站指纹技术对于用户隐私保护和网络空间安全治理的重要意义，系统梳理网站指纹相关技术的研究成果和发展动态显得尤其重要。然而，目前该领域的综述文献还比较少见。基于该认识，本文对网站指纹识别和防御按照主体研究和辅助研究顺序分别进行了较为系统的总结和分析。在主体研究方面，本文首次对网站指纹识别方法和防御方法的研究成果根据其技术特点进行分类梳理，形成如图 2 的结构脉络。对于网站指纹识别方法，本文首先根据分类器判别决策的内在数学原理将其分为相似度判别法、传统机器学习方法和深度学习方法三种。进一步，通过深入分析各类识别方法，本文依据其具体使用的数学模型不同进行二次细分。类似地，对于网站指纹防御方法，本文首先根据防御技术应用的网络层次不同，将其分为应用层、网络层和复合层防御。在深入分析防御技术原理的基础上，本文根据防御方法采用的防御模式的不同，对其再次进行分类，以便于进行对比分析。在辅助研究方面，本文先后讨论了网站指纹识别如何在先验知识不足、低基础发生率、复杂背景流量的场景下实施，如何解决训练数据过时等问题，并总结了站内网站指纹和新型网站指纹的相关工作。此外，网站指纹防御也探讨了安全和带宽界限的评估、分析特征的信息泄露和系统化的防御评估方法等问题。最后，本文总结了网站指纹研究面临的三大挑战，并对未来研究进行了展望，提出了五个可能的发展趋势。

本文的组织结构如下：第 2 节概述了网站指纹研究的基本概念、识别假设、威胁模型、防御模型和研究意义。在此基础上，本文利用第 3、4 节对网站指纹识别和防御的国内外研究进展进行系统化总结和分析。然后，本文在第 5 节进行了总结与展望，着重分析了当前网站指纹研究面临的问题和挑战，并给出了网站指纹未来研究方向的几点思考和前瞻。

2 网站指纹研究概述

网站指纹研究主要解决两个问题。一是如何利用指纹特征识别出用户访问的网站，即识别问题。二是如何隐藏指纹特征，使得识别失效，即防御问题。相应地，网站指纹研究包括网站指纹识别和防御研究两个方面。为了便于叙述，下面在引入一些基本概念的基础上，分别对网站指纹识别研究和防御研究的基本涵义进行了阐释。针对网站指纹识别，本章第 1、2 节分别对其识别假设和威胁模型等问题场景的定义进行了论述。对于网站指纹防御研究，本章第 3 节重点介绍了网站指纹防御模型。最后，第 4 节对识别和防御的辩证关系和研究意义进行了总结。

网站指纹研究涉及到一些基本概念，如上行包、下行包、突发(Burst)、迹(Trace)、实例(Instance)、指纹(Fingerprint)、模板(Template)、元数据(Metadata)、敌手(Adversary)等。上行（去）包是指从客户端到服务器的数据包，下行（来）包则相反。突发是一组同一方向（上行或下行）的连续包序列。迹是指一次网页访问期间产生的一组连续的数据包序列。数据包一般用时间戳和带方向包长的元组来表示。如果把网页看成是一个类别，对应于某个类别的一条迹称为一个实例。指纹即网站指纹。它是指从迹中直接或间接提取的一系列特征形成的一个数组或向量。模板假设是网站指纹研究的一个

图 2 网站指纹识别与防御方法体系脉络

基本假设，表示同一个网页的模型是一致的。对于网页的不同次访问产生的迹符合网页同一模板假设。该假设在一定的时期内是近似成立的。元数据是网站指纹领域最小的研究对象，即数据包。在网站指纹研究中，敌手即是网站指纹识别者。

网站指纹识别本质是一种有监督分类的问题。分类问题的输出是网页类别，输入是每次网络访问产生的连续包序列。经典的网站指纹识别的场景模型如图 3 所示。该场景模型包括用户、识别者、加密流量、网络和网站服务器等五种实体。用户是访问网页的普通网民。大多数网站指纹识别通常是定向的，即针对特定的用户。如果针对同一个局域网的多个用户，就是非定向的网站指纹识别[2]。识别者位于离用户最近的出口路由器之内，可以关联加密流量和目标用户，提高识别的准确率。根据实际的网络场景，加密流量使用的加密协议会有所区别。传统的网站指纹研究并不需要对载荷进行解密，一般是通过网络数据包的时间戳、包大小和方向等信息提取指纹特征。网络场景可以是匿名网络、加密代理或者 VPN 的任何一种，甚至可能是多种场景的复合。网页服务器可以是普通网站，也可能是其他的 Web 服务，比如，著名的 Tor 的隐藏服务。一般地，网站指纹识别包括三个步骤。首先，识别者截获用户和网络入口之间的流量数据，并从流量数据中提取出元数据特征信息；其次，识别者利用提取的特征数据对一定的数学模型进行训练，得到必要的参数，从而完成模型的构建；最后，识别者再次截获用户与网络入口的未知流量数据，并利用训练得到的模型，判断流量数据所属的网页是否为识别者关注的目标网页。如果是，则给出告警。其中，流量数据通常是加密的。

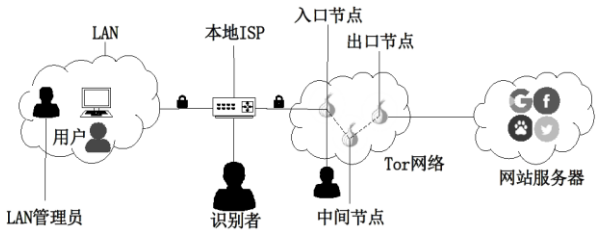

图 3 网站指纹识别研究一般场景模型

由于受到用户行为的多样性、复杂的背景流量、敌手对用户的先验知识不足等多种因素的影响，在实际网络环境下开展网站指纹识别研究的难度较高。为此，网站指纹识别研究需要设置一些前提条件，即识别假设。另外，网站指纹识别研究一

个重要的问题就是敌手位置。敌手位置需要在隐私增强技术的覆盖范围之内，识别者才能够截获到受防护的流量，使得网站指纹识别研究具有意义。威胁模型主要反映了敌手能力、位置等相关信息。本章的第 1、2 节将对识别假设和威胁模型详细展开讨论。与网站指纹识别相对应的技术是网站指纹防御技术。它是通过对网络流量特征施以混淆、变形和延时发送等操作，达到隐藏或覆盖流量的部分甚至全部有区分能力的特征的技术。从网站指纹的研究历史来看，网站指纹防御与识别研究是相互促进，且共同发展。网站指纹防御方法的实现大部分需要网络通信的参与者协作完成。这些参与者包括（用户）客户端、防御代理节点和（网站）服务器等。防御代理节点包括防御客户端和防御服务器两种。针对去包的防御措施通常由客户端或防御客户端实现。针对来包的防御措施一般由网站服务器或防御服务器实现。因此，如果仅在客户端或服务器一侧部署防御，肯定会存在一个方向的数据包未被防御保护，导致防御效果不佳。

2.1 识别假设

网站指纹识别研究通常需要在一定的前提假设之下开展。这样做一方面便于研究人员对不同的识别方法进行对照实验，另一方面有利于将研究的重点放在对流量本身的可区分特征的挖掘和发现任务之上。

2.1.1 场景假设

鉴于科学研究循序渐进的一般规律，减少网站指纹研究过程中复杂因素的影响，研究人员普遍基于两种场景假设开展网站指纹方法评估和验证，即封闭世界和开放世界假设。

在封闭世界假设下，用户仅可以访问给定范围内的网页。这一部分网页集也称为监控集，即识别者感兴趣的目标网页集合。与实际可访问的网页数量相比，该网页集只是一小部分。由于封闭世界的场景假设较为理想，一般仅用于初步判断网站指纹识别方法的可行性并进行调参，以及初步比较不同网站指纹识别方法的性能。此外，大多数文献评估防御方法时，也一般采用封闭世界假设。显然，封闭世界假设场景下的网站指纹识别是一种多分类问题。在开放世界假设下，用户不仅可以访问识别者设置的监控集的网页，也可以访问相当数量的其他网页集合。该网页集合一般称为非监控集。而且，非监控集比监控集的网页数量要大的多。显然，开放世界场景假设比封闭世界假设更加接近实际情况。开放世界场景下的性能评估是分析网站指纹识别方法有效性不可缺少的环节。在开放世界假设场景下，网站指纹识别研究可以看成二分类问题，也可以看成是多分类问题。

2.1.2 网页假设

大部分的网站指纹研究默认对网页做了两种假设。一是用户访问的网页是网站的主页。显然，这是一个比较强的假设。二是同一网页在不同的时刻适用于同一个模型。该假设与实际情况是有偏差的。这种偏差受到网页的动态性、网站安全机制和网站改版等多种因素的影响。比如，网站改版后，网页包含的内容可能会有较大的差别，导致加载的包序列不再符合同一模型。然而，在足够小的一段时间内，这一假设是有其合理性的。显然该窗口时间长短跟具体的网站是紧密相关的。于是，在特定时长的实验进程中，可能出现大部分网站适用该假设，但个别网站并不适用的情况。从该意义上说，该假设存在额外的误差。

2.1.3 用户行为假设

在用户行为假设方面，主要是针对用户的上网行为进行规定。一是单标签访问。用户只能逐个网页进行访问，即每次只能打开一个网页。二是多标签访问。即用户先打开一个网页浏览，间隔规定的时间或者等第一个网页完全加载后，再打开第二个网页进行浏览。虽然，用户行为假设较为理想，却是必要的和有意义的。因为只有这样，需要研究的问题才变得较为确定。实际上，在网站指纹的研究中，可以根据研究的具体问题进一步抽象出合理的用户行为假设。

2.1.4 识别者能力假设

大部分网站指纹研究对于识别者能力做了两种假设。一是识别者具有用户足够多的先验知识。这一假设使得识别者可以完全复制用户的网络环境、浏览器配置等条件进行流量截获和模型训练。这种先验知识的获得不属于网站指纹研究的范畴。二是识别者具有流量分割的能力。即识别者可以分离噪音流量和网页流量，这里的噪声流量指的是除浏览器流量之外的其他应用流量。而且，该假设同时认为识别者也可以从网页流量中准确提取每一次网页访问的完整流量，即识别者可以获取每一个流实例的完整包序列。这在实际中是可行的。因为通常情况下用户在发起网页请求之前会有一个短暂的思考时间。从而可以利用这个间隔时间判别流实例的开始与结束[3]。

本质上，网站指纹识别假设可以理解成是实际场景的一种简化。它的积极意义在于可以使研究人员将重点放在关键特征的发现和利用之上。另一方面，通过识别假设降低实际网站指纹问题的难度，符合人类认识事物从简单到复杂的一般规律。为了系统地把握网站指纹研究使用的识别假设情况，我们将其梳理总结在表 1。为了缩小与实际场景的差距，越来越多的研究人员开始尝试在弱化识别假设的前提下开展网站指纹研究[2][4][5][6][7]。

2.2威胁模型

威胁模型反映了网站指纹识别的基本情景。威胁模型不仅指出了识别者对用户具备的先验知识

的多少（即识别者知识），也说明了识别者具备哪些网站指纹识别必须的能力（即识别者能力）和意欲识别的目标对象和数量（即识别者目标）等信息。除了识别者知识、识别者能力和识别者目标之外，威胁模型主要给出了识别者的可能位置和来源。网站指纹识别是在一定的威胁模型的基础上进行的，即网站指纹识别方法都是基于某种威胁模型而成立的。威胁模型不同，识别的难度不同。按照典型网站指纹识别的定义，网站指纹识别研究主要考虑本地的被动式识别者，即识别者位于接近本地网络的位置。此外，该识别者位置必须可以截获到隐私增强技术保护的加密流量。于是，网站指纹识别研究对识别者的位置提出了相应的要求。从识别者的来源看，符合该要求的识别者可能是本地 ISP、恶意的网络入口节点和本地局域网(Local Area Network, LAN)管理者等。

表 1 网站指纹识别文献使用的识别假设汇总

识别假设	网站指纹识别文献
封闭世界	文献[2, 5, 8-27]
开放世界	文献[2, 8-10, 12, 13, 17-24, 27]
网页假设 1（主页）	文献[2, 5, 8-27]
网页假设 2（模板）	文献[2, 5, 8-27]
行为假设 1（单标签）	文献[2, 8-27]
行为假设 2（多标签）	文献[2, 5, 7, 19, 27]
识别者假设 1（先验知识）	文献[2, 5, 8-27]
识别者假设 2（流量分割）	文献[28]

值得一提的是，对于不同的网络场景，由于其应用的隐私增强技术不同，传输链路的流量包转化过程是不同的，相应所需要的中继结点数量、服务器协议、数量和位置等也是不同的。因此，只有确定了应用场景，才能讨论威胁模型。威胁模型确定了，才能研究网站指纹识别。以 Tor 网络场景为例，图 4 给出了相应的威胁模型。

可以看到，对于采用多中继的 Tor 网络，除了 LAN 管理员和本地 ISP 之外，潜在的识别者一般在入口洋葱路由器(Onion Router, OR)。虽然研究人员设计的防御技术也有在 Tor 的中间节点和出口节点部署的，但是并不是主流[29]。由于用户直接使用 Tor 浏览器访问网站服务器，在三个潜在识别位置截获的流量都是经过 Tor 防御机制保护的流量。

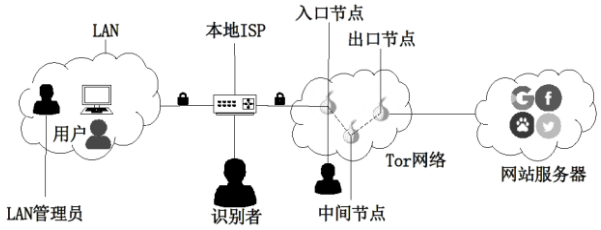

图 4 基于 Tor 应用场景的网站指纹识别威胁模型

2.3防御模型

与威胁模型定义了敌手的位置类似，防御模型刻画了防御代理的部署位置，包括防御客户端和防御服务器。防御位置决定了防御代理可以获得什么样的流量数据以及可以实施什么样的防御操作。网站指纹防御方法一般需要防御客户端和防御服务器的支持。防御客户端与防御服务器之间的链路应该覆盖潜在识别者可能的识别位置，从而可以保证识别者截获到的是受到防御技术保护的流量。网站指纹防御的通用模型如图 5 所示：

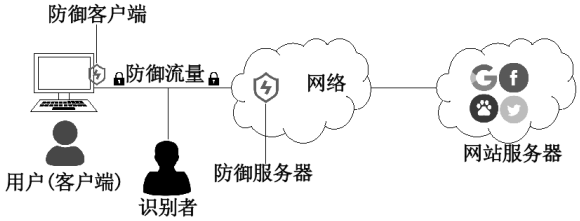

图 5 网站指纹防御的通用模型

防御客户端和防御服务器配合完成了某一特定的网站指纹防御技术。在实际部署中，由于防御技术的功能相对比较单一，防御客户端和防御服务器一般与网络中的已有的中间节点集成在一起。如果配置独立的防御客户端或防御服务器，不仅额外增加成本，也需要额外增加通信开销。无论从成本还是效率来看，这样部署都不值得。

一般情况下，防御客户端部署于用户终端，甚至可以做为浏览器的插件与浏览器软件集成在一起。防御服务器通常位于通信网络中间。单纯从技术实现的角度上看，虽然防御服务器可以部署在网站服务器这一侧，但是这样做存在明显的问题。原因主要有两点。一是如果防御服务器部署在某个网站服务器上，则该防御技术只能对该网站施加防护。此外，一般网站服务器通常都希望网站被广泛传播。因此，它们并没有动力部署防御服务器。因此，防御服务器一般部署于网络的某个中继节点上，例如，在 Tor 匿名网络场景中，防御服务器可位于中间节点。

实施防御时，防御客户端截获用户的加密流量，对其应用防御技术后形成防御流量，并输入到链路中传输。防御服务器接收数据后，进行逆防御操作，恢复出原始的加密流量。

2.4研究意义

网站指纹研究包括识别和防御两个方面。它们是相辅相成、互相促进和交替发展的。开展网站指纹识别与防御研究对构建一个安全健康的网络环境具有重要的意义。

首先，网站指纹识别技术的研究具有重要的意义。依托隐私增强技术提供的安全防护，大量不法之徒大肆进行毒品枪支、人口贩卖、色情恐怖、敲诈勒索、反动谣言和洗钱等黑色交易和违法活动，

给世界各国带来了严重的社会问题和安全问题。例如，Tor 黑市站点“丝稠之路”在短短的 2 年半时间内，不法交易规模大致 15 亿美元，聚集了 4000 个以上的不法商家和 15 万名的匿名用户[30]。如何尽可能准确地探测到诸如此类的不法交易是网站指纹识别研究的目的之一。其次，网站指纹识别技术研究对于隐私增强技术的发展具有重要意义。通过该技术的研究可以发现隐私增强技术的不足和缺陷，从而推动研究人员改进相应的隐私增强技术。最后，网站指纹识别研究对于每个使用网络的用户也具有较强的指导作用。用户了解该技术，就可以知道如何改变自己的浏览习惯或者网络、浏览器的设置，使得自己浏览的网站难以被识别，从而减少隐私泄露的风险。

同样地，网站指纹防御技术的研究必不可少。网站指纹识别的研究成果也表明即使在具有 PETs 的网络环境下，人们的隐私仍然存在泄漏的风险。例如，不法分子可能利用网站指纹识别技术对正常浏览网页的用户隐私进行窥探，给用户的个人信息安全造成了较大的危胁。而且，网站指纹识别技术对计算资源要求不高，进一步扩大了潜在识别者的范围，也增加了潜在的安全隐患。因此，研究网站指纹防御技术，提出新的更高效的流量延迟、伪装、变形和填充等防御技术，积极应对网络黑手利用网站指纹识别技术窃取用户隐私的现实危胁显得尤其重要。

3 网站指纹识别研究进展

网站指纹识别研究是网站指纹主要研究方向之一。本节在概述网站指纹识别研究的基本情况、指纹特征、评价指标和分类方法的基础上，分两个层次对网站指纹识别的主体研究，即网站指纹识别方法，进行分类并详细进行比较综述，随后也总结了网站指纹识别的其他相关研究动态。

3.1概述

网站指纹识别最早可以追溯到 1996 年。Wagner 和 Schneier 首先注意到包长度特征泄露了流量数据的信息[31]。随后，研究人员相继通过分析每个 TCP 连接，提取到 Web 页面的各种资源长度特征，通过简单的统计计算方法，取得了较高的准确率。这一阶段的网站指纹识别是利用资源长度特征实施的。随着 HTTP 流水线技术、持久化技术、Traffic Morphing[32]等网站指纹防御技术的应用，资源长度特征不再有效。为此，研究人员寻求新的包元数据特征，并应用机器学习方法进行网站指纹识别研究，也取得了较好的效果。随着 PETs 的进一步发展，网络流量在源端即被加密保护。而且，通过单跳或者多跳中转，加密流量在传输过程中再次被加工处理，进一步增加了识别的难度。于是，研究人员对流量的特征进行了更深入的挖掘。网站指纹识别方法也从传统机器学习过渡到了深度学习，从而持续保持了网站指纹识别较高的准确率。

3.1.1 指纹特征

指纹特征在不同的领域有不同的具体涵义。在网站指纹研究领域，指纹特征即网站指纹。它可以理解成是用来描述流实例的一个向量或数组。在网站指纹研究中，指纹特征是识别与防御方法所处理的直接数据对象。对于网站指纹识别方法而言，识别者利用训练数据的指纹特征构建出分类器，并将该分类器应用于测试数据的指纹特征，从而推断出未知流量的所属网站类别。对于网站指纹防御而言，防御代理需通过一定的策略方法隐藏有区分能力的特征，从而达到防御的目的。可见，指纹特征是网站指纹研究的关键所在。

指纹特征具有狭义和广义的区别。狭义的指纹特征仅考虑元数据的时间戳和包长度，其中元数据可以是 Cell 单元、IP 包、TCP 包或者是同一个方向的连续包组成的突发。广义的指纹特征可以进行一定程度的包头部分析，但不能脱密。即广义的特征包括了 TCP/IP 头部的信息。

提取狭义的特征所基于的流实例模型不含数据包头部信息，仅包括时间戳和包长度。一般地，首先可以把流实例定义成$< t_1, l_1 >$、$< t_2, l_2 >$、...、$< t_n, l_n >$的形式。其中，t_i表示时间戳，l_i表示带方向包长度。一般地，上行包大小表示为正整数，下行包则相反。理论上，只要是该序列数据通过映射、统计或某种运算得到的特性值都可以做为特征。大致地，狭义的特征可以分为五类。一是包长度特征，即与长度有关的统计信息。二是包时间特征，主要包括包到达率、包到达间隔时间和包传输时间等。三是包顺序特征，即与包到达的先后次序有关的顺序信息。四是包统计特征，主要包括总体统计和头部、尾部统计信息。五是突发特征，即与突发有关的统计信息。

对于广义的特征，学者 Yan 和 Kaur 进行了系统的分析，共梳理了 109 个类别，共 35683 种特征并进行了分类，将它们分为包层级、突发层级、TCP 层级、端口层级和 IP 地址层级[33]。包层级特征包括从数据包的计数、长度、顺序、时间和方向直接得到的特性值；突发层级特征则包括了从计数、持续时间、字节、连续数据包以及到达时间中获得的特性值；在 TCP 层级，则是根据 TCP 子流的计数、持续时间、字节数、数据包和突发以及到达时间来定义特征；在端口层级，需经由统计流实例中不同端口的使用情况以及特殊端口(443、80)的数量来提取特征；在 IP 地址层级，可以通过计算不同的 IP 地址和主机为一个网页下载而通信的频率，以及不同 IP 地址和主机名之间的通信量来提取特征。

流量的指纹特征受到多种因素的影响，包括浏览器应用软件、浏览器软件版本和缓存策略等。这

些因素通过影响流量的指纹特征，进而影响网站指纹识别的准确率性能。

浏览器应用软件或浏览器软件版本对流量指纹特征产生影响，根本原因在于不同浏览器软件或同一浏览器软件的不同版本对于各种web技术特性的支持是不同的。典型的影响流量模式的 web 技术特性包括 JavaScript 解释器、异步执行脚本、导航计时 API、ActiveX 特性、本机 Flash 阻塞特性和缓存编译程序等。上述特性对流量模式的影响原因与其各自的功能相关。例如，导航计时 API 特性为网页加载提供了精准的定时机制。假设同一网页被多次加载，支持该特性的浏览器产生的流量模式偏差势必要比其他的浏览器要小。

浏览器缓存策略通过在本地缓存网页资源，减少了浏览器从服务器加载资源的频率，从而加快了网页的加载。从浏览器缓存策略的功能不难看出，是否启用缓存策略对流量模式的影响是很大的。而不同浏览器缓存机制的实现细节也会影响流量模式。例如，缓存的空间大小、缓存的最大文件大小等。这些差异都会影响到流量模式，进而影响指纹特征。

3.1.2 评价指标

网站指纹识别的评价方法与实验场景相关。在封闭世界场景下，一般以准确率(Accuracy)为评价指标，即召回率(Recall)或真阳率。具体的定义如公式(1)所示。其中，TP指检测为阳性且实际为阳性的样本数量；FN指检测为阴性且实际为阳性的样本数量。

$$Accuracy = Recall = TPR = \frac{TP}{TP + FN} \#(1)$$

在开放世界场景下，评价指标更加多样，包括精度(Precision)、贝叶斯检测率(Bayers Detection Rate, BDR)、召回率、准确率和 F1-Measure 等。其中，精度、贝叶斯检测率的定义如公式(2)和(3)所示。在精度计算公式中，FP指检测为阳性且实际为阴性的样本数量。在贝叶斯检测率计算公式中，$p(mon)$表示监控集测试样本数量与总样本数量的比例。$p(unmon)$表示非监控集测试样本数量与总样本数量的比例。召回率的计算公式与封闭世界场景一致。需要指出的是，开放世界场景的准确率并不等于召回率，其定义如公式(4)所示。其中，TN指检测为阴性且实际为阴性的样本数量；P表示实际总的阳性样本数据；N表示实际总的阴性样本数量。在实际研究中，常常画出精度-召回率(Precision-Recall, P-R)曲线图直观比较不同分类器的性能。理论上，精度和召回率同时达到 1 是最理想的情况。在实际研究中，往往需要在两者中权衡。F1-Measure 又称为 F1-Score，它是精度和召回率加权调和平均，是信息检索领域的常用的一个评价标准，常用于评价分类模型的好坏，其定义如公式(5)所示。

$$Precision = \frac{TP}{TP + FP} \#(2)$$

$$BDR = \frac{TPR \times p(mon)}{TP \times p(mon) + FN \times p(unmon)} \#(3)$$

$$Accuracy = \frac{TP + TN}{P + N} \#(4)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \#(5)$$

3.1.3 分类方法

网站指纹识别研究可以分为主体研究和辅助研究两大类。网站指纹识别主体研究的根本目标是在面对各种不同的网络场景条件下依然可以对测试实例做出准确的识别。为了实现这一根本目标，研究人员往往通过精心选择特征和合理匹配数学模型等角度，在不同的应用场景中不断实现更高准确率的分类器。实际上，网站指纹识别的主体研究即是识别方法。

从不同的角度上看，网站指纹识别方法可以有多种分类方法。一是根据使用的特征的维度划分，即包层级、突发层级、TCP 层级、端口层级和 IP 地址层级进行分类。然而，很多文献使用的特征往往跨越多个层级，导致这种分类方法不够清晰。二是根据识别方法针对的隐私增强技术的不同，包括 Tor、JAP、Shadowsocks、SSH 和 VPN 等。该分类方法同样存在同一文献涵盖不同类别的情况出现，导致各类别不够独立。三是根据具体使用的方法不同。纵观网站指纹研究历程，可以大致分为相似度判别方法、传统机器学习方法和深度学习方法等。除了上述网站指纹识别主体研究之外，研究人员也开展了辅助网站指纹识别的其他研究，包括如何解决数据过时问题、如何解决“基础发生率谬误”问题、如何分离背景流量等。

3.1.4 公开数据集

在网站指纹识别的研究过程中，研究人员为了确保其识别方法的可复现性，在 Github 或者个人主页上发布了实验数据集。这些数据集是网站指纹识别研究的重要基础，陆续被后续的学者所引用。为了便于查阅，本文将典型的公开数据集整理成表 2。

3.2网站指纹识别方法研究进展

考虑到基于特征层级和隐私增强技术的分类方法存在不够清晰的问题，本文基于具体所使用的方法对网站指纹识别主体研究的文献进行分类。具体到每类方法时，本文再进一步根据使用的数学模型进行细分。最后，分别从隐私增强技术场景、特征的使用、数据集情况和具体的性能等方面对其中文献加以比较分析。

3.2.1 相似度判别法

相似度判别法是指通过计算测试实例指纹与各监控网站(类别)指纹之间的相似度，根据相似度

与设置的阀值的比较结果，以决定测试实例的类别的方法。该方法的关键技术在于选择哪些特征、如何选择合适的网页模型和如何选择相似度策略。本文根据不同的相似度计算方法对文献进行综述，并总结归纳在表 3 中。

特征匹配法。特征匹配法是指直接将两个特征向量之间相同的特征值的数量作为两个向量之间相似度的度量。这种度量方法不需要复杂的计算，只需要进行简单的比较即可。该方法一般用在网站指纹研究的早期阶段，即资源长度识别。

Cheng 和 Avnur 最早进行了资源长度识别的研究。他们仅利用了 HTML 长度和总对象长度信息作为特征，并建立了特征对数据库。在网页数量不多的情况下，数据库里的特征对不会重复。在进行匹配查询时，最多只有一次匹配。当网页数量多时，特征对可能重复。为此，作者提出了链接分析方法。在开启缓存的 92 次测试中，准确率为 96%。在不开启缓存的 71 次测试中，准确率为 94%[36]。同样基于特征匹配法，Hinz 在其资源长度识别研究中，提取了每个访问实例的所有文件大小及其计数信息作为特征，并与各网页模板的特征进行比较，统计出准确匹配的大小次数。如果匹配次数大于某个门限值则判断为匹配成功[37]。在初步概念验证时，作者基于 SafeWeb 数据进行测试。结果表明，来自同一个网站的两次不同访问至少有 45%是精确匹配的。

表 2 网站指纹识别研究公开数据集概览

数据集名称	监控集[1]	开放世界集[1]	PETs	链接
Libratore06[3]	2000×100	NA[2]	SSH	http://traces.cs.umass.edu/
Cai12[34]	100×40	NA	Tor	http://home.cse.ust.hk/~taow/wf/
Wang14[24]	100×90	9000×1	Tor	http://home.cse.ust.hk/~taow/wf/
Rimmer17[20]	900×2500	400000×1	Tor	https://distrinet.cs.kuleuven.be/software/tor-wf-dl/
Sirinam18[21]	95×1000	9000×1	Tor	https://github.com/deep-fingerprinting/df
Zhuo18[27]	200×25	2000×1	Shadowsocks	https://github.com/jason-zhuo/PHMM
Wang20[4]	100×200	80000×1	Tor	https://github.com/OpenWF/openwf.git
Smith21[35]	100×200	16000×6	QUIC, TCP	https://github.com/jpcsmith/wf-in-the-age-of-quic

[1]监控集和开放世界集描述统一格式化为：网站数量×每个网站实例数量。
[2]NA 表示不存在对应的数据集。

综上，在特征使用上，Cheng 和 Avnur 使用了 HTML 长度和总对象长度构成特征对，且突出了 HTML 长度特征的区分能力。Hinz 使用了资源对象长度及计数特征，而不是计算对象总长度为特征；在识别方法上，文献[36]和[37]均使用简单的匹配方法；在方法优化上，两者分别使用了链接分析算法和范围匹配策略；在实验场景上，两篇文章都是面向以 SSL 加密技术来强化隐私的网络场景。但是，Hinz 所基于的 SafeWeb 是一种网页代理。它不仅使用 SSL，而且使用 JavaScript 来加密页面内容和网页地址。

特征匹配法有效的原因在于早期的 HTTP 协议泄露了资源长度、HTML 长度和总对象长度等特征。而且与当时的网页以静态网页为主有关。这种方法在现在的网页技术条件下已经不再适用。

雅卡尔系数。雅卡尔系数(Jaccard Coefficient, JC)主要用于衡量两个集合之间的相似程度。对于集合 A 和集合 B，它们的 JC 相似度定义为：

$$J(A,B)=\frac{|A\cap B|}{|A\cup B|}\#(6)$$

使用雅卡尔系数作为相似度的研究工作主要有两项。以资源对象的长度和数量为特征，Sun 等把网页实例看成了不同长度对象的集合，计算网页实例之间的相似度使用了雅卡尔系数[38]。在 2191 个监控目标网页集和 98496 个非监控网页集的数据集上，识别的真阳率约为 75%，假阳率低于 1.5%。同样地，Liberatore 和 Levine 也用雅卡尔系数来度量不同网页实例之间的相似度。与 Sun 等的工作不同的是，Liberatore 和 Levine 以包长度和方向作为特征。所有不同的带方向包长构成实例的属性。属性值定义为每个实例具有对应属性的包的数量。网页的属性集为属于该网页大多数实例的属性的集合。实验数据在 SSH 的网络环境中收集。在 1000 个网站规模的数据集中，该方法取得了 73%的准确率[3]。

综上，在特征使用方面，Sun 使用了资源对象长度及计数特征。Liberatore 和 Levine 则采用了包长度和方向；在实验场景上，前者在 SSL 加密的网络场景上应用，后者则是基于 SSH 代理环境；在性能结果方面，两者相差不大，但是数据集的规模有所不同。相较特征匹配法，基于雅卡尔系数的相似度策略忽略了对于特征元素顺序的限制。实际上，这两种策略都无法辨别特征的微小扰动，从而可能引起错误，正在逐渐被淘汰。

余弦相似度。余弦相似度利用两个向量之间夹角的余弦值来度量相似程度。它最先被学者 Shi 和 Matsuura 用于研究基于 Tor 场景的网站指纹识别问题[39]。他们利用了连续来包计数构成的向量作为网页实例的代表。对于网页的代表向量，使用了属于该网页的多个实例其中之一的代表向量作为类别代表。该向量与所有其他实例的相似度的连乘积是最大的。在距离的计算上，作者使用了加权的余弦相似度作为相似度得分。权重为两个向量所对应的网页实例具有的间隔数量的比值。在 20 个网站规模的实际测试中，该方法的准确率 50%。

余弦相似度策略是一种广泛使用的距离度量

方法。Shi 和 Matsuura 的工作是直接利用了该策略进行分类。然而在网站指纹研究中，余弦相似度策略可以与机器学习方法结合起来使用。

表 3 基于相似度判别的网站指纹识别方法

分类器[1]	相似度策略	PETs	特征	数据集[2]	性能
LL_JAC[3]	雅卡尔系数	SSH	带方向包长	1000	73%
CW[36]	特征匹配	HTTPS (SSL)	HTML 长度和总对象长度	1(92 个页面)	94%-96%
HZ[37]	特征匹配	SafeWeb	资源文件大小及计数	1	45%
SS[38]	雅卡尔系数	HTTP	资源文件大小及计数	2191，98496	TPR>75%，FPR<1.5%
SK[39]	余弦相似度	Tor	连续的来包计数	20	50%
LC[40]	编辑距离	VPN/SSH 等	TCP 尾包	1000，2000	81%-97%
BL[41]	互相关系数	SSH/tinyproxy	包间时隔和包大小	25	40%
QZ[42]	汉明距离	VPN/Tor	带方向包长度	100，550	55%-97%

[1]分类器默认使用原文名称或已发表文献中使用的名称。若无可查名称，本文以作者姓氏首字母结合分类器特点自拟。
[2]数据集默认单位是网站数量。

编辑距离。编辑距离是一种度量两个字符串差异程度的方法。具体方法是计算一个字符串经过多少次数的处理操作之后可以变成另一个字符串。得到的处理操作的次数反映了两个字符串的差异性。常用的编辑距离是莱文斯坦距离。

Lu 等借鉴编辑距离对 SSL(SSH)隧道环境中的网站指纹进行研究。作者考虑到 HTTP 传输方式是成块传输，即除了最后一个包，其他包都是以 MTU 大小传输。因此，本文利用了最后一个包作为资源对象的可识别特征。在剔除了 MTU 数据包和长度在 300 字节以下的数据包之后，考察了两个特征向量，即 HTTP 请求特征向量和 HTTP 回复特征向量。在计算网页实例之间的相似度时，分别使用两个向量计算编辑距离，并进行加权求和。求和结果作为最终的相似度。在 OpenVPN 和 OpenSSH 的环境中，作者分别在 1000 和 2000 个网站的数据集上实验，准确率分别达到了 97%和 81%[40]。

互相关系数。互相关系数最初用来表征不同通信信号之间的相互依赖关系。现在也逐渐被其他领域借鉴。Bissias 等首次使用互相关系数[41]研究网站指纹识别。针对无法从 TCP 连接获取资源对象大小的实际场景，作者考察了流量实例的两种特征序列，一是包间间隔时间，二是每个包的大小。在相似度度量上，采用两种特征序列分别计算互相关系数，并计算其乘积的方法。在实验环境设置上，浏览器先经 SSH 隧道连接到 tinyproxy 代理，再经因特网与 Web 服务器连接。在可识别性最好的头 25 个网站中，猜测 1 次的成功率为 40%，猜测 2 次的成功率为 70%。

汉明距离。汉明距离是信息论中比较两个等长的字符串的差异的一种方法。其定义为两个字符串对应位置字符不相同的数量之和。

Qasem 等利用改进的汉明距离作为相似度度量方法进行了网站指纹识别的研究。该文章的特点是引入了布隆过滤器对流实例进行摘要表示。基于布隆过滤器，作者为每个实例定义一个相似度摘要作为其特有表示。相似度摘要定义为布隆过滤器的数量与每个布隆过滤器的连接组成。在计算相似度摘要之间的距离时则使用了汉明相似度的方法。在 VPN 的数据集中取得了很好的效果，但在 Tor 数据集上效果却一般[42]。

3.2.2 传统机器学习方法

机器学习算法是一类从数据中自动分析获得规律，并利用规律对未知数据进行预测的算法。机器学习在近 30 多年已发展为一门多领域交叉学科。网站指纹识别研究中大量应用了机器学习的模型和算法。虽然深度学习方法属于机器学习方法的概念范畴，然而由于深度学习方法的出现相对较晚也相对独立，本文把除深度学习方法之外的其他机器学习方法称为传统机器学习方法。根据数学模型的不同，本文对基于传统机器学习的识别方法进行了综述，并归纳总结在表 4。

朴素贝叶斯。朴素贝叶斯(Naive Bayers, NB)方法是贝叶斯分类方法的一种，其数学理论基础是贝叶斯原理[49]。NB 方法通过高斯核估计测试实例属于各个类别的概率，并以最高概率的类别为判断结果。

Liberatore 和 Levine 在 SSH 网络环境中，不仅使用雅卡尔系数的相似度方法，也使用了 NB 的分类方法[3]。同样针对 SSH 网络环境，Herrmann 等使用了多项式朴素贝叶斯(Multinomial Naive Bayers, MNB)方法进行网站指纹研究。同时，本文使用了三种属性值转换方式，即 TF(Term Frequence)变换、IDF(Inverse Document Frequence)变换和 TF-IDF 变换，并应用余弦正则化对属性向量进行归一处理。在总共 775 个网站的数据集中，MNB 方法取得了 94.31%的准确率。同时，作者也针对 OpenVPN、CiscoVPN、Stunnel、Tor 和 JonDonym 等应用场景进行评估。在所有单跳代理场景中，MNB 方法的准确率均超过了 94%。在 Tor 和 JonDonym 环境中，准确率仅分别达到了 3%和 19.97%[15]。

此外，在没有包长信息的情况下，Dyer 等仅仅利用粗粒度特征，包括总带宽、总加载时间和突发带宽等，使用最简单的 NB 分类器即可以得到与

Panchenko 等的 SVM 方法相似的准确率性能。该文献揭示了网页流量的另一类有区分能力的特征，即粗粒度特征。由于该类特征是基于网页实例总体信息提取的，因此，所有包层级的防御技术无法有效掩盖该特征[11]。

表 4 基于传统机器学习的网站指纹识别方法

分类器[1]	模型	PETs	特征	数据集[2]	性能
LL_NB[3]	NB	SSH	包长度和方向	1000	<73%
XW[6]	RF	SSH, Tor	RTT 等	50	56.2%-95.84%
G_MA[7]	NB	SSH	TCP 连接数等	50	40%-75.9%
VNG++[11]	NB	HTTPS 等	粗粒度特征	512	约 80%
FL[12]	KNN	HTTPS,Tor	时间戳序列	100	91%
KFP[13]	RF	Tor	包统计等 150 个特征	100	91%
HY[14]	SVM	Tor	请求对象大小与包数量	100	约 65%
H[15]	MNB	VPN 等	带方向包长及计数	775	3%-94%
P[18]	SVM	Tor, JAP	Herrmann 特征等	775	54.61%-80%
CUMUL[19]	SVM	Tor	CUMUL 特征等	100	>90%
OSAD[22]	SVM	Tor	取整包长序列	100	约 90%
KNN[24]	KNN	Tor	唯一包长等 3736 个特征	100	约 90%
ZM[26]	RF	Shadowsocks, Tor	包计数等 50 个特征	100	67%
PHMM[27]	PHMM	SSH, Shadowsocks	带方向包长	200	93-99%
YK[33]	ET	8 种通信场景	广泛的精炼特征	100, 2000	95-98%
DLSVM[34]	SVM	Tor, SSH	取整包长序列	100	>80%
G_SCS[43]	NB	SSH	包长序列等	1000	81.7%
AIS[44]	SVM	Tor	带方向包长	20	41.6-74%
DTW[45]	KNN	DSL	RTT 时间	24	>80%
WPF[46]	KNN	HTTPS	包长的累积和	100	>91.63%
SL[47]	RF	Tor, HTTPS	精炼特征	100-712	69.8%-86.1%
YZ[48]	HMM	HTTP	侧信道特征	6, 1000	75%

[1]分类器默认使用原文名称或已发表文献中使用的名称。若无可查名称，本文以作者姓氏首字母结合分类器特点自拟。
[2]数据集默认单位是网站数量。

Gu 等利用 NB 分类器在 SSH 流量中进行了网站指纹研究。在特征的提取上，作者通过分析 SSH 代理的特性，对上下行流量分别提取了不同的特征。在分类时，先利用上行流量特征的相似度进行初步筛选，再根据下行流量特征进行最终判别。上行流量特征相似度的计算是由三种特征的相似度加权得到。最终判别则是把下行流量特征直接输入到 NB 分类器得到。在 OpenSSH 的数据集上，该方法取得了 81.7%的准确率，优于 Liberatore 和 Levine 的 NB 和 JC 方法[43]。

Gu 等同样利用 NB 分类器研究了 SSH 网络环境下的多标签网站指纹问题。文章首先通过数据预处理提取了 TCP 连接数、单向总带宽和包间时间间隔等三种特征，并计算马氏距离来综合判别流实例是否包含两个网页。一旦判别为两个网页流实例，则分别对首页和第二页提取不同的特征，并分别使用 NB 分类器进行判别。在 SSH 网络环境下，实验采集了 50 个网站的数据集进行评估，在第一个页面上取得了 75.9%的准确率，在第二个页面上准确率为 40.5%[7]。

综上，在特征的使用上，上述研究使用的特征种类和数量都不相同。由于 NB 分类器的属性独立性要求，各研究所选择的特征之间的依赖性强弱是影响分类器性能是原因之一。在应用场景上，除了 Dyer 等和 Herrmann 等研究了 Tor 场景，其他的研究主要基于 SSH 场景。此外，大部分都是研究单标签网站指纹，只有 Gu 等在文献[7]研究了二标签网站指纹。从数据集上看，各研究使用的网站数量均较少。虽然 Herrmann 等使用 775 个网页，但是样本数量太少。

支持向量机。支持向量机(Support Vector Machines, SVM)是机器学习中一种经典的有监督二分类方法[50]。SVM 通过选择超平面，使得空间中两种类别的训练实例尽可能地分离开。该超平面可以用来对新的实例进行类别判别。

针对 Herrmann 等的 MNB 方法准确率不高的问题，Panchenko 等使用 SVM 方法应用 Herrmann 等在文献[15]提出的特征在 Tor 数据中进行网站指纹识别实验，准确率从 2.96%提升到了 30.98%。进一步地，本文不断增加使用特征的种类。相应地，识别准确率也不断提高，最高达到 47.36%。作者又通过对数据进行了去噪，再次将准确率提高到 54.61%。在 JAP 的数据集上，准确率最高达到 80%。在开放世界的评估中，作者采集了 5 个网页作为监控训练集。其中，每个网页有 35 个实例。此外，训练集也包括了 1000 个非监控页面实例。测试集包括了 5 个监控页面和 4000 个非监控页面。其中，每个监控页面有 25 个实例。每个非监控页面有 1 个实例。在 Tor 网络环境下，最好情况下，SVM 方法在开放世界的真阳率是 73%，假阳率仅为 0.05%[18]。

Cai 等同样使用 SVM 进行网站指纹识别研究。与传统方法不同，Cai 等直接利用包序列信息作为

SVM 分类器的输入。相似度的计算则是使用受限版本的 DL(Damerau-Levenshtein)编辑距离[22]。这是因为 DL 距离的插入、删除、替换和调换等操作对应着包或请求的重新排序、请求的忽略和请求回复包大小的轻微扰动。在包序列信息的处理上，如果是 Tor 包，则去除其中的 ACK 包；如果是 SSH 包，则去除长度小于 84 的包。在此基础上，将序列的包大小四舍五入到 600 的倍数。在基于 Tor 和 SSH 的网络环境中，本文综合评估了 8 种不同场景的性能指标。结果表明，DLSVM 方法优于 MNB 方法和 Panchenko 等的 SVM 方法[34][51]。

Wang 等采用 SVM 技术进一步改进了 Cai 等的方法[22]。Wang 等的主要改进体现在实例之间的距离计算和实例的表示方面。在距离计算方面，考虑到多次访问同一网页可能产生的扰动情况，作者改进了编辑距离的计算，即删除了替换的操作。在代价的计算上，本文对于调换操作赋予了更小的代价，而对于插入与删除则赋予了相同的代价，但是对于出包赋予了更大的代价。在实例的表示方面，与 Cai 等不同的是，Wang 等使用了 Tor 流实例的 Cell 方向序列。在同样的数据集中，Wang 等的方法比 Cai 等的方法准确率提高了 3%到 5%。

与传统的被动式网站指纹识别方法不同，He 等使用 SVM 方法研究主动式网站指纹识别。该方法的主要思想是通过主动延迟 HTTP 请求的发送，使得识别者可以获得各个请求对象的大小和包数量等信息。为了延迟 HTTP 请求，需要解决一些关键问题：包括延迟包的正确定位、避免来包的重传、识别 Tor 的控制 Cell 以重启 HTTP 请求等。在特征的选择上，这里提取的特征主要是针对从服务器返回的流量。在 SVM 方法的实现上，本文改进了距离的计算。最后，作者通过在 PlanetLab 上模拟 Tor 网络进行评估验证，在与 Panchenko 等的 SVM 方法的比较中，取得了更好的效果[14]。

Jahani 等通过引进快速傅立叶变换(Fast Fourier Transform, FFT)来计算实例之间的相似度，采用的特征是带方向的包长。FFT 使得特征从时域变换到了频域以方便和加速分析。在分类器的选择上，本文借助 SVM 方法并采用“一对一”的多分类策略。在多个 Tor 数据集的评估中，该方法准确率高于 95%，优于 Wang 等和 Cai 等的 DLD 方法[16]。

Panchenko 等提出的 CUMUL 识别方法也使用了 SVM 分类器。该方法仅使用了 104 个特征，但是识别的准确率却很高。除了四个统计特征之外，剩下的 100 个插值特征是从带方向包长序列的累积表示中抽样得到的。在 Tor 数据集的评估中，CUMUL 方法准确率为 91.38%，优于 KNN 方法的准确率。同时，该文章还评估了开放世界场景中非监控网站包含非主页实例、封闭世界场景中监控网站包含多个非主页实例等情况[19]。

综上，在特征使用上，Panchenko 等在文献[18]使用了一些定制的特征，在文献[19]创造性地提出了 CUMUL 特征。Jahani 等则使用了带方向包长特征。He 等是通过主动延迟请求来获取对象的大小和包数量做为特征[14]。Wang 等和 Cai 等使用的特征的区别在于前者是采用 Cell 单元的方向信息[22]，而后者使用的是数据包四舍五入到 600 的长度值[34]。可以看到，除了文献[18]由于使用的特征种类复杂导致引入了较多的特征提取工作量之外，其他文献使用的特征种类比较简单，识别准确率却更高。这表明除了通过提取更多特征来提高网站指纹识别的准确率之外，还可以通过优化特征向量之间距离的计算和创造一类更有区分能力的特征来实现。在应用的场景上，上述研究都是基于 Tor。特别地，He 等的实验场景是模拟的，而非实际的。此外，He 等的主动式网站指纹虽然不符合网站指纹的一般定义，但是仍然具有现实意义。

最近邻居算法。最近邻居算法即 KNN(K-Nearest Neighbor)算法。该算法的核心思想是测试样本的分类输出由训练样本中与其最近的 K 个邻居样本的多数决定。

Gong 等采用了 KNN 邻近算法进行了远程流量分析识别。该研究中的识别者位于远程有利位置，并不在网站指纹通常规定的识别位置场景，即从用户到网络的入口之间的本地位置。因此，该方法不能利用传统网站指纹识别者所能观察到的流量特征。作者通过向用户主机发送 ping 包，根据 RTT 的时间反推恢复流量的模式，并利用完整的时间序列数据进行识别。在距离度量上，本文采用了 DTW(Dynamic Time Warping)的计算方法，并使用 KNN 方法进行分类。在 12 个和 24 个网站规模的数据集上，均取得了 80%以上的准确率[45]。

Shen 等同样采用 KNN 方法进行同一网站的站内网站指纹识别研究[46]。在特征的选择上，本文的 WPF 方法利用了包长的累积和特征。实验数据是在京东网站随机提取 100 个网页，每个网页 60 个实例。结果表明，WPF 方法的准确率达到 91.6%，优于 DTW 方法和 appscaner 方法。

Wang 等对 KNN 进行了改进，用来在 Tor 网络环境实施网站指纹识别。该方法在时间复杂性和准确率方面较以往的方法有了提升。作者对 KNN 的改进主要体现在对每种特征对相似度的贡献设置了权重，而且该权重是可以动态调整的。在特征的使用上，本文提取了将近 4000 个的流量特征。在 Tor 的实际数据集上，该方法取得了比文献[22][34]的方法更好的性能[24]。

Feghhi 等仅利用上行流量的时间戳信息进行网站指纹识别，通过引入微分动态时间翘曲的方法，计算两个流实例之间时间序列的 F-距离。在分类器的选择上，本文使用了 KNN 方法。该方法可以在 5

天的时间跨度内保持 91%的预测成功率[12]。

Al-Shehari 等借鉴 Cai 等的 SVM 方法研究了不同浏览器软件对于流量分析识别的抵抗问题。通过数据预处理提取了流量带方向的包长特征。在多种主流浏览器采集的数据实验中，分别获得了 41.6%-74%不等的准确率。同时，本文对产生不同准确率的根本原因进行了分析[44]。

综上，在特征的使用上，Gong 等突破了网站指纹定义的限制，通过主动方式获取特征。Shen 等和 Al-Shehari 等使用了与包长有关的特征。Wang 等则提取了近 4000 种与包长度、包时间和包顺序相关的自定义特征。在距离的计算上，Feghhi 等使用了 F-距离，Gong 等使用了 DTW，Al-Shehari 等使用了编辑距离，Wang 等和 Shen 等则使用欧式距离。从以上对比情况可以看到，虽然各识别方法使用的特征不同，距离策略也各异，然而实验结果均表明了网站指纹识别的可行性。这些研究工作表明，网站指纹识别之所以可行，本质还是在于网页流量模式泄露了一些个性化信息。从应用场景上看，文献[12]、[44]、[45]和[46]主要考虑 HTTPS 场景。Wang 等则考虑了 Tor 场景。除了 HTTPS 场景之外，Feghhi 等也考虑了 Tor 场景。

随机森林。随机森林(Random Forest, RF)是由一组相互独立的决策树构成的分类器[52]。决策树节点的分裂可以基于信息增益等不同的策略。随机森林实现方法决定了其具有便于进行并行化、可以生成特征的重要性等优点。

Hayes 等提出 KFP 网站指纹识别方法。该方法利用 RF 作为分类器，提取了流量的 150 个重要性得分最高的特征作为模型输入。在封闭世界评估中，KFP 使用 RF 的输出作为分类结果。然而在开放世界的评估中，KFP 不直接使用 RF 的输出作为分类的结果，而是利用 RF 的所有叶子形成的特征向量，并通过计算特征向量之间的马氏距离进行分类决策。在判别类别上，只有与测试实例最近的 K 个训练实例都属于同一个类别时，该测试实例才被分配给该类别。在对网页和隐藏服务的不同数据集上，KFP 均取得了比其他识别方法更好的性能[13]。

Zhao 等应用 RF 研究了实际 Shadowsocks 网络环境下的网站指纹识别问题。结果表明，基于 Shadowsocks 的网站指纹不比 Tor 更简单。为了提高对实际 Shadowsocks 流量的识别性能，文中提出了基于特征重要性和集体特征效用的特征精炼方法，并抽取了关键特征。通过特征精炼，本文的识别方法在保证准确率的前提下，提高了效率。另外，通过基于 OOB(Out of Bag)得分调节每个类别的训练实例数量，得到最佳的训练实例数量，并将对实际 Shadowsocks 流量的识别精度从 55%提高到 67%[26]。

Xu 等利用 RF 对两标签网站指纹进行研究。在两标签场景下，文章的测试实例是由两个网页实例组成，而且两者之间有重叠。作者通过特征选择与 RF 结合，把测试实例划分到输出标签向量累加和最大的类别。在定位第二个重叠页面的起点时，本文使用了 XGBoost 的方法[53]。在 Tor 和 SSH 的数据集上，该工作对重叠页面分别取得了 95.84%和 56.2%的准确率[6]。

Shen 等基于随机森林模型提出了一种系统化的网站指纹方法。该方法的主要特点是对特征进行了一系列的操作以精炼特征的数量，包括特征预处理、特征评价和特征组合等。通过去除无意义的、低贡献的特征，在不损失准确率的前提下，提高了时间效率。在多个实验数据集上，该方法均得到了比经典方法更高的准确率[47]。

综上，在特征使用上，Shen 等提出了一套特征精炼的策略，包括预处理、评估与组合等。Zhao 等则从包长度、包时间、包顺序、总体统计和头尾部统计等信息中提取了约 800 种特征。Xu 等仅提取了第一页的起始包块特征。Hayes 等则不直接利用提取到的特征，而是创造性地使用随机森林的输出作为直接特征。在研究的场景上，Zhao 等考虑了 Shadowsocks 场景，特别是其中应用了实际采集的 Shadowsocks 数据集是该工作的一个亮点。其他研究都是基于 Tor 应用场景。值得一提的是，在分类器原理上，Hayes 等的 KFP 识别除应用随机森林之外，也综合应用了 KNN 的决策思想。

隐马尔可夫模型。隐马尔可夫模型(Hidden Markov Model, HMM)是一种统计模型，实质是个马尔可夫过程[54]。

Yu 等在匿名网络场景中利用用户浏览网页的时长这一侧信道信息进行网站指纹识别的研究，并对网站建立了 HMM 进行求解。在构建 HMM 时，作者把网页看成是 HMM 的一种隐藏状态，观测的浏览时长作为 HMM 的观测变量。在 HMM 发射概率计算上，作者通过多次模拟访问网页的浏览时长的概率分布获得。另外，本文基于齐普夫分布和模拟估计，得到 HMM 的初始概率和转换概率。最后，利用估计的参数计算隐藏状态序列，即用户的浏览网页顺序。在实际数据集测试中，网站指纹识别的准确率达到了 80%以上[48]。

Zhuo 等考虑到网络包序列与基因序列的相似性，采用了生物信息学中基因序列比对常用的 PHMM(Profile Hidden Markov Model)模型对网站指纹问题进行建模。为了应用 PHMM 模型，作者对带方向包长特征进行了符号化操作，将包序列转化为符号序列。在具体的网站分类中，借鉴了生物基因序列比对研究中使用的序列对齐和比对工具，直接输出分类的结果。在 SSH 和 Shadowsocks 的数据集上，取得了比传统方法更好的结果[27]。

HMM 应用于网站指纹识别关键是求出模型的

参数。为了求出模型的参数，Yu 等使用了模拟法，Zhuo 等使用了基因序列对齐工具 ClustalO[55]和比对工具 HMMER[56]。从求解原理上看，Zhuo 等是通过将包长特征转化为基因蛋白质符号，从而可以借鉴基因比对工具直接算出参数。这一方法高度依赖于包长特征，产生了一定的局限性。而模拟法具有坚实的理论基础，适用性更广。在实验场景上，Yu 等基于 SSL/TLS 链路，而 Zhuo 等基于 SSH 和 Shadowsocks。

极端随机树模型。与 RF 类似，极端随机树(Extremely Randomized Trees，ET)也是由多个决策树构成的分类器[57]。它们都具有特征随机、模型随机和参数随机等特性。但是，与 RF 不同的是，ET 使用的是全部的样本进行分类，而 RF 使用随机抽样的样本进行分类。

Yan 和 Kaur 在研究网站指纹中特征选择问题时使用了 ET 树模型。在该文章中，作者对特征进行了广泛的提取和分析，并将特征分为包层、突发层、TCP 层、端口层和 IP 地址层等五个层次 109 个类别共 35683 种。在输入到 ET 分类器前，对特征进行了筛选和特征聚类等操作。在与六种经典网站指纹分类器的对比中，该文的 Wfin 方法在所有 8 种通信场景中均取得了最好的效果。该结果表明了特征工程在网站指纹研究中发挥了重要的作用，也证明了 ET 分类器的有效性[33]。

3.2.3 深度学习方法

深度学习是机器学习的分支，是一种以人工神经网络为架构，对数据进行表征学习的算法。表征学习的目标是寻求更好的表示方法并创建更好的模型以便从大规模未标记数据中学习这些表示方法。深度学习的好处是用非监督式或半监督式的特征学习和分层特征提取高效算法来替代手工获取特征。深度学习方法在计算机视觉、语音识别等领域取得了较好的效果。

鉴于深度学习方法优异的性能，研究人员尝试将深度学习方法应用于网站指纹识别的研究中。与传统机器学习分类方法直接依赖于输入的特征数据不同，深度学习方法则依靠其多层的神经网络结构不断地通过线性和非线性变化逐一自动学习更深层特征，从而不断地逼近数据的真实分布。这种自动学习特征的能力使得深度学习方法具有了更多的潜力。根据所基于的神经网络结构不同，现有网站指纹识别方法主要基于堆叠降噪自编码器、卷积神经网络和循环神经网络等三种神经网络结构。直观起见，相关的研究工作归纳总结在表 5。

堆叠降噪自编码器。自编码器是由编码器和解码器组成。堆叠去噪自编码器(Stacked Denoising AutoEncoder, SDAE)是在自编码器的基础上增加堆叠和去噪功能而形成的。

Abe 和 Goto 首次尝试使用深度学习方法研究网站指纹。他采用的模型是堆叠降噪自编码器，通过引入堆叠和降噪降低了过拟合，提高了准确率。本文中使用的特征很简单，仅需要 Cell 元胞的方向序列即可。在封闭世界设置下，使用两层和三层堆叠的最高准确率均是 88%，略低于 OSAD 方法和 KNN 方法。在开放世界设置下，SDAE 方法的真阳率高于对比方法，假阳率性能却低于对比方法[8]。

Rimmer 等对深度学习在网站指纹的应用进行了系统性的探讨。其中，基于 SDAE 的网站指纹识别方法在各数据集上取得了 94.25 至 95.766%的准确率，优于传统机器学习方法中表现最好的 CUMUL[20]。研究成果表明，在增加每个类别训练样本的数量情况下，基于深度学习的网站指纹识别方法的性能可以得到提高。

Abe 等和 Rimmer 等都使用了 SDAE 模型，但是主要有以下区别：在降噪策略上，Abe 等是通过引入白噪声的方式，Rimmer 等是通过 Dropout；在模型训练上，Rimmer 等借鉴 GE Hinton 等的研究成果分二步训练，先是输入无标签数据实施贪婪的、分层的和无监督预训练来初始化模型，再使用有标签数据，通过反向传播分类错误再细粒度调优模型[58]。Abe 等并没有进行预训练。在实验评估上，Abe 等评估了二层和三层的堆叠式结构，而 Rimmer 等仅实验了三层的网络结构。在实验效果上，Rimmer 等的 SDAE 模型取得了更好的准确率性能。但这一结果可能与实验数据集的规模有关。Rimmer 等使用了迄今为止最大的网站指纹数据集，而 Abe 等使用的数据集小得多。

卷积神经网络。卷积神经网络(Convolutional Neural Networks, CNN)是一种前馈神经网络[59]，常用于特征提取。

Rimmer 等在文献[20]也提出了基于 CNN 的网站指纹识别方法。在 100 个网站的数据集中取得了 96.66%的准确率，优于其他对比方法。该结果表明 CNN 在网站指纹识别应用中具有可行性。

Sirinam 等基于 CNN 神经网络提出了 DF (Deep Fingerprinting)网站指纹识别方法。与 Rimmer 等的 CNN 结构不同的是，Sirinam 等没有在 1 个卷积层后直接接上池化层，而是至少经过 2 个卷积层后才接上池化层。这样可以加深神经网络的层数从而更有效的提取特征。为了防止过拟合，DF 使用了批处理标准化(Batch Normalization, BN)层和 Dropout 层。在过滤器数量的使用上，DF 随着层数的加深不断增加过滤器的数量以提高获取特征的能力。在激活函数上，DF 在接近输入的层级上使用 ELU 函数，减少了信息的丢失。通过一系列的设计，DF 在封闭世界和开放世界场景中均取得了很好的效果[21]。

Bhat 等在 ResNet 的基础上，提出了 Var-CNN 的网站指纹识别方法。该方法从两个角度对 ResNet 进

行了增强。首先，通过引入扩展因果卷积，扩大了各层的感受野。此外，Var-CNN 在网络设计中有效结合七种累积统计特性，实现利用累积统计特性的弱预测来加强神经网络的强预测的目的。在与 DF 方法的比较中，Var-CNN 在开放和封闭世界实验中均取得了更好的效果。同时，Var-CNN 仅需要较小的训练样本，减少了数据过时问题的影响[60]。

Rahman 等基于 DF 分类器研究了包时间信息对网站指纹的作用。本文作者提出了七种突发层级的时间特征，并对其信息泄露进行了研究。结果发现提出的特征冗余较小，对分类器的鲁棒性有积极意义。利用传统机器学习方法，基于这些特征的组合，网站指纹识别的准确率可达到 60%以上。在 KFP 方法使用的时间特征基础上增加这些新特征之后，深度学习方法的准确率可以提高 2%以上，而机器学习方法准确率变化不大[61][62]。Wang 等提出的 2ch-TCN 网站指纹识别方法同样使用卷积神经网络。2ch-TCN 的特点是采用了双通道设计，分别用以处理时间特征和 Cell 方向特征。这种设计提高了识别的准确率。在典型公开数据集和作者自采数据集上均取得了良好的效果。该工作的另一个贡献是首次发现传统的Cell方向序列提取方法存在误差。这也表明网站指纹识别方法的准确率有进一步提升的空间[63]。国内学者 Ma 等也提出了基于卷积神经网络的指纹识别方法 DBF。该方法包括三个模块，即突发特征提取模块、突发特征抽象学习模块

表 5 基于深度学习的网站指纹识别方法

分类器[1]	神经网络	PETs	输入	数据集[2]	性能
SDAE[8]	SDAE	Tor	Cell 方向序列	90	88%
Var-CNN[10]	CNN	Tor	Cell 方向序列，包时间间隔序列	900	98.8%
AWF_LSTM[20]	LSTM	Tor	Cell 方向序列	900	88.04%
AWF_SDAE[20]	SDAE	Tor	Cell 方向序列	900	94.25%
AWF_CNN[20]	CNN	Tor	Cell 方向序列	900	91.79%
DF[21]	CNN	Tor	Cell 方向序列	95	98%
Tik-Tok[61]	CNN	Tor	突发层时间序列，原始时间序列	95	84.3%, 96.5%
2ch-TCN[62]	CNN	Tor	原始时间戳特征、Cell 方向序列	100	86.31%, 97.15%
DBF[64]	CNN	Tor	Cell 方向序列	900	98.31%

[1]分类器默认使用原文名称或已发表文献中使用的名称。若无可查名称，本文以作者姓氏首字母结合分类器特点自拟。
[2]数据集默认单位是网站数量。

和突发特征深度分析模块。DBF 另一个特点是对开放世界和封闭世界分别采用不同的模型。对于开放世界不是基于阀值进行分类，而是将封闭世界的输出向量及其统计向量合并后引入随机森林模型，最终利用多数投票法决定测试样本的类别。实验结果表明，该方法对缓解概念飘移、抵抗网站指纹防御等方面具有良好的性能[64]。

综上，Rimmer 等首先探索了 CNN 在网站指纹的可行性。在此基础上，Sirinam 等通过优化 Rimmer 等提出的网络结构将分类性能提升到一个新的水平。考虑到深度学习模型对训练数据数量要求较高的缺点，Bhat 等通过引入 ResNet[65]，提出了 Var-CNN 模型，较好地解决了该问题并进一步提高了分类性能。。同时，Var-CNN 也是首个结合自动特征学习与手工提取特征的网站指纹识别方法。与以往创新深度神经网络模型的研究工作不同的是，Rahman 等重点研究了时间特征与已有 CNN 网络模型结合在网站指纹研究的效果，通过结合文献[13]的时间特征和作者提出的时间特征，DF 深度学习分类器的性能提高了 2%。Wang 等同样强调了时间信息的重要性，但是在利用方法上与 Rahman 等不同。Rahman 等是把时间信息和 Cell 方向信息整合成一个序列。Wang 等的 2ch-TCN 识别则把两种信息看成是双通道特征。Ma 等提出的 DBF 深度学习模型最主要的特点是对不同的特征应用不同的卷积核进行特征提取，并使用深度学习与传统机器学习方法结合的方式来评估开放世界性能。

循环神经网络。循环神经网络(Recurrent Neural Network, RNN)是一种考虑输入序列的时序关系的神经网络，具有记忆性和参数共享等特性。网站指纹研究常用的 RNN 结构是长短期记忆网络(Long Short-Term Memory Networks, LSTM)[66]。

Rimmer 等借鉴 LSTM 网络结构对网站指纹识别进行研究。由于 LSTM 方法训练的时间复杂性，作者仅提取了每次访问的头 150 个 Cell 单元以提高效率。尽管提取的 Cell 单元数量远低于其他深度学习方法，仍然取得了与之相当的准确率。这主要得益于 LSTM 考虑了单元之间的关联性[20]。

RNN 结构考虑了输入序列之间的时序关系。实际上，对于同一网页不同次访问产生的包序列而言，大量数据包之间的顺序是有固定的时序关系。这种内在本质解释了 RNN 模型在网站指纹研究的可行性。Rimmer 等的工作为 RNN 在网站指纹识别研究中的应用拉开了序幕。

3.3其他网站指纹识别研究动态

除了网站指纹识别方法的主体研究之外，网站指纹识别还有其他重要问题的研究。例如，如何在弱化识别假设的场景下提高识别的准确率？如何解决“基础发生率谬误”问题？如何去除无关的背景流量？如何解决训练数据过时问题？如何使用少量的训练样本进行网站指纹识别？等等。

(一)先验知识不足

先验知识不足，是对识别者能力假设的放宽。即识别者不一定具有和用户完全一样的知识信息。这意味着识别者无法完全复制用户的网络环境、浏览器配置等信息，从而导致识别的难度增加，甚至造成分类器的失效。针对该问题，Marc 等研究了TBB(Tor Browser Bundle)版本、TBB 配置和网络位置等三种先验知识不足的情况下网站指纹识别准确率变化的情况[2]。实验结果表明，TBB 版本不同将导致识别准确率最大下降 50%以上；不同的 TBB 入口节点配置，识别的准确率最大相差 11.7%；网络位置的差别，识别的准确率最大相差接近 60%。

类似地，Marks 和 Halvemaan 则研究了 HTTP 版本对于网站指纹识别性能的影响[67]。HTTP 版本主要包括 HTTP1.1 和 HTTP2.0 两种。结果表明，当训练集与测试集的 HTTP 版本不同时，将使识别准确率下降 20%至 40%。此外，Miller 等的研究则表明了缓存和 Cookie 对网站指纹识别准确率的影响可达 17%[68]。

此外，Reddy 也在他的学位论文里研究了网站配置对网站指纹识别的影响。作者考虑了协议配置、网站内容、网站大小和托管策略等因素。结果表明，相比 HTTP1.1，HTTP2.0 和服务器推送可以显著降低指纹识别的准确率。但是对于其他因素，并没有给出结论[69]。

虽然以上四篇文章的研究工作都揭示了识别者的先验知识不足对于网站指纹识别准确率的显著影响，但是考虑的因素是不同的。它们的共同之处在于均没有给出相应的解决方案。

本质上，识别者对用户的先验知识不足之所以会导致网站指纹识别准确率下降，原因在于先验知识不足将会使得训练集和测试集中同一网页的流量模式发生显著的改变。先验知识不足的因素很多，已有研究考虑的情况还比较少，而且并没有给出解决方案。从这个角度出发，针对该问题的网站指纹识别研究还有很多值得开展的工作。

(二)低基础发生率

所谓低基础发生率，是指阳性事件的发生概率在全部事件中占有的比例极低[4][19][21][70]。在网站指纹识别研究中，它表示用户访问识别者监控的网站的概率极低。在低基础发生率的情况下，网站指纹分类器的精度性能会出现严重下降。这个问题就是“基础发生率谬误”(The Base Rate Fallacy)。相关的研究工作总结如下。

Cai 等在文献[70]首次指出监控网站的流行度将会影响分类器的性能。网站流行度越高，分类器性能越好。实际上监控网站的流行度反映了阳性事件的发生率。即监控网站越流行，被访问的机会越大，阳性事件的发生率就比较高。Juarez 等首先在文献提出了“基础发生率谬误”概念，并通过实验说明了该问题。作者结合召回率和假阳率，及其监控样本和非监控样本的比例等指标，提出了 BDR 指标来评价网站指纹在开放世界评估实验的效果。结果表明，在阳性事件先验发生概率为 0.005 的情况下，分类器的 BDR 只有 0.13%。

针对“基础发生率谬误”问题，Sirinam 等在文献[21]使用 P-R 曲线来评价分类器的性能。然而，该文献中阳性事件发生的概率比实际场景的基础发生率仍然要高很多。Panchenko 等在文献[19]中使用精度和召回率的 CCDF(Complementary Cumulative Distribution Functions)图来评价两种不同的开放世界场景，即开放世界的网页是否包含与监控页面属于同一网站的网页。

精度和贝叶斯检出概率虽然综合考虑了召回率和假阳率的影响，但是仍然没有显式的引入反映基础发生率的参数。因此，分类器是否可以克服“基础发生率谬误”的问题，还需要结合实验的数据集进行判断。

$$r-precision=\pi_r=\frac{R_{TP}}{R_{TP}+R_{WP}+rR_{FP}}\#(7)$$

针对此问题，Wang 等在文献[4][25]创造性地在精度定义中考虑了基础发生率的因素，提出了$r-precision$的计算公式，如公式(7)所示。其中，R_{TP}、R_{FP}和R_{WP}分别表示真阳率、假阳率和错阳率，r定义为阴性事件与阳性事件的比值。研究发现以前被认为最好的分类器在$r=1000$时，其$r-precision$只有 0.14。因此，已有研究的分类器无法直接在实际场景中进行网站指纹识别。基于这一认识，作者提出了三种精度优化器，并对六种经典的分类器进行精度优化。在$r=1000$时，经优化后 KFP 分类器，精度达到了 86%，比未优化前的提高了 6 倍。

“基础发生率谬误”是网站指纹识别方法实用化过程必然面临的一个问题，即已有网站指纹识别方法必须针对该问题提出合理的精度优化策略，使其在实际应用场景中可以保持较高的准确率，否则将导致识别方法完全失效。Wang 等的工作考虑了基础发生率的因素并创造性地提出了$r-precision$评价指标，也首次提出该问题的解决方案。然而，各识别方法经优化后的精度仍然低于 90%，因此，针对该问题的研究仍有广阔的空间。

(三)复杂背景流量

网站指纹识别实际应用时面临的另一个问题是复杂的背景流量。复杂背景流量存在两种情况，即背景流量包含其他应用的流量或者同一浏览器不同标签访问的流量。一方面，识别者在截取网络流量时，很难保证获得的全部是浏览器应用的流量数据。针对这一问题，大部分已有研究是采用理想化的假设进行保证的。学者 Wang 等首次在文献[28]探索了噪音流量去除问题的研究。在该文献中，作者提出了 counting-based 和 classification-based 两种

方法。然而，实验表明该两种方法均在分离噪声流量的同时，也大量地丢失了真正有用的流量数据。这说明净化背景流量仍然存在较大的困难。

另一方面，用户在浏览网站时，大量的使用多标签访问方式，从而造成了同一时刻流量数据里夹杂着不定数量的网站流量。完全剥离出单独的网站流量是一项额外的工作，而且难度较大。大部分历史研究采用了理想化的假设，即用户一次仅访问一个页面。当然，也有部分研究探讨了多标签流量的情况，甚至给出了特定解决方案。

Marc 等首先考虑了多标签访问的情况[2]。在 Tor 网络环境下，作者以单标签访问数据作为训练集，以多标签访问数据作为测试集，选取了五种不同分类器，测试了两个标签访问在不同间隔时的识别成功率。实验结果表明，在不加其他优化的情况下，所有情况下的识别准确率均下降到了 10%以下。由此，直接将已有识别方法应用于多标签环境下进行工作是不可行的。该文献仅说明了问题的存在，并没有给出解决方案。此外，它也仅仅考虑了二标签访问的情况。

学者 Gu 等研究了 OpenSSH 场景下二标签访问的情况[7]。作者通过提取不同的特征，在第一和第二次访问延迟为 2 秒的情况下，识别第一次访问页面准确率为 75.9%，第二次访问页面识别准确率为 40%。该结果比文献[2]的准确率有所提高。然而，该文献的数据集仅为 50 个网站，数据规模较小。同时，识别准确率和假阳率距离实用仍有较大的差距。与文献[2]类似，该文献也仅考虑了二标签访问的情况。

学者 Xu 等同样研究了二标签访问的网站指纹识别问题[6]。该文献通过识别第一页与第二页起点之间的包块信息并利用该信息进行网站指纹识别，在 SSH 流量数据中取得了 92.58%的召回率，在 Tor 流量数据中得到了 64.4%的召回率。该方法要求对开始重叠点的检测准确率要较高，同时要求第一页与第二页之间的间隔时间不能太短，才能有较高的识别准确率。在流量分割算法上，该文献使用的 BalanceCascade-XGBoost 方法优于文献[28]提出的 time-kNN 方法。

与上述文献不同的是，学者 Cui 等不仅考虑二标签重叠的情况，也考虑了二标签连续的问题[5][71]。该文章的总体思想是将流量实例切分成多个部分，对每个部分进行预测，最后以多数投票的方式决定流量实例的分类。通过使用切分，准确率由 22.94%提高到 70%。

不同于其他文献仅考虑二标签访问的情况，Ramezani 等在 HTTPS 场景下首次研究了三标签和四标签同时访问的网站指纹识别问题[72]。该文献利用的特征是 TLS 握手协议中的非加密特征，即服务器名称。作者分别实验了 LSTM 和全连接模型两种神经网络结构，在实际的二标签数据集测试中，分别获得了 94.8%和 86.4%准确率。在实际的多标签数据集测试中，LSTM 获得了 87.2%的准确率。

综上，在有背景流量条件下的网站指纹识别研究中，针对多应用流量混合的研究只有 Wang 等的工作，然而效果不佳。针对多标签流量混合的研究较多。其中，Marc 等的工作首次提出了多标签网站指纹识别的问题，但是没有给出解决方案。随后的研究中，研究人员针对该不足，均提出了具体的方法，得到了相应的准确率性能。从多标签的问题模型上看，Gu 等和 Xu 等考虑的是二标签重叠问题；Cui 等除了考虑二标签重叠，也考虑了二标签连续的情况；Ramezani 等研究的特色是他们首次考虑了三、四标签同时访问的情况。从应用场景上看，各研究工作的应用场景覆盖了 HTTPS、SSH 和 Tor 等常见 PETs。

(四)训练数据过时

数据过时问题是指当训练样本数据生成时间与测试样本数据的生成时间间隔太长，用来训练的样本数据就不适合用来进行预测。因此，网站指纹识别的分类器必须定期进行更新，以保证其准确率。数据过时问题的成因是网页流量受到网页本身内容变化、网络条件变化等因素的影响，同一网页的流量分布情况会随着时间发生变化。

数据过时的问题将更严重影响分类器的性能。Marc 等在文献[2]发现 KNN 分类器在 10 天之内准确率从 80%下降到了 30%。Rimmer 等在文献[20]也证实了类似的发现，其 SDAE 分类器在 28 天内从 95%下降到 81%。因此，一个分类器最多只能在数周内有效[73]。对于一个深度学习模型，往往需要的样本数据较大，需要的数据采集时间也较多。以采集 50 万个样本为例，单终端资源识别者需要 250 天完成训练模型生成[73]。

因此，对于低资源识别者，数据过时问题是一大障碍。用了过旧数据训练出的分类器的性能必然是很差的。解决数据过时问题，主要的思路是寻找只需要少量训练样本，而且分类性能好的分类器。Sirinam 等在文献[73]利用 TF(Triplet Fingerprinting)以解决该问题。TF 利用了 N-shot learning 的技术，仅需要少量的训练样本即可进行分类。即使在训练集与测试集通过不同网络收集，而且收集时间间隔数年，TF 仍然可以达到 85%的准确率。这篇文献同时也提出了一项具有挑战性的应用场景，即如何利用陈旧的数据来训练分类器并进行指纹识别。

Oh 等提出 GANDaLF 模型，其中使用了生成对抗网络进行小样本网站指纹识别研究。GAN 模型的优点在于不需要像 TF 那样使用有标签数据集进行预训练。同时，该项研究同时考虑了站内网站指纹识别的情况。在包含站内网页的网站指纹识别实验中，GANDaLF 识别准确率高出 TF 识别

16-43%[74]。

针对训练数据过时问题，Marc 等首次提出了该问题的存在，并没有提出应对方法。Sirinam 等提出的 TF 识别方法首次使用 N-shot learning 技术较好地解决了该问题，但是准确率仍然有提升的空间。同时，TF 需要事先使用一个大的公开数据集预训练一个特征提取器。该特征提取器必须动态更新，才能保持 TF 的有效性。这个要求极大地限制了 TF 的实用性。Oh 等在小样本网站指纹识别研究上更进了一步。他们提出的 GANDaLF 模型首次引入了生成对抗网络，并且对预训练的数据集降低了更求，不需要数据集具有标签。同时，实验表明 GANDaLF 的准确率远高于 TF。

(五)站内网站指纹识别

大部分的网站指纹研究基于主页假设，即网站的分类实际上是网站主页的分类。在这种假设下，网站之间的区分性是比较大的。站内网站指纹则是在一个网站内，监控用户访问的是哪个网页。显然，由于同一个网站的网页相似度较高，此时网站指纹难度更大。Shen 等在 HTTPS 场景中进行站内网站指纹研究，并提出了 WPF 方法[46]。该方法识别准确率高于 90%，性能远优于 Appscaner[75]和 DTW 方法[12]。

Miller 等针对站内网站指纹提出了一种新的识别方法 BoG[68]。该方法利用聚类和高斯相似技术将变长流量转换为固定宽度表示。由于该表示方法与文本分析的词袋假设类似，所以也被称为高斯袋(Bag of Guassian, BoG)。同时作者使用站内链接结构构建 HMM 模型以进一步提高性能。BoG 识别在 HTTPS 场景中站内指纹的准确率比以前提高了 30%。

站内网站指纹虽然不符合主页假设，但具有实际意义。例如，假设某一恶意网页潜伏在某个合法的网站之内。如果要将其识别出来，就需要更细致的数学模型。已有工作均基于相对简单的 HTTPS 场景，并没有考虑更为复杂的 Tor、Shadowsocks 和 SSH 等网络场景。

(六)新型网站指纹识别

新型网站指纹研究突破了传统网站指纹的定义和假设限制，尝试以不同的方法和特征实现网站指纹识别。主要的工作包括以下几项。

与传统的被动网站指纹相比，新型网站指纹识别可以采用主动式。例如，He 等通过主动延迟 HTTP 请求的发送来获取所需要的请求对象大小和包数据等特征来实现网站指纹识别，取得了较好的效果[14]。其次，新型网站指纹识别可以不局限于使用流量特征。例如，Gong 等则通过测算 RTTs(Round-trip times)估计流量模式以进行网站指纹识别，在对家庭 DSL 用户的识别上取得了一定的效果[45]。此外，还有一种网站指纹研究的新型思路是重新考察并利用被忽略的流量信息以提高识别性能。例如，Al-Naami 在其研究中发现了确认包可以用来提高识别的准确率，而确认包往往在网站指纹研究中被首先过滤掉[76]。值得一提的是，在针对 Tor 隐藏服务的网站指纹识别研究中，Kwon 等提出了电路指纹识别方法[77]。该方法利用了隐藏服务的电路与普通的 Tor 电路在通信与互动模式的显著区别，无论在用户端或服务端，准确率达到 98%以上。

除了突破传统网站指纹识别定义限制的新型研究之外，针对网络新协议和新技术的进展，强化对网站指纹识别研究的未来前景的探索也是十分有意义的。

随着 QUIC 协议和 HTTP/3 协议的出现，已有学者开始对基于 QUIC 场景的网站指纹识别进行研究。Zhan 等利用两种特征对 HTTPS 协议和 QUIC 协议进行了定量的脆弱性比较分析。同时，作者又利用网页加载的前部包数据对两个协议进行了定性分析。结果均表明 QUIC 协议比 HTTPS 协议更加脆弱。该文的早期场景的网站指纹识别研究实际上也是突破了传统网站指纹识别的定义限制[78]。Smith 等针对 QUIC 协议的出现，研究了 QUIC 协议与 TCP 协议共存下的网站指纹识别问题，提出了 mixed 和 split 两种方法。该研究也表明利用 TCP 数据集训练的模型对 QUIC 测试集的进行预测的效果很差，但是两者的特征具有迁移性[35]。这一结论与 Zhan 等是类似的。不同的是，Zhan 等指出这种迁移性在早期场景的网站指纹识别中并不可靠。

(七) 网站指纹识别应用研究

网站指纹识别的应用研究不仅包括在实际网络场景中应用网站指纹识别技术，而且包括利用网站指纹识别技术为其他的研究任务服务，即迁移应用研究。例如，近年来，有学者尝试使用网站指纹识别技术分析协议的安全性。下面分别介绍这两方面的应用。

Kim 等设计了一种在实际网络场景中收集和提炼 Tor 流量的框架。基于此框架，作者从实际场景中收集了普通网站和隐藏服务的流量，采用 XGBoost、决策树和随机森林进行评估。结果表明，隐藏服务的指纹识别难度更大。同时，该项研究表明，对于小规模网站的实际场景网站指纹识别是有可能达成的[79]。Ma 等则提出 2 阶段情境感知的系统方法解决实际加密代理环境中网站指纹识别问题。与基于访问的传统方法不同，作者采用基于流的方法。在基于 Shadowsocks 和 V2Ray 的模拟数据集测试中，本文的方法取得了比传统方法更好的效果[80]。该方法仅在较小的数据集规模取得成功，然而却佐证了 Kim 等的研究结论。

对比两项工作，Kim 等的研究对象仍然基于完整的访问。然而，Ma 等的研究对象则是基于流。这也是他们方法在实际应用场景中可行的关键所

在。在研究场景上，与 Kim 等针对 Tor 开展研究不同，Ma 等的工作针对加密代理。在实验效果上，Ma 等的效果远优于 Kim 等。从网站类型上看，Kim 等不仅研究了一般网站，也研究了隐藏服务。而 Ma 等仅针对一般网站。

在网站指纹技术的迁移应用方面，Siavoshani 等利用网站指纹识别对 TLS 协议的安全性进行了研究。具体地，作者提出一种迭代框架，通过基于机器学习方法的网站指纹识别找到信息泄露的关键数据域。结果表明，TLS 握手、TLS 记录长度和初始化向量字段是该协议的关键泄露部分。该框架也可以推广用在分析其他协议的安全性，有助于设计出更安全的协议[81]。

4 网站指纹防御研究进展

网站指纹防御研究是网站指纹研究的另一重要研究方向。本节首先介绍了网站指纹防御研究的基本情况。在此基础上，分别引入了评价指标和分类方法的相关内容。进一步地，本节分别对网站指纹防御的主体研究和辅助研究进行了综述。

4.1概述

网站指纹防御研究是通过设计改变流量特征的技术，使得网站指纹识别的准确率下降到安全的范围内。衡量网站指纹防御技术的好坏有其相应的评价指标。随着网站指纹识别研究的发展，网站指纹防御研究也不断推陈出新，逐渐集成了丰富的技术种类，产生了一系列的防御方法。这些防御方法可以从不同的角度进行分类。下面分别对网站指纹防御研究的评价指标和分类方法展开阐述。

4.1.1 评价指标

网站指纹防御方法的主要评价指标有两种：一是性能，即网站指纹识别的准确率下降了多少；二是开销，反映了实现防御方法所需要的代价。此外，网站指纹防御方法还有其他的评价指标，如是否需要修改浏览器或覆盖网络（如 Tor、SSH 等）实现代码、是否需要额外的先验知识或基础设施（如数据库或额外的服务器等）支持和是否具有灵活可配置性等。例如，有的防御方法需要修改浏览器代码以提供支持，如 Walkie-Talkie、HTTPOS 等。有的防御方法需要提前知道页面的先验知识，因此需要数据库的支持以维持页面的动态变化所需要的存储资源，如 Supersequence、Glove 等。有的防御需要额外的辅助服务器支持，如 BuFLO、LLaMA 等。所有的额外要求均增加了网络防御方法的可应用和可部署的难度，也是研究人员需要考虑的重要因素。最后，是否具有灵活可配置性则反映了防御方法在性能和开销之间的权衡能力。

网站指纹防御方法的性能评价指标与其实验设置有关。大部分网站指纹防御方法的实验评估是在封闭世界假设下进行的，其性能评价一般采用封闭世界的准确率指标，即在应用防御方法之后，同一网站指纹识别方法的准确率下降了多少。显然，下降得越多，防御性能越好。网站指纹防御开销包括时间开销和带宽开销两个部分。时间开销也就是时延大小。不同的防御方法的本质差异是在开销和性能之间的权衡策略。

4.1.2 分类方法

与网站指纹识别研究类似，网站指纹防御的主体研究是对防御方法的探索。防御方法的研究目标是隐藏流量的关键特征，同时保证较低的时间和带宽开销。除了防御方法的研究之外，研究人员同样也做了相关的辅助研究，如研究防御方法的安全上下界分析、流量特征信息泄露的分析等。

网站指纹防御方法根据不同的角度，可以有不同的分类方法。按照防御技术所部署的网络层次，可以将网站指纹防御方法分为网络层、应用层和复合层等。其中，复合层防御需要同时在网络层和应用层实现一定的防御技术。此外，网站指纹防御也可以划分为可模拟防御和不可模拟防御、有限防御和通用防御、确定性防御和随机防御等[24]。

4.2网站指纹防御方法研究

本文根据防御技术所部署的网络层次，将网站指纹防御方法分为网络层、应用层和复合层防御，并整理在表 6 中。应用层防御方法需要直接对未加密应用数据即明文数据进行操作，一般运行在浏览器或者网站服务器。网络层防御方法一般需要防御客户端和服务器的支持。复合层防御方法则是结合了网络层防御和应用层防御技术。

网络层防御和应用层防御各有优势和不足。一方面，由于应用层防御方法直接面向明文数据，所以在应用填充防御技术时，相对网络层防御方法更加精确，也更简单。然而，应用层防御要修改浏览器或网站服务器代码，影响了其实际应用范围。另一方面，网络层防御方法施加的防御仅限于防御客户端和服务器之间。因此，防御客户端和服务器之间通信链路之外的流量不再有防御技术带来的各种开销。这种优势对于带宽有限的网络场景(如 Tor 等)是十分重要的。然而，网络层防御方法需要防御客户端和服务器内部运行协议转化，影响了一定的效率。

网络层和应用层防御在具体实施时，虽然总体目标都是为了隐藏有区分能力的特征，然而各自都分别基于不同的思路和技术原理展开设计，因此产生了各种各样的防御模式。此外，无论是网络层或是应用层防御方法，防御技术既可能只部署于防御客户端，也可能只部署于防御服务器，甚至可能同时部署于防御客户端和服务器。

4.2.1 应用层防御

为了有效地隐藏流量特征，应用层防御通过采用明文数据重构、HTTP 请求重构和流量分割等思

路来实现。明文数据重构包括数据填充、数据分割等防御模式。HTTP 请求重构包括延迟请求、调整请求次序和插入冗余请求等防御模式。防御方法往往使用了其中的一种或多种防御模式。

（一）基于明文数据重构的防御方法

基于明文数据重构的防御方法是通过对应用数据即明文数据直接操作，使得在明文数据这一层级便消除了一些可区分的流量特征。属于这一类别的防御方法有明文填充、ALPaCA 和 HTTPOS。

明文填充。明文填充是指应用层数据在加密之前即使用一定的填充机制，达到混淆明文长度的目的。常见的明文填充有两种，即 Session Random 255 Padding 和 Packet Random 255 Padding[11]。前者是对整个会话使用同一个随机填充值。后者则是对每个包生成一个随机填充值。两种防御方法的随机填充值都是从集合{0,8,16,..., 248}中随机选择一个数字得到。

ALPaCA。ALPaCA 是由 Cherubin 等提出的一种基于 Web 服务器的防御技术[82]。它是通过填充和创建网页新内容来达到隐藏应用层的特征的。

表 6 网站指纹防御方法概览

防御类别	防御方法	防御模式
应用层防御	明文填充[11]	数据填充
应用层防御	ALPaCA[82]	数据填充
应用层防御	HTTPOS[83]	数据分割
应用层防御	LLaMA[82]	请求延迟
应用层防御	RP[84]	请求乱序
应用层防御	TrafficSliver-App[29]	流量分割
网络层防御	Splitting[85]	流量分割
网络层防御	HyWF[86]	流量分割
网络层防御	TrafficSliver-Net[29]	流量分割
网络层防御	Camouflage[18]	流量合成
网络层防御	Glue[87]	流量合成
网络层防御	Morphing[32]	流量模拟
网络层防御	Padding[11][27]	包填充
网络层防御	PMP[27]	包填充
网络层防御	PDP[27]	链路填充
网络层防御	AP[88]	链路填充
网络层防御	WTF-PAD[89]	链路填充
网络层防御	DFD[90]	链路填充
网络层防御	Front[87]	链路填充
网络层防御	BuFLO[11]	包填充、请求延迟、链路填充
网络层防御	CS-BuFLO[91]	同上
网络层防御	Tamaraw[23][70]	同上
网络层防御	DynaFlow[92]	请求延迟、链路填充
网络层防御	Supersequence[24]	SCS
网络层防御	Glove[93]	SCS
网络层防御	Mockingbird[94]	对抗样本
网络层防御	BANP[95]	对抗样本
网络层防御	Dolos[96]	对抗补丁
复合层防御	W-T[97]	链路填充

该技术不仅有效而且便于部署。ALPaCA 包含 P-ALPaCA 和 D-ALPaCA 两种变种。给定一个页面大小集合，两种变种防御都是从该集合中为原始的页面对象选择一个目标页面大小，并根据该目标页面大小对原始页面对象进行填充。对于不同的对象类型，防御技术分别填充不同的内容，以保证对象可读性。最后，对整个 HTML 的大小再次进行填充，以保证一致性。于是，原始页面将与集合的来源页面类似，从而起到了保护作用。P-ALPaCA 和 D-ALPaCA 的不同在于，前者是基于概率的填充方案。后者是确定性的填充方案，实现技术较为简单，但具有较小的开销。

HTTPOS。HTTPOS(HTTP Obfuscation)是由 Luo 等于 2011 年提出的，实现在浏览器代码里的防御方法[83]。HTTPOS 包括四个模块。实现防御功能主要在第二个模块。模块二的原理是通过对每个非 MTU 的来包进行随机划分为二个子包来实现的。总的来讲，HTTPOS 巧妙地利用了 TCP 通知窗口、HTTP 管道和 HTTP 范围等协议机制来控制来去包的大小。

上述防御方法虽然都是基于明文数据重构，但是重构方法却不同，既有数据填充，也有数据分割。具体地，采用数据填充模式的防御方法有明文填充、ALPaCA 等。采用数据分割模式的防御方法有 HTTPOS 等。明文填充、ALPaCA 防御方法虽然都是对应用数据直接填充，但是实现方式不同。前者是随机值填充，后者是参照目标页面构造填充。从实现机制上讲，前者比后者简单的多。值得一提的是，ALPaCA 是基于 web 服务器的防御技术。这一点有悖于防御服务器一般不会部署在 web 服务器的常识。实际上，ALPaCA 防御是针对 Tor 的隐藏服务的。而在隐藏服务的服务端部署防御服务器是有现实需求的。原因在于 Tor 的隐藏服务对安全性要求更高，传播途径也与一般 web 服务不同。

（二）基于 HTTP 请求重构的防御方法

基于 HTTP 请求重构的防御方法是通过延迟请求、调整请求次序和插入冗余请求等操作使得 HTTP 请求序列的特征得到隐藏，从而实现防御的目的。该类别的防御方法有 LLaMA 和 Randomized Pipelining。

LLaMA。LLaMA 是 Cherubin 等提出的一种基于客户端的防御方法，其设计是受到 Randomized Pipelining 的启发[82]。两者的策略都是改变 HTTP 请求发送的次序。LLaMA 是以 Tor 浏览器插件的方式实现防御的。LLaMA 主要有两个技术要点。一是引入了请求延迟。二是增加了额外请求。同时，这些请求被发送到一个提供定制长度资源的服务器上。

Randomized Pipelining (RP)。RP 防御基于客户端，也是 Tor 浏览器上已经实现的一种防御技术[84]。它是通过随机化 HTTP 管道的深度来决定。即该技术对于每个连接包含的请求数量进行了随机化。因此，当请求的并发数量大于管道深度时，对于同一网站的多次访问的请求顺序可能发生变化，从而起到了一定的保护作用。

显然，RP 防御没有带宽开销，因为它没有引入额外的数据包。RP 目前已经在 Tor 客户端处理默认开启状态。然而，研究表明该防御并没有有效抵抗网站指纹识别[98]。

RP 和 LLaMA 防御方法虽然都是通过对 HTTP 请求重构达到防御目的，但是具体的策略不同。即 LLaMA 采用了延迟请求和增加冗余请求的策略，而 RP 则采用调整请求次序的策略。从防御效果上，LLaMA 通过引入额外的干扰流量以及随机请求的时延，使得识别的难度更大。

（三）基于流量分割的防御方法

在应用层进行流量分割，可选的方案是对 HTTP 请求进行分解。由于 HTTP 请求的对象通常包含多个，如果可以发起不同的 HTTP 请求分别获取这些对象，将对流量起到混淆作用。TrafficSliver-App 防御方法采用了该技术。

TrafficSliver-App。TrafficSliver-App 是基于流量分割的应用层防御方法，由 Cadena 等在 2020 提出[29]。该防御方法的原理是在客户端和 Tor 入口节点之间部署代理，截获客户端的流量并分给多条 Tor 链路发送或者汇聚多条来自服务端的多条链路的流量。该防御利用 HTTP 协议的特性将页面请求分割成多个不同对象请求，并确保通过不同链路传送，从而破坏了流量特征。

特别地，TrafficSliver-App 防御仅处理 GET 请求，不需要更改浏览器或 Tor 代码。然而，由于该防御需要看到明文内容，防御代理为了去除 TLS 加密，需要执行中间人识别。这是与一般的防御方法最大的不同。

4.2.2 网络层防御

与应用层防御方法类似，网络层防御方法也利用了多种不同的防御模式，如流量分割、流量合成、流量模拟、包填充、链路填充、延迟请求、最短公共超序列(Shortest Common Supersequence, SCS)和对抗样本等。基于流量分割的防御技术有 Traffic Splitting、HyWF；基于流量合成的防御技术有 Camouflage、Glue；基于流量模拟的防御方法有 Traffic Morphing；基于包填充的防御方法有 Padding、Probabilistic MTU Padding、BuFLO、CS-BuFLO、Tamaraw 和 DynaFlow 等；基于链路填充的防御方法有 Probabilistic Dummy Packet、AP、WTF-PAD、DFD、BuFLO、CS-BuFLO、Tamaraw、DynaFlow 和 Front；基于延迟请求的防御方法有 BuFLO、CS-BuFLO、Tamaraw 和 DynaFlow；基于 SCS 的防御方法有 Supersequence、Glove；基于对抗样本的防御方法有 Mockingbird、BANP 和 Dolos。其中，BuFLO、CS-BuFLO、Tamaraw 和 DynaFlow 防御使用了多种防御模式，且 CS-BuFLO、Tamaraw 均是由 BuFLO 改进而来的。所以，本文统称这几种防御方法为类 BuFLO 防御。

（一）基于流量分割的防御方法

与应用层的流量分割思想类似，基于网络层流量分割的防御方法通过将流量分散到多条链路上发送，来达到混淆流量特征的目的。该防御方法通常假定识别者往往只位于其中的一条链路上，或者不同链路上的识别者不会串通共谋[86]。不同于应用层流量分割，网络层流量分割处理的对象是数据包。

Traffic Splitting。该防御简称 Splitting，其基本思想是将一条链路的流量分担在多条链路上，使得识别者难以获得完整的流量，从而提高安全性。基于此思想，Cadena 等人在 Tor 上提出一种新的防御方法。即通过在 Tor 上建立多条电路，并按一定的策略分配流量到各条电路分别传输，最后在出口节点再把流量合并成在一起输送到用户。这样使得识别者大概率只能获取部分的流量，从而降低其识别的成功率。研究结果发现使用加权随机的分配策略可以使识别准确率降低 60%以上[85]。该防御方法几乎没有引入额外的时间和带宽开销。

HyWF。HyWF 是一种基于多址技术的网站指纹防御方法。HyWF 由 Henry 等提出，其实质是一种二路径调度策略[86]。该策略有三个“不固定”，即每次选择的网络路径不固定；每次分别使用不固定的概率选择网络路径；每次连续传输的包的数量是不固定的。

TrafficSliver-NET。TrafficSliver-NET 是由 Cadena 等提出一种网络层防御方法[29]。该防御方法基于 Tor 构建了新型的网络场景模型，即网络只有一个中间节点和出口节点，但是入口节点有多个。在链路建立阶段，它先建立传统三跳链路，再建立多个到中间节点的二跳链路。为了保存链路状态，TrafficSliver-NET 使用了 Cookie 机制并在防御客户端与服务器之间进行通信确认。为了保证防御链路之外的正常传输，中间节点需要有重排序机制。该防御还有个特点是用户定制策略，即流量要怎么分配是用户通过指令定制的。

综上，虽然 Splitting、HyWF 和 TrafficSliver-NET 都是采用流量分割的防御模式，但是在具体实现上存在区别。具体表现在：在防御模型上，与 Splitting 和 HyWF 相比，TrafficSliver-NET 所基于的 Tor 应用模型不同，即包括了多个入口节点而不是通常的一个。其次，TrafficSliver-NET 将防御代理扩展部署到了中间节点，这是前两个防御没有的；在路径数量上，HyWF 是固定的二种路径，其他两种都是可配置；在分配策略上，三种防御方法各不相同；在调度策略上，TrafficSliver-NET 实现了用户定制，其他两种则不具备。然而，三种防御均使用较少的开销，实现了较好的安全目标。

（二）基于流量合成的防御方法

基于流量合成的防御方法是指防御代理通过

技术手段干扰真实网页流量的准确提取，从而使得识别者无法有效实施网站指纹识别的防御技术。该类别的防御方法有 Camouflage 和 Glue。

Camouflage。Panchenko 等在 2011 年提出 Camouflage 防御，通过释放诱饵页面来干扰识别者[18]。该防御的特点是利用诱饵页面产生干扰流量对原始流量进行混淆，达到防御的目的。只要客户访问一个页面，也自动同时加载另一个诱饵页面。由于识别者获取的不是纯净的页面流量，从而对客户访问的页面起到了保护作用。因此，该防御也称为 Decoy Page 防御。

Glue。Gong 和 Wang 首次提出了 Glue 防御方法[87]。Glue 的设计思路与传统的防御方法不同，其主要思想源于流量切分的困难。Glue 是通过将连续访问的页面无缝地连接成一个整体，使得识别者无法知道其中有几个页面，准确的分割点在何处，从而无法实施有效的识别。Glue 防御实现具体包括三种模式，即 Front 模式、Glue 模式和 Back 模式。

Camouflage 和 Glue 的主要区别之一是流量合成的实现方式不同。前者是在真实的网站访问流量上添加背景流量，即真实流量与虚假流量的合成。这里产生背景流量的诱饵页面是从背景页面集合中随机产生的。这种随机性增加了同一页面多次访问的流量模式的差异性。后者是将真实的流量无缝地连接在一起，即真实流量本身的合成。

（三）基于流量模拟的防御方法

基于流量模拟的防御方法是通过对真实流量进行变形，使得识别者无法进行有效区分的技术。为了最大化模糊特征，所有的页面访问流量都是参考同一目标进行变形。典型的该类防御方法是 Traffic Morphing。

Traffic Morphing。该防御也简称为 Morphing[32]。它是通过对网页流实例包序列的唯一包长度进行随机填充使得包序列看起来像来自另一个网页，从而达到混淆的目的。假设 Morphing 要伪装一个包序列T。它首先需要从另一个包序列D中学习到其包长的分布概率P_r。基于该分布概率，Morphing 随机抽样并发送假包，直到防御后的T序列与D序列相似度达到某个阀值以下。

Morphing 防御只改变包序列的唯一包长度。它并没有覆盖序列长度、包顺序或包间间隔时间等特征。因此，它对依赖唯一数据包长度的识别方法特别有效，但对于依赖其他特征的识别方法就不再有效[24][34]。

（四）基于包填充的防御方法

基于包填充的防御方法与应用层防御的明文填充方法的基本思想是一致的。但是由于网络层防御看不到明文数据，只能对加密后的数据操作。然而，由于防御方无法得知加密方式，故在实施填充时，一般是需要防御代理自定义加密方式，附在原先网络包载荷之后。因此，基于网络层的填充相比应用层更复杂。

Padding。与明文填充不同，包填充是直接对网络层封装的数据包进行填充。根据填充的方法不同，可以分为 Linear Padding、Exponential Padding、Mice-Elephants Padding、Pad to MTU 和 Packet Random MTU Padding 等六种[11]。具体地，Linear Padding 是将所有包填充到与其长度最近的 128 的倍数的大小，或者是填充到 MTU。Exponential Padding 是将所有包填充到其长度的 2 次幂的大小，或者是填充到 MTU。Mice-Elephants Padding 是将所有小于 128 字节的包填充到 128 个字节，大于 128 字节的包填充到 MTU。Pad to MTU 是将所有包的大小填充到 MTU。Packet Random MTU Padding 则是从集合{0,8,16,...,MTU-L}中随机抽样一个数字作为数据包的填充值。其中，L为数据包长度。

Probabilistic MTU Padding (PMP)。与 Padding 防御不同，PMP 是一种基于概率的包填充方法，由 Zhuo 等于 2018 年提出[27]。它的基本原理是对于每个包，随机生成[0,1]之间的一个数q。如果q小于预先设定的概率p，则对该数据包进行填充。反之，则不填充。

可以看到，对于包填充方法，主要区别在于如何设计填充策略以平衡开销与性能。包填充方法只隐藏了包长度特征。对于包时间、包顺序和包间间隔时间等特征并不受影响。因此，它只能抵抗部分的识别。

（五）基于链路填充的防御方法

链路填充是指通过插入哑包使得不同网站的页面访问产生的包序列的总包数量是固定的，或者是某个参数的倍数等，从而使得包总数量这一重要的区分特征被隐藏，达到了防御的目的。使用链路填充的防御方法有 Probabilistic Dummy Packet、Adaptive Padding 和 WTF-PAD 等。

Probabilistic Dummy Packet (PDP)。除了 PMP 之外，Zhuo 等也提出了另一种防御方法 PDP[27]。两者都是基于概率的一种防御方法。所不同的是，前者是以概率p决定是否将包填充至 MTU。后者则是以概率p决定是否在当前包之后插入哑包，并不对当前包进行填充。

Adaptive Padding (AP)。该防御是 Shmatikov 和 Wang 于 2006 年提出的一种防御方法[88]。AP 定义了一系列的箱子b_i。其中，每个b_i所规定的时间范围是2^{i-1}至2^i。特别地，b_1所对应的时间范围是 0-2ms。在实现时，AP 根据下一个包的到达时间与箱子规定的时间范围的关系，决定是否发送哑包。AP 有两种工作模式，即 burst 模式和 gap 模式。每种模式使用不同的延迟分布。AP 的优点是不延时数据包的发送，而是立即发送。

WTF-PAD。WTF-PAD(Website Traffic

Fingerprinting Protection with Adaptive Defense)是一种专门针对 Tor 的防御方法，由 Juarez Marc 等提出[89]。WTF-PAD 防御方法基于 AP 防御。但是，WTF-PAD 比 AP 增加了一种状态机，即当防御客户端接收到防御服务器的消息事件发生时，防御客户端将触发填充动作，进一步混淆了特征。其次，用户发起请求后，WTF-PAD 将唤起防御服务器进行填充，以混淆页面大小特征。

DFD。针对 WTF-PAD 抵御基于深度学习技术的识别效果不佳的问题，Abusnaina 等提出 DFD(Deep Fingerprinting Defender)防御方法，较好地解决了该问题。DFD 的改进主要在于哑包注入的机制不同。DFD 包含注入监视和注入缓冲两个模块。注入监视用来记录上一次的突发长度，并推算出当前注入的包数量。注入缓冲则用来记录上一次确认包，用以当前注入。结果表明，DFD 仅用 14%的带宽开销就可以导致 86%以上的误分类率[90]。

Front。Front(Front Randomized Obfuscation of Network Traffic)防御方法是由 Gong 和 Wang 提出的[87]。考虑到流头部含有丰富的特征，Front 的设计思想是通过隐藏流头部特征来实现防御的。Front 的关键技术有四点：一是对于防御客户端和服务器分别设置不同的哑包数量，且该数量是随着流实例的不同而动态变化。当页面停止加载后，防御客户端将发送一个通知。所有未发送哑包将被丢弃，从而增加了不确定性，提高了安全性。二是哑包的插入窗口也是与流实例相关的。窗口的位置也是从一个均匀分布里抽样的。三是哑包的插入时间同样需要随着流实例的变化分别从给定的瑞利分布里抽样得到的。四是所有包的发送都没有延迟。

在链路填充的作用下，流实例的包顺序、包时间和包间时间间隔特征发生了变化。由于引入了额外数据包，混淆了原来的包长度特征。所以，链路填充的安全性比包填充更好。链路填充策略可以分解为二个子策略，即如何设计包序列？如何确定包时间戳？对于第一个问题，AP 和 WTF-PAD 都是基于某个目标网页的包序列分布。Front 则根据不同的流实例动态变化包序列。PDP 的设计比较简单，即通过预设的随机概率来控制包序列的生成。DFD 的哑包数量则由上一次的突发长度决定。对于如何确定包时间戳，AP 和 WTF-PAD 通过设置定时器来实现。Front 则是从动态的瑞利分布里抽样得到。PDP 和 DFD 的哑包时间戳均是由具体的插入时间决定。

此外，WTF-PAD 实际上是在 AP 基础上进行了升级，增加了对页面大小等特征的混淆，同时也引入了更多不确定性，提高了安全性。WTF-PAD 防御对基于传统机器学习技术的识别方法防御效果较好，但却无法有效抵御基于深度学习技术的识别方法。原因在于该防御的填充机制启动时机较有规律，而且可能泄露两个连续到来的实际突发。这种信息容易被神经网络学习到。DFD 则针对 WTF-PAD 的这些不足，在注入机制上进行了改进，有效地抵抗了基于 DNN、CNN 等典型神经网络结构的指纹识别。

（六）类 BuFLO 防御方法

BuFLO 是最早的通用防御。它集成了包填充、延迟请求和链路填充等防御模式。其中，延迟请求的目的是为了调节包的发送速率，实际上可以看成是包的调度。BuFLO 是一种高开销的防御。为此，研究人员对其进行改进，提出了 CS-BuFLO 和 Tamaraw 防御方法。由于 DynaFlow 同样集成了延迟请求和链路填充防御模式，只是具体的实现方式不同，故可划分到该类别。

BuFLO 。 BuFLO(Buffered Fixed-Length Obfuscator)是 Dyer 等于 2012 年提出的[11]。它的目的是通过消除所有的侧信道信息来抵御任意的流量分析攻击。BuFLO 有三种固定，即包长固定、固定速率和固定时长。三种固定是分别通过d、ρ、τ三个参数来调整的。具体地，d决定了固定包长的大小，ρ决定了发包的频率，τ决定了发包的最小时长。协议假定填充的字节有标记以便接收端可以丢弃它们。虽然 BuFLO 提供了较高的安全性，但也存在一些问题。首先，它对于不同的网站开销也是不一样。对于包长小、加载时间短的网站，BuFLO 的开销很大。此外，BuFLO 无法自适应网速的变化。于是，对于不同的网络连接，BuFLO 需要重新调优参数。

CS-BuFLO。针对 BuFLO 不能适应不同网络连接的问题，Cai 等于 2014 年提出了 CS-BuFLO 防御方法[91][99]。首先，CS-BuFLO 引入了速率适配机制，以适应网络变化。在包填充上，CS-BuFLO 不是采用固定长度填充，而是将长度填充到 2 的幂次方大小，同时提供了载荷填充和总填充两种模式。防御客户端和服务器可以分别选择其中一种填充模式。对于如何判断页面已经加载完成，CS-BuFLO 提供了两种机制。为了进一步提高效率，CS-BuFLO 提供了防御客户端积极填充和服务器早期中止的填充交互模式。

Tamaraw。为了提高 BuFLO 的实用性，研究人员提出了 Tamaraw 防御[23][70]。Tamaraw 实际上是 BuFLO 的改进。Tamaraw 主要做的改进有：首先，与 BuFLO 对于上下行包使用相同的固定包长不同，Tamaraw 对下行包和上行包赋予了不同大小的固定包长。这样设计目的是为了减少带宽的开销。同样地，对于固定速率的设置，由于上行包的数量远小于下行包的数量，Tamaraw 对于下行包设置了更高的固定速率。最后，对于固定时长的设置，Tamaraw 不是设置唯一的固定时长，而是给出了一系列可选的固定时长，从而给调节开销提供了较大的灵活性。

DynaFlow。针对网站指纹防御存在的缺少安全性的形式化证明、高带宽和延迟和需要动态更新的数据库支持等问题，Lu 等提出了 DynaFlow 防御方法。该防御方法在保证相当安全性的同时，效率提高了 40%以上[92]。同时，DynaFlow 不需要数据库的支持，也可以将保护扩展到动态生成的网站。但是 DynaFlow 需要更改 Tor 的代码，影响了其实际部署。

DynaFlow 包含三个部分，即突发模式变形、动态变化间隔的恒定流量和突发数量的填充。在突发模式变形部分，DynaFlow 通过引入哑包填充将流整理成 4 个来包，1 个去包的连续循环模式。同时，DynaFlow 通过插入哑包，可以支持动态间隔的恒定流量。为了隐藏突发数量信息，DynaFlow 将突发数量填充到参数m的幂次方倍。此外，为了进一步优化性能，DynaFlow 将 2 个连续的且总长度小于 512 字节的包组成一个新的更大的包，以降低时延开销。

综上，CS-BuFLO 和 Tamaraw 都是为了各自的优化目标由 BuFLO 发展而来的。它们与 BuFLO 均包括了包填充、包调度和链路填充三种防御模式。CS-BuFLO 最重要的特点是适应不同网络速率的变化，且具有拥塞敏感性。Tamaraw 则大大减少了防御的开销，提高了实用性。DynaFlow 不使用包填充，但是在链路填充策略上比其他三种防御更加复杂。另外，DynaFlow 增加了对连续的小数据包的合成功能，这也是其他三种防御所没有的。然而 DynaFlow 为了降低开销需要修改 Tor 代码，一定程度上提高了部署的难度。

（七）基于 SCS 的防御方法

基于 SCS 的防御方法利用了集合的 SCS 具有最优带宽的特点。该防御方法的难点在于求解 SCS 的困难性[100]。典型的该类防御有 Supersequence 和 Glove。

Supersequence。Wang 等在 2014 年提出了 Supersequence 防御以抵抗 KNN 识别[24]。Supersequence 目标是寻求一种带宽最优的可模拟防御技术。Supersequence 防御的基本原理是通过计算包序列集合的 SCS 得到最优防御的输出包序列。该防御首先把包序列划分到不同的匿名集中。通过计算每个匿名集合中的 SCS，即可得到在给定准确率前提下的整个数据集上的带宽最优防御方法。然而，计算多元素集合的 SCS 是一种 NP 难问题。因此，Supersequence 采用了近似的方法。Supersequence 防御需要数据先验知识，而且匿名集的选择是其一大困难。

Glove。Glove 是一种基于 SSH 的防御方法，由 Nithyanand 等提出[93]。由于之前防御方法不知道在何时增加覆盖流量，导致通常的策略是全部增加，从而引起了大量的不必要的开销。Glove 则是充分利用已有流量的知识，在保证安全性的情况下，减少了开销。具体地，Glove 首先对每个网页选择一个代表实例，利用网页的代表实例对网页使用 k-medoids 方法进行聚类，并对每个类别选取一个 super-trace，最后联合各类 super-trace 求出一个最终的 super-trace。Glove 防御具有可证明的防御上界，并具有高度可调节性。

Supersequence 和 Glove 都是基于集合 SCS 的特点而设计。因此，它们都需要对数据集有一定的先验知识。它们主要的区别在于匿名集的选择方法。

（八）基于对抗样本的防御方法

对抗样本这一概念最早由 Szegedy 等在 2014 年提出[101]。它的基本原理是对输入样本故意添加一些人无法察觉的细微的干扰，导致模型以高置信度给出一个错误的输出。典型的该类别防御方法包括 Mockingbird、BANP 和 Dolos 等。

Mockingbird。Mockingbird 是基于对抗样本思想的防御方法，由 Imani 等提出[94]。它实际上是一种产生可以抵御对抗性训练的对抗样本的技术。把每个实例看成是突发序列，Mockingbird 是通过对该序列施加最小的扰动，使得对它的判别偏差到错误的类别，从而降低分类器的准确率，达到防御的目的。实验表明，在存在对抗性训练的情况下，Mockingbird 仅引入了 56%带宽开销却仍能使识别准确率由 98%下降到 29%。

为了实时生成扰动样本，Nasr 等提出了盲对抗性网络扰动(Blind Adversarial Network Perturbations, BANP)防御。该防御通过解决特定的优化问题使得扰动生成与目标输入无关。此外，BANP 防御引入了重映射函数和正则化项，使得扰动的生成符合了必要的约束条件。针对扰动现有包时间、包长，注入扰动包等情况分别使用不同的映射方法。在流相关识别和网站指纹识别的实验中，BANP 取得了比其他与输入相关的防御方法更好的性能。然而，在对抗性训练下，BANP 的防御性能急剧下降[95]。

与对抗样本的思想类似，学者 Shan 等提出的 Dolos 防御方法利用了计算机视觉里对抗补丁的思想。与 BANP 防御一样，Dolos 防御的对抗补丁是通用的，无需预先知道全局输入。同时，该对抗补丁也是位置不可知，即对抗补丁的插入位置可以是任意的，而不影响结果。Dolos 防御包括补丁生成和补丁注入两个模块。补丁生成模块的目标是在给定带宽开销约束的前提下，对一条源网站的迹注入补丁，使得该新迹的特征与源迹的特征接近以至于无法区分。补丁注入模块则是对补丁进行实时注入和混淆，以保证对抗补丁不依赖输入与位置，也可抵抗逆向分析。实验结果表明，在对各种分类器的防御表现中，Dolos 防御可以提供 94%以上的保护

性能，即便分类器进行了对抗性训练[96]。

对抗样本是对抗机器学习研究里的一种重要思想。在网站指纹防御领域应用对抗样本的思想是近几年新出现的研究方向。基于对抗样本思想的防御方法除了考虑带宽开销和防御性能之外，还需要考虑两个方面。一是实时性，即防御序列的生成是否与输入有关。二是抵御对抗训练的能力，即在敌手进行对抗训练的情况下，防御性能受到影响的程度。

以已有的研究工作为例，Mockingbird 的关键是提出了一种扰动源实例的策略，其防御序列的生成是需要在输入已知的情况下。该策略可以使得识别准确率显著下降，即便在存在对抗性训练的前提下。Nasr 等的 BANP 防御突出了实时性，即防御样本的生成与输入无关。为了不影响底层流量的正常通信，作者定义了重映射函数和正则化项以控制扰动的生成约束。由于其扰动生成方法相对稳定，也缺少随机数的调节，使得该防御方法抵抗带对抗性训练的识别的能力较弱。Shan 等的 Dolos 防御借鉴了对抗补丁的思想，弥补了 BANP 和 Mockingbird 防御各自的不足。Dolos 防御可以抵御对抗性训练的关键在于对抗补丁的生成依赖了用户端的秘钥。即使敌手获得了 Dolos 的源码，也无法获取用户端秘钥。于是，敌手对抗性训练基于的样本与用户生成的防御样本是完全不同的。此外，Dolos 防御具备抵抗逆向分析的能力。

4.2.3 复合层防御

复合型防御是指兼具网络层和应用层防御模式的防御方法。该类防御的典型代表是 Walkie-Talkie。

Walkie-Talkie(W-T)。该防御是一种带宽可调，时间开销很低的防御方法[97]。W-T 以半双工方式工作，即客户端仅在 web 服务器响应之前的所有请求之后才继续发送请求。因此，该防御大大降低了识别者可获取的特征集合。W-T 同时支持随机序列填充机制，包括填充突发和添加假突发。填充突发是指对每个突发插入随机数量的数据包，以此隐藏了每个突发真实的数据包。添加假突发是指在部分真实的突发之间插入假突发，以掩盖真实突发位置所反映出的规律。

W-T 需要修改浏览器的代码，以实现半双工通信。显然，W-T 的引入将使得浏览器的用户体验下降，这增加了其应用部署的难度。

4.3其他网站指纹防御研究

除了上述网站指纹防御的主体研究之外，研究人员也从其他角度开展了网站指纹防御的相关研究，主要包括安全性和带宽界限分析、特征信息泄露分析、网站指纹识别性分析和系统化防御评估方法等。下面本文将对其中典型的相关研究进行介绍。

（一）安全性与带宽界限分析

安全性与带宽是防御方法重要的评价指标。研究安全性、带宽和防御方法这三者之间的内在关系是安全性与带宽界限分析研究的目标。

针对防御方法缺少形式化的安全证明，基于给定的特征集，Cherubin 提出了一种评估防御方法的安全下界的方法[102]。该下界与使用的识别方法无关。根据得到的安全下界，该文提出了安全度量指标。这个度量指标仅与特征集有关。于是，该成果促使网站指纹研究人员聚焦于最佳特征发现的问题。Cai 等开展了类似的研究[70]。该研究工作主要有两项成果。一是在给定安全水平的前提下证明了所有可能防御的带宽下界。二是在给定封闭世界的性能情况下，提出一种评估防御的数学框架。

（二）特征信息泄露分析

随着网站指纹研究的发展，研究人员逐渐提出新的指纹特征。现有的网站指纹研究文献提出了多达 35683 种特征[33]。实际上，特征是可以通过各种方法迭代构造的。因此，从某种程度上讲，特征的数量是无限的。针对流量的特征数量庞大的问题，如何判断各特征的区分能力和重要性是网站指纹研究的一个问题。一直以来，对特征的评估往往是通过对照实验验证推断出来的。这样不仅工作量巨大，而且比较笼统。为此，Li 等提出了 WeFDE 的特征信息泄露分析方法，可以定量计算每个特征泄露的信息量，为网站指纹防御方法的研究指明了方向[103]。

（三）网站指纹识别性分析

网站指纹识别性分析是通过分析单一网站抵御识别的能力，找到导致其脆弱性的关键特征。网站指纹识别性分析对于网站指纹防御的设计可以起到指导作用。相关的工作主要有两项。

因应 Tor 洋葱服务敏感性高、数量少而容易被识别的问题，Overdorf 等通过网络层和网站层特征分析，采用集成分类器研究了 482 个洋葱服务的指纹识别性[104]。研究结果表明不同洋葱服务的指纹识别性差异很大，即网站指纹识别对所有的洋葱服务并不具有同样的威胁。网站指纹识别的准确率是一种平均性能。

Oh 等同样分析了网站的指纹识别性[17]。与 Overdorf 等不同的是，Oh 等采用多层感知机对各网站的样本进行训练。其次，在特征的使用上，Oh 等仅采用了网站 HTML 文件里的链接和内嵌内容的统计信息。在评估方法上，Overdorf 等和 Oh 等分别使用了 F1 得分和准确率做为评价指标。

（四）系统化防御评估方法

为了评估防御技术对识别方法的抵抗能力，常规的做法是通过实验验证。然而，Wang 和 Goldberg 首次提出一种基于特征的比较方法，促进了网站指纹研究的系统化[23]。首先，作者分析了识别方法对

各特征的敏感性。进一步，本文通过分析防御方法隐藏了哪些特征，从而判断出防御方法对识别的抵抗能力。该系统化分析方法相较实验验证更加有针对性。

5 研究挑战与展望

网站指纹研究经过二十几年的发展，无论是在网站指纹识别还是防御，均取得了一些重要的成果。对于网站指纹识别来说，分类器的数学模型涵盖了大多数的传统机器学习方法和深度学习方法，用于分类的特征数量多达数万种，研究的应用场景覆盖了主流的具有隐私增强技术的网络场景，也取得了令人瞩目的性能。同样，对于网站指纹防御来说，防御模式种类多样，防御技术的部署更加灵活多样，防御的效果也是较为显著。甚至，部分防御方法已经成为可应用产品，如 RP、WTF-PAD 防御。然而，网站指纹的实际应用仍然存在较多的挑战和必须解决的问题，主要包括以下几个方面。

5.1问题与挑战

5.1.1 网站指纹理论研究不足

无论是网站指纹识别或防御，目前都是主要以实验验证为研究方法。网站指纹识别通常的研究过程是：基于给定的数据集，选择或提取流量的一组特征，再应用多种机器学习方法对数据集进行分类。根据相应的性能指标评价各分类器的优劣。该研究方法主要存在两点不足。一是特征选择没有指导性理论。选择哪类特征、多少特征往往需要多次尝试验证得到。虽然 Li 等在文献[103]提出评估特征信息泄露的方法，但是该方法对于不同的数据集，特征的排序会有所不同。因此，该研究仍然缺少与数据集无关的评估方法。此外，选择多少特征可以得到最佳的分类效果也无法确定。因此，在同一个分类器和数据集的情况下，随着特征数量的增加，性能不一定是单调增加的。二是对于应该使用哪种机器学习模型缺少理论牵引，同样只能通过实验测试。实际上，应该把机器学习模型与特征分析结合起来，提出一套模型与特征如何匹配的理论分析方法。

对于网站指纹防御，同样存在类似的不足。评估不同防御方法对识别的抵抗能力，一般都是通过实验评估得到的。Wang 虽然提出了一种系统化的研究方法，但是该方法研究的特征有限，而且仅分析了防御方法与特征的关系，并没有研究抗识别性能[23]。仅仅依靠实验评估防御方法的抗识别性能存在的主要问题有：一是由于特征的数量和识别模型种类繁多，仅通过给定数据集的几个实验，无法证明结论的普遍性，不利于防御方法的应用部署。二是由于缺少对现有实验结论的理论提炼与总结，使得对分类器和防御方法改进的研究没有方向性，不利于网站指纹研究的发展。

5.1.2 网站指纹研究假设过强

为了便于学术研究，网站指纹定义了场景假设、网页假设、用户行为假设和识别者假设等前提条件。为此，基于上述假设的网站指纹研究受到诟病，其离实际部署和应用也还有较大的差距。

在场景假设方面，有部分研究文献突破了封闭世界和开放世界的假设，开始考虑更贴近实际场景的情况。例如，Wang 在文献[4]不仅考虑了低基础发生率的情况，而且考虑了动态的基础发生率的应用场景，比一般的开放世界场景假设更加符合实际，也更加困难。然而，研究发现现有的顶尖分类器在接近实际的应用场景中表现差强人意，除非进行针对性的精度优化设计。

在网页假设方面，研究人员通常认为一段不太长时间的同一网页符合模板假设[2]。实际上，目前大多数的网页是动态，其模型是会变化的，而且不同的网页变化规律不同。所以，在同一段时期内，有些网页适合模板假设，有些网页则并不适合。网站指纹研究人为地假定所有网页均符合模板假设，势必会产生一定的模型偏差。其次，网页假设默认网站指纹研究对象是网站的首页。这种限制大大影响了网站指纹研究的适用性。

在用户行为假设，网站指纹研究对用户的访问行为进行了规定。无论是单标签浏览方式还是指定动作的多标签浏览方式均较为理想。而且，已有的多标签浏览方式的研究基本上都是考虑二标签的场景，但在实际生活中，个体的浏览习惯差异很大，大量存在多标签浏览的方式。

在识别者假设方面，完全先验知识的假设，即识别者具有和用户完全一样的知识信息，显然过于理想。已有研究表明，在识别者未有用户浏览器及其版本、浏览器设置、网络位置和浏览范围等先验知识的场景下，分类器的性能将大大降低。那么，研究如何在非完全先验知识场景中的网站指纹则是必须面对的现实问题。针对流量分割的假设，除了少量研究之外[5][6][28]，大部分的文献都默认没有背景流量，而且用户是逐页访问的。这种假设都不符合实际情况。

5.1.3 网站指纹实际应用部署进展缓慢

无论是网站指纹识别或防御，距离实用性的部署和应用仍然有较大的差距。网站指纹识别之所以未能实际部署和应用，除了网站指纹研究假设太强之外，还因为有两大问题未能解决，即数据过时和增量式网站指纹。

数据过时问题是指训练数据的采集时间距离应用时间太远，导致分类器模型性能降低或失效。那么如何利用旧的数据训练出一个可用的分类器则对网站指纹实用化具有重要的意义。实际上，学者 Sirinam 等在文献[73]已经提出了这一具有挑战性的课题，并取得了初步的效果。在跨度三年的训练

集和测试集上取得了 85%的准确率。然而，类似的工作数量较少。

网站指纹识别实用化另一个问题就是如何实现增量式网站指纹。增量式网站指纹实际也是数据过时问题的一种解决方法，即自动地增量采集时新数据并训练新模型用于实时预测。以什么样的方式增量采集新数据，以什么样的频率丢弃旧数据，如何应用旧模型逐渐迭代产生新模型都是需要研究的问题。然而，目前该研究仍是一片空白。

网站指纹防御方法的部署困难来自两方面的原因。一是需要多方的协同配合。对于应用层防御方法涉及到修改浏览器或网站服务器代码，或者在其上面嵌入插件，必须得到浏览器厂家和网站服务提供商的支持。对于网络层防御方法，同样需要在覆盖网络的中继节点部署防御代理，而网络上潜在的中继节点数量太过庞大。显然，无论是应用层还是网络层防御，落实其部署需要牵扯的面较广，推动部署十分困难。二是许多的防御方法天然不适合部署，缘于其开销代价过大，只能用于实验室研究，如 BuFLO 防御。

虽然目前 Tor 浏览器已经部署了 RP 防御，然而却被验证无法有效抵抗识别。而且，另外在 Tor 完成实验性部署的 WTF-PAD 防御，实际上也迟迟没有作为默认配置。这与其额外引入的开销代价是紧密相关的。因为 Tor 本身就已经具有较高的延迟了，所以开发人员和用户都没有很强的动力去推动其产品化。

5.2未来研究方向

综合以上网站指纹研究面临的问题与挑战，未来可以尝试通过发展网站指纹理论研究解决网站指纹理论研究不足的问题，提高理论在研究中的牵引作用。针对网站指纹研究假设过强的研究现状，通过弱化网站指纹的某些识别假设，开展相应的理论与实验研究，提高网站指纹研究的实用性。考虑到网站指纹实用部署的进展缓慢的问题，学者需要逐一解决必要的关键问题。具体地，对于网站指纹识别应用，除了深入开展弱化识别假设研究之外，还需要置重点于训练数据过时、增量式网站指纹和精度优化等问题的研究。此外，未来网站指纹识别研究可以突破其典型定义，虽然同样以准确识别关注的网页为目标，但可以不拘泥于实现的方式、场景和特征范畴等限制。

网站指纹防御的实用化则需要不断优化防御技术，在保证安全性的同时尽可能降低时延和带宽开销，以此提高相关方部署的意愿和动力。尤其要注意，不能以牺牲用户体验为代价。此外，随着基于深度学习的网站指纹识别方法的出现，传统手工提取特征的方法不再是唯一选项。越来越多的研究利用了神经网络的自动特征学习能力。面临这样的新变化，网站指纹防御技术也必然要推陈出新。下面对以上潜在的研究方向进行详细讨论。

5.2.1 发展网站指纹理论研究

针对网站指纹理论研究不足，为了使网站指纹识别和防御的研究更加有指导性，需要在以下研究方向上着手：一是强化特征理论的研究，包括特征的生成和扩展方法、特征的物理意义研究、特征独立性和冗余特征的排除、特征之间的相互干扰分析等。二是强化跨学科研究并借鉴研究成果。从机器学习模型本身出发，研究其对不同特征的敏感性，尝试探索由特征和模型之间的某些固定的匹配关系及其原因。例如，基于深度学习的识别方法对包长、包方向和包时间戳等原始特征比较敏感。而基于传统机器学习的识别方法对统计特征比较敏感。三是从防御方法原理出发，尝试推导其安全性和开销的上下界。四是联合防御方法开展特征的重要性分析、分类器对特征的敏感性分析，尝试从理论推导出防御方法对不同分类器的抵抗能力，并与实验结果对比。五是尝试从数据集本身出发推导识别准确率的上界。已有研究表明，针对同一数据集的不同识别方法的准确率可能存在一个上界。该上界对于网站指纹识别与识别的研究都有重要的理论的指导作用。

5.2.2 弱化网站指纹研究假设

网站指纹研究假设的弱化研究是实用化网站指纹的关键。该问题可研究的方向很多。总体思路是由易到难，从单一条件到多个条件组合的逐渐弱化，循序渐进地接近实际的应用场景。在场景假设上，除了向低基础发生率和动态的场景方向研究，也可以从应用场景的组合入手，即用户所依赖的 PETs 不是单一的覆盖网络，而是多个覆盖网络的级联。在网页假设上，要从主页向任意页的研究发展。在用户行为假设上，不仅要研究二标签访问情形，而且要向多标签访问和更复杂的用户动作设置情形发展。在识别者能力假设方面，在非完全先验知识场景的网站指纹识别研究上，可以探索更多的先验知识不足的可能。在复杂背景流量问题上，探索更准确的多应用混合流量分离和多标签访问流量的切割方法。

5.2.3 新型网站指纹识别研究

新型网站指纹识别研究是指以识别用户访问的网站为目标，但是实现方式突破了传统的网站指纹识别的定义限制。新型网站指纹识别的目标与传统网站指纹识别是一样的，但是在其他方面可以不一样。例如，在特征的获取方式上，传统网站指纹识别默认通过被动方式获取。新型网站指纹识别则可以采用主动方式获取；在特征的来源上，传统网站指纹识别通常基于单一网页的完整访问。新型网站指纹识别则可以基于流对象或者单一网页的完整访问的某一部分；在敌手的数量上，传统网站指纹识别通常假定只有一个。新型网站指纹识别可以

拓展研究两个以上共谋或非共谋敌手的情况；在关注的目标上，传统网站指纹识别一般基于主页假设。新型网站指纹识别可以延伸研究同一网站的网页或者同一类别的网页；在关联研究方面，新型网站指纹识别可以尝试与其他的流量分析技术结合起来；在新技术跟踪方面，新型网站指纹识别需要考虑如何在 QUIC 协议的潮流下，解决 QUIC 协议与 TCP 协议并存场景下，甚至是未来 QUIC 协议取代 TCP 协议场景下的网站指纹识别可能面临的问题。总体上看，该课题具有十分广阔的研究空间。

5.2.4 新型网站指纹防御研究

深度学习识别方法的出现改变了传统网站指纹识别方法的特征提取方式。在传统网站指纹识别方法中，敌手需要自己通过特征工程提取所需的流量特征，并输入到分类器进行训练和预测。然而，深度学习识别方法的特征学习是由深度神经网络结构中自动学习中获取的。这一特点不仅大大减少了敌手的工作量，也使得传统网站指纹防御方法无法知道真正重要的特征是什么，需要隐藏哪些特征。因此，针对深度学习识别方法的新型网站指纹防御方法显得迫在眉睫。目前，有学者提出使用对抗样本或对抗补丁的方法应对深度学习识别方法，并取得了一定的效果[94][95][96]。然而，研究人员可以尝试集成其他的方法以进一步提高效果，如，生成对抗网络等[105]。可以预见，随着深度学习识别方法的不断发展，探索新型网站指纹防御技术是未来重要的研究方向。

5.2.5 推动网站指纹研究的应用部署

针对网站指纹识别应用部署的痛点，除了弱化网站指纹研究假设开展相关研究之外，还需要从增量式分类器的构建和旧数据的利用两个方向同时研究，尽可能多地研究利用旧数据训练模型的有效方法，更要在构建自适应的增量式网站指纹分类器的研究上发力。此外，实际场景的网站指纹识别很多是阳性事件发生概率极低的情况，网站指纹识别方法必须进行精度优化后才能应用，以避免“基础发生率谬误”。因此，针对识别方法的精度优化也是需要重点研究的方向。

最后，下一步研究应该尝试推进一些网站指纹识别原型系统的落地，在实际中发现新的问题，并逐步加以解决。针对网站指纹防御部署的困难，未来可以尝试推动一些轻量级的防御方法在小规模的网络场景中部署和试用，以评估其实际效果与理论的差距，并加以改进，进而实现在更大规模的网络场景中应用。此外，有针对性地强化对网站指纹识别行为的监管也具有重要意义[106]。

6 结束语

网站指纹研究揭示了即使在隐私增强技术的保护下，用户的浏览兴趣隐私仍然存在被泄露的风险。同时也表明，通过精心设计防御方法，用户的隐私可以得到更可靠的保障。无论是网站指纹识别技术还是防御技术，只有善意的人们运用它，才能保护用户的隐私安全，维护健康的网络环境。网站指纹研究历经几十年的发展，通过一代代学者的努力，取得了十分丰富的成果，为下一步研究打下了坚实的基础。本文从网站指纹的基本概念着手，阐述了定义、识别假设、威胁模型、防御模型、分类方法和研究意义等内容。在此基础上分别对网站指纹识别和防御的研究进展和动态进行了分类综述，并进行比较分析。最后，本文提出了网站指纹研究的三大挑战，并指出了未来的研究方向。

参考文献

[1] Dingledine R, Mathewson N, SyversonP. Tor: The second-generation onion router//Usenix Association Proceedings of the 13th Usenix Security Symposium. San Francisco, USA, 2004:303-319.

[2] Juarez M, Afroz S, Acar G, et al. A critical evaluation of website fingerprinting attacks//Computer and Communications Security. Scottsdale, USA, 2014:263-274.

[3] LiberatoreM, Levine B N. Inferring the source of encrypted http connections//Computer and Communications Security. Alexandria, USA, 2006:255-263.

[4] WangT. High precision open-world website fingerprinting. IEEE Symposium on Security & Privacy,2020:152-167.

[5] CuiW, ChenT, Fields C, et al. Revisiting assumptions for website fingerprinting attacks//Computer and Communications Security. London, UK, 2019:328-339.

[6] XuY, Wang T, Li Q, et al. A multi-tab website fingerprinting attack//Proceedingsofthe34thAnnualComputerSecurityApplications Conference. San Juan, USA, 2018:327-341.

[7] Gu X, YangM, Luo J. A novel website fingerprinting attack against multi-tab browsing behavior//Proceedings of the 2015 ieee 19th international conference on computer supported cooperative work in design. Calabria, Italy, 2015:234-239.

[8] Abe K, Goto S. Fingerprinting attack on tor anonymity using deep learning//Proceedings of the Asia Pacifc Advanced Network. Hongkong, China, 2016:15-20.

[9] BhatS, Lu D, Kwon A, et al. Var-cnn and dynaflow: Improved attacks and defenses for website fingerprinting. arXiv preprint arxiv. org/abs/1802.10215,2018:1-19.

[10] BhatS, Lu D, Kwon A, et al. Var-cnn: A data-efficient website fingerprinting attack based on deep learning.Proceedings on Privacy Enhancing Technologies. 2019,2019(4):292-310.

[11] Dyer K P, Coull S E, RistenpartT, et al. Peek-a-boo, i still seeyou:Whyefficienttrafficanalysiscountermeasuresfail//IEEE Symposium on Security and Privacy.SanFrancisco, USA, 2012:332-346.

[12] FeghhiS,LeithDJ.Awebtrafficanalysisattackusingonlytiming information. IEEE Transactions on Information Forensics &Security, 2016,11(8):1747-1759.

[13] HayesJ,DanezisG.K-fingerprinting: A robust scalable website fingerprinting technique//Proceedingsofthe25thusenixsecuritysymposium. Vancouver, Canada, 2016:1187-1203.

[14] He G, YangM, Gu X, et al. Anovelactivewebsitefingerprintingattackagainsttoranonymous system//Proceedings of the 2014 ieee 18th international conference on computer supported cooperative work in design. 2014:112-117.

[15] Herrmann D, Wendolsky R, FederrathH. Website fingerprinting: attacking popular privacy enhancing technologies with themultinomialna ïve-bayesclassifier//IEEEInternationalConferenc e on Cloud Computing Technology and Science. Beijing, China. 2009:31-42.

[16] Jahani H, Jalili S. A novel passive website fingerprinting attack on tor using fast fourier transform. Computer Communications, 2016,96:43-51.

[17] Oh S E, Sunkam S, Hopper N. p-fp: Extraction,classifcation, andpredictionofwebsitefingerprintswithdeeplearning.Proceedings on Privacy Enhancing Technologies. 2019(3):191-209.

[18] PanchenkoA, Niessen L, Zinnen A, et al. Website fingerprinting in onion routing based anonymization networks//Workshop on Privacy in the Electronic Society. Waterloo. Canada, 2011: 103-114.

[19] PancenkoA, LanzeF, Zinnen A, et al. Website fingerprinting at internet scale//Network and Distributed System Security Sym- posium. San Diego, USA,2016:1-15.

[20] Rimmer V, PreuveneersD,JuarezM,etal.Automatedwebsite fingerprinting through deep learning//Network and Distributed System Security Symposium. San Diego, USA,2018:1-15.

[21] SirinamP, Imani M, Juarez M, et al. Deep fingerprinting: Undermining website fingerprinting defenses with deep learning//Proceedings of the 2018 acm sigsac conference on computer and communications security. Toronto,Cannada, 2018:1928-1943.

[22] WangT, Goldberg I. Improved website fingerprinting on tor// Workshop on Privacy in the Electronic Society. Bloomington, USA, 2013:201-212.

[23] WangT, Goldberg I. Comparing website fingerprinting attacks and defenses. Amersterdam, Netherlands: Technical Report 2013-30,2013.

[24] WangT,CaiX,NithyanandR,etal.Effectiveattacksandprovable defenses for website fingerprinting//Usenix Security Symposium. San Diego, USA, 2014: 143-157.

[25] WangT.Optimizingprecisionforopen-worldwebsitefingerprinting.ar Xiv preprint arXiv:1802.05409, 2018:1-16.

[26] ZhaoY, Ma X, Li J, et al. Revisiting website fingerprinting attacks in real-world scenarios: A case study ofshadowsocks//InternationalConference on Network and System Security. Hong Kong, China, 2018: 319-336.

[27] ZhuoZ,ZhangY,ZhangZL,etal.Websitefingerprintingattack on anonymity networks based on profile hidden markov model. IEEE TransactionsonInformationForensicsandSecurity,2018,13(5):10 81- 1095.

[28] WangT, Goldberg I. On realistically attacking tor with website fingerprinting. ProceedingsonPrivacyEnhancingTechnologies.2016 (4):21-36.

[29] CadenaWDL,MitsevaA,HillerJ,etal.Trafficsliver:Fighting website fingerprinting attacks with traffic splitting//2020 ACM SIGSAC Conference on Computer and Communications Security. Online, 2020:1971-1985.

[30] Luo Jun-Zhou, YangMing, Ling Zhen, et al. Anonymous communicationanddarknet: A survey. Journal of Computer Research and Development, 2019, 56(1):103-130(in chinese) (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(01):103-130).

[31] WagnerD, Schneier B. Analysis of the ssl 3.0 protocol//Proceedings of the Second Unix Workshop on Electronic Commerce. San Jose, USA, 1996:29-40.

[32] Wright C V, Coull S E, MonroseF. Traffic morphing: An efficientdefenseagainststatisticaltrafficanalysis//Proceedings of the Network and Distributed System Symposium.San Diego, USA,2009: 237-250.

[33] YanJ,KaurJ.Featureselectionforwebsitefingerprinting//Privacy Enhancing Technologies: volume 2018. Barcelona, Spain, 2018:200-219.

[34] CaiX,ZhangXC,JoshiB,etal.Touchingfromadistance:web- site fingerprinting attacks and defenses.//Computer and Communications Security. Raleigh, USA, 2012:605-616.

[35] Smith J P, MittalP, Perrig A. Website fingerprinting in the age of quic//Proceedings on Privacy Enhancing Technologies. Online, 2021 (2):48-69.

[36] Cheng H, AvnurR. Traffic analysis of ssl encrypted web browsing, arXiv preprint, 1998:1-12.

[37] HintzA.Fingerprintingwebsitesusingtrafficanalysis.//Internation al Workshop on Privacy Enhancing Technologies. Copenhagen, Denmark,2002:171-178.

[38] Sun Q, Simon D R, WangY M, et al. Statistical identification of encrypted web browsing traffic//Proceedings of IEEE Symposium on Security and Privacy. California,USA, 2002:19-30.

[39] ShiY, MatsuuraK. Fingerprinting attack on the tor anonymity system//LectureNotesinComputerScience(includingsubseriesLe ctureNotesinArtificialIntelligenceandLectureNotesinBioinform

atics). 2009:425-438.

[40] Lu L, Chang E C, Chan M C. Website fingerprinting and identification using ordered feature sequences//European Conference on Research in Computer Security. Athens, Greece,2010:199-214.

[41] Bissias G D, LiberatoreM, Jensen D, et al. Privacyvulnerabilities in encrypted http streams//Proceedings of Privacy Enhancing Technologies Workshop. Cavtat, Croatia, 2005:1-11.

[42] Qasem A, Zhioua S, Makhlouf K. Finding a needle in a haystack: The traffic analysis version//Proceeding of Privacy Enhancing Technologies: volume 2019. Stockholm, Sweden, 2019:270-290.

[43] Gu Xiao-Dan, YangMing, Luo Jun-Zhou, et al. Website fingerprinting attack based on hyperlink relations. Chinese Journal of Computers, 2015, 38(04): 833-845(in chinese)
(顾晓丹, 杨明, 罗军舟, 蒋平. 针对 SSH 匿名流量的网站指纹攻击方法. 计算机学报, 2015,38(04):833-845).

[44] AlshehariT, Zhioua S. An empirical study of web browsers' resistance to traffic analysis and website fingerprinting attacks. Cluster Computing, 2018, 21(4):1917-1931.

[45] GongX,KiyavashN,BorisovN.Fingerprintingwebsitesusing remote traffic analysis//Proceedings of the 17th ACM Conference on ComputerandCommunicationsSecurity.Chicago,USA,2010: 684-686.

[46] Shen M, LiuY, Chen S, et al. Webpage fingerprinting using only packet length information// 2019 IEEE International Conference on Communications. Shanghai, China, 2019: 1-6.

[47] Shen M, LiuY, Zhu L, et al. Optimizing feature selection for efficient encrypted traffic classification: A systematic approach. IEEE Network, 2020,34(4):20-27.

[48] Yu S, Zhou W, Jia W, et al. Attacking anonymous web browsing at local area networks through browsing dynamics. Computer Journal,2012(4):410-421.

[49] Rish I. An empirical study of the naive bayes classifier//IJCAI 2001 workshop on empirical methods in artificial intelligence: volume 3. Seattle, USA, 2001:41-46.

[50] ChangCC,LinCJ.Libsvm:Alibraryforsupportvectormachines. ACMtransactionsonintelligentsystemsandtechnology,2011, 2(3):1-27.

[51] CaiX.Websitefingerprintingattacksanddefensesonanonymitynetworks[Ph.D. dissertation]. New York, USA: Stony Brook University,2014.

[52] Breiman L. Random forests. Machine Learning, 2001,45(1):5-32.

[53] ChenT, Guestrin C. Xgboost: A scalable tree boosting system//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. San Francisco, USA, 2016:785-794.

[54] BlunsomP. Hidden markov models. Lecture notes, August, 2004, 15(18-19):48.

[55] SieversF, Wilm A, Dineen D, et al. Fast, scalable generationofhigh-qualityproteinmultiplesequencealignmentsusing clustal omega. Molecular Systems Biology, 2011,7(1):1-6.

[56] Eddy S. Hmmer user's guide: Biological sequence analysis using profile hidden markov models, version 2.2. HMMER User's Guide, 1998.

[57] GeurtsP, Ernst D, Wehenkel L. Extremely randomizedtrees.Machine learning, 2006, 63(1):3-42.

[58] Hinton G E, Osindero S, Teh Y W.A fast learning algorithm for deep belief nets. Neural Computation, 2006,18(7):p.1527-1554.

[59] LecunY,BottouL.Gradient-basedlearningappliedtodocument recognition. Proceedings of the IEEE, 1998,86(11):2278-2324.

[60] BhatS, Lu D, Kwon A, et al. Var-CNN: A data-efficient website fingerprinting attack based on deep learning.Proceedings on Privacy Enhancing Technologies. 2019(4):292-310.

[61] Rahman M S, SirinamP, MatthewsN, et al. Tik-tok: The utilityofpackettiminginwebsitefingerprintingattacks.arXivpreprint arXiv:1902.06421, 2019.

[62] Rahman M S. Using packet timing information in website fingerprinting. Rochester, USA: Rochester Institute of Technology,2018.

[63] WangM, LiY, WangX, et al. 2ch-tcn: A website fingerprinting attack over tor using 2-channel temporal convolutionalnetworks//2020 IEEE Symposium on Computers and Communications. Rennes, France, 2020:1-7.

[64] MaCheng-Cheng, Du Xue-Hui, Cao Li-Feng, et al. Burst-analysis website fingerprinting attack based on deep neural network. Journal of Computer Research and Development, 2020, 057(4):746-766(inchinese)
(马陈城, 杜学绘, 曹利峰, 吴蓓. 基于深度神经网络 burst 特征分析的网站指纹攻击方法. 计算机研究与发展,2020,57(4):746-766).

[65] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//IEEE Conference on Computer Vision & Pattern Recog- nition. Las Vegas, USA, 2016:770-778.

[66] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation, 1997,9(8):1735-1780.

[67] Marks S T, HalvemaanK. Website fingerprinting attacks against tor browser bundle: a comparison between http/1.1 and http/2. Technical report, 2017.

[68] MillerB,HuangL,JosephAD,etal.Iknowwhyyouwentto the clinic: Risks and realization of https traffic analysis//Proceeding of Privacy Enhancing Technologies.Amsterdam, Netherlands,2014:143-163.

[69] ReddyS.Measuringtheimpactofsiteconfigurationsonsitefingerprinting over the web and tor. Https://www.ideals.illinois.edu/handle/2142/108024, 2020

[70] Cai X, NithyanandR, Wang, etal. A systematic approach to

developing and evaluating website fingerprinting defenses// Proceedingsofthe2014ACMSIGSACConferenceonComputeran d Communications Security. Scottsdale, USA, 2014:227-238.

[71] CuiW. Website fingerprinting attacks[doctoral dissertation]. Oklahoma, USA: Oklahoma State University,2019.

[72] Ramezani A, Khajehpour A, SiavoshaniM J. On multi- session website fingerprinting over tls handshake. arXiv preprint arXiv:2009.09284, 2020.

[73] SirinamP, MathewsN, Rahman M S, et al. Triplet finger-printing: More practical and portable website fingerprinting with n- shot learning//Computer and Communications Security. London,UK, 2019:1131-1148.

[74] Oh S E, MathewsN, Rahman M S, et al. Gandalf: Gan for data-limited fingerprinting//Proceedings on Privacy EnhancingTech-nologies. Online, 2021(2):305-322.

[75] Taylor V F, Spolaor R, Conti M, et al. Appscanner: Automatic fingerprinting of smartphone apps from encrypted network traffic//2016 IEEE European Symposium on Security and Privacy. Saarbrücken, Germany, 2016:439-454.

[76] AlnaamiKM.Enhancingcybersecuritywithencryptedtrafficfingerpri nting[Ph.D. dissertation]. Dallas, USA: university of Texas,2017.

[77] Kwon A, Alsabah M, Lazar D, et al. Circuit fingerprinting attacks: passive deanonymization of tor hidden services//Usenix Security Symposium. Washington , USA, 2015:287-302.

[78] ZhanP, WangL, TangY. Website fingerprinting on early quic traffic. arXiv preprint arXiv:2101.11871,2021:1-30.

[79] Kim Y H, Ho L, Kim W G, et al. Poster: A pilot study on real-time fingerprinting for tor onion services//Proceeding of the Network and Distributed System Symposium. Online, 2021:1-3.

[80] MaX,ShiM,AnB,etal.Context-awarewebsitefingerprintingover encrypted proxies.//the Proceeding of IEEE Conference on Computer Communications. Online, 2021:1-10.

[81] SiavoshaniM J, Khajepour A H, Ziaei A, et al. Machine learning interpretability meets tls fingerprinting. arXiv preprint arXiv:2011.06304,2020:1-17.

[82] GiovanniC,JamieH,MarcJ.Websitefingerprintingdefensesat theapplicationlayer//ProceedingsonPrivacyEnhancingTechnologies. Minneapolis, USA. 2017, 2017(2):186-203.

[83] Luo X, ZhouP, Chan E W W, et al. Httpos: Sealing information leaks with browser-side obfuscation of encrypted flows.//Proceedings of the Network and Distributed System Symposium.San Diego, USA, 2011:1-20.

[84] PerryM. Experimental defense for website traffic fingerprinting [EB/OL]. Online, 2011.

[85] CadenaWDL, MitsevaA,PennekampJ,etal.Poster:Trafficsplittingtocounterwebsit efingerprinting//Proceedings of the2019ACM SIGSAC Conference on Computer and Communications Security. London, UK,2019:1-3.

[86] Henry S, GarciaavilesG, SerranoP, et al. Protecting against website fingerprinting with multihoming. Proceedings on Privacy Enhancing Technologies, 2020(2):89-110.

[87] Gong J, WangT. Zero-delay lightweight defenses against website fingerprinting//Proceeding of the 29th USENIX Security Symposium. Online: USENIX Association. Online, 2020:717-734.

[88] ShmatikovV, WangM H. Timing analysis in low-latency mix networks: Attacks and defenses//European Conference on Research in Computer Security. Hamburg, Germany, 2006:18-33.

[89] JuarezM,ImaniM,PerryM,etal.Towardanefficientwebsite fingerprinting defense//European Symposium on Research in Com-puter Security. Heraklion, Greece,2016:27-46.

[90] Abusnaina A, Jang R, Khormali A, et al. Dfd: Adversarial learning-based approach to defend against website fingerprinting//Proceedings of IEEE Conference on Computer Communications.Online, 2020:2459-2468.

[91] Cai X, NithyanandR, Johnson R. Cs-buflo:a congestion sensitive website fingerprinting defense//Workshop on Privacy in the Electronic Society. Amsterdam, Netherlands,2014:121-130.

[92] Lu D, BhatS, Kwon A, et al. Dynaflow: An efficient website fingerprinting defense based on dynamically-adjusting flows// Proceedings of the 2018 Workshop on Privacy in the Electronic Society. Barcelona, Spain, 2018:109-113.

[93] NithyanandR, Cai X, Johnson R. Glove:A bespoke website fingerprintingdefense//WorkshoponPrivacyintheElectronicSoci ety. Amsterdam, Netherlands,2014:131-134.

[94] ImaniM,RahmanMS,MathewsN,etal.Adv-dwf:Defending againstdeeplearning-basedwebsitefingerprintingattackswithadv ersarial traces. arXiv preprint arXiv:1902.06626,2019:1-16.

[95] Nasr M, Bahramali A, Houmansadr A. Blind adversarial network perturbations. arXiv preprint arXiv:2002.06495,2020:1-17.

[96] ShanS,BhagojiAN,ZhengH,etal.Areal-timedefenseagainst website fingerprinting attacks. arXiv preprint arXiv:2102.04291, 2021:1-18.

[97] WangT, Goldberg I. Walkie-talkie: An efficient defense againstpassivewebsitefingerprintingattacks//Proceedingsofthe2 6th USENIX Security Symposium. Dallas, USA, 2017: 1375-1390.

[98] WangT. Website fingerprinting: attacks and defenses[doctoral dissertation].Waterloo, Canada: Waterloo University,2015.

[99] X Cai, R Nithyanand, R Johnson. New approaches to website fingerprinting defenses. arXiv preprint arXiv:1401.6022,2014:1-13.

[100] JiangT andLi M. On the approximation of shortest common supersequencesandlongestcommonsubsequences//International Colloquium onAutomata,Languages,andProgramming.Szeged,Hungary,

1994:191-202.

[101] Szegedy C, ZarembaW, Sutskever I, et al. Intriguing properties of neural networks.//International Conference on Learning Rep-resentations.Banff, Canada, 2014.

[102] Cherubin G. Bayes, not naive: Security bounds on website finger-printing defenses//Proceedings on Privacy Enhancing Technologies. Minneapolis, USA, 2017:215-231.

[103] Li S, Guo H, Hopper N. Measuring information leakage in web-sitefingerprintingattacksanddefenses//Proceedingsofthe2018ACM SIGSAC Conference on Computer and Communications Security. Toronto, Canada, 2018:1977-1992.

[104] OverdorfR,JuarezM,AcarG,etal.Howuniqueisyour.onion? an analysis of the fingerprintability of tor onion services// Proceedingsofthe2017ACMSIGSACConferenceonComputeran d Communications Security. Dallas, USA, 2017:2021-2036.

[105] Juarez M. Design and evaluation of wfp techniques[Ph.D. dissertation]. Leuven, Belgium: Arenberg Doctoral School,2019.

[106] Sun Xue-Liang, Huang An-Xin, Luo Xia-Pu, et al. Webpage fingerprinting identification on tor: A survey. Journal of Computer Research andDevelopment, 2021, 58(8):1773-1788(in chinese)

(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(08):1773-1788).

ZOU Hong-Cheng,Ph.D. candidate. His current research interests include cyberspace security and machine learning.

SU Jin-Shu, Ph.D., professor, Ph.D. supervisor. His current research interests include cyberspace security and Internet architecture.

WEI Zi-Ling, Ph.D., assistant professor. His current research interests include cyberspace security,wireless communications, and edge computing.

ZHAOBao-Kang, Ph.D., associate professor. His current research interests include cyberspace security and computer networks.

XIAYu-Sheng, Ph.D.candidate. His current research interests include cyberspace security and anonymous communication systems..

ZHAONa, Ph.D. candidate.Her current research interests include cyberspace security and anonymous communication systems.

Background

Website fingerprinting research includes two aspects, namely website fingerprinting identification and defense. The former is a kind of traffic analysis attack. It aims to identify the website which a target user is surfing under the protection of privacy-enhancing technologies, e.g., Tor, SSH, Shadowsocks, VPN, and et.al., by collecting a lot of traffic samples, extracting all kinds of traffic features, training a reasonable mathematical model or DNN, such as SVM, KNN, RF, HMM, SDAE, CNN, LSTM, and so forth, and then do predicting. The traffic features could be obtained in a variety of ways, and contain different useful information for distinguishing websites. Finding the most informative features is of critical importance to reach the final goal. Website fingerprinting identification is passive, thus it does not require decrypting, modifying, delaying, or dropping the traffic.

As to website fingerprinting defense, it is the opposite side of website fingerprinting identification, makes every effort to hide or obfuscate informative features of traffic, which lessens the effect of website fingerprinting identification. Hence, the research of website fingerprinting defense can be classified into the topic of privacy protection. The defenses can be implemented in the application level and network layer, and deployed both in the client-side and server-side, even in the relay nodes of an overlay network. In summary, there are many kinds of defense modes, including data filling, data cutting, HTTP requests delaying and disrupting, packet padding, link padding, traffic splitting and merging, adversarial examples, SCS, and so on.

Since Wagner and Schneier first uncovered that the packet length feature reveals the information of traffic data, researchers have carried out extensive research in this area. They consider different communication scenarios, such as HTTPS, Tor, SSH, Shadowsocks, and so on, which protect user privacy at different levels. In every scenario, scholars devise different informative features, such as packet length, CUMUL, and et.al., to identify the websites. In all, the number of all the features amounts to more than 35,000. As for the classifiers, it begins with simply comparing the similarity of two traffic instances to make a decision, then further integrates with modern machine learning methods, especially the DNNs. The main obstacles for website fingerprinting identification originate from the lack of knowledge from the attacker to the target user. The related research is on the way.

In this work, to systematize website fingerprinting research, we begin by introducing its basic concepts, identification hypotheses, threat models, and research significance. After that, the latest and comprehensive achievements of website fingerprinting identification and defense are respectively carefully organized, detailed described, and compared. Finally, we summarize three challenges of website fingerprinting research and point out future research directions.

This work is supported by the National Key Research and Development Program of China (No.2018YFB0204301), the National Natural Science Foundation of China (No. 61972412),Science and Technology Innovation Plan of Hunan Province (No. 2020RC2047), General Scientific Research Project of Hunan Education Department (No. 19C0140), and General Project of Changsha Normal University (No. 2019xjzkpy16).